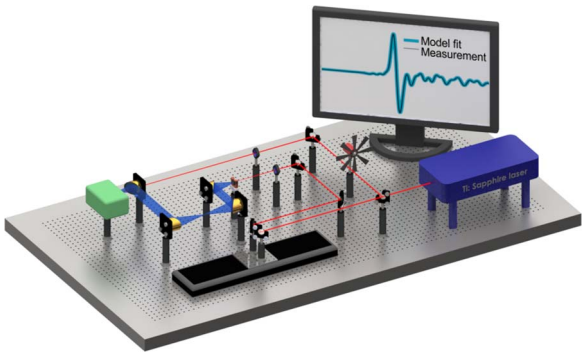
IEEE Photonics Journal
An IEEE Photonics Society Publication

Open Access

A Simple Transfer-Function-Based Approach for Estimating Material Parameters From Terahertz Time-Domain Data

Volume 6, Number 1, February 2014

K. M. Tych
C. D. Wood
W. Tych



DOI: 10.1109/JPHOT.2013.2292337
1943-0655

IEEE photonics SOCIETY

IEEE

A Simple Transfer-Function-Based Approach for Estimating Material Parameters From Terahertz Time-Domain Data

K. M. Tych,[1] C. D. Wood,[2] and W. Tych[3]

[1]School of Physics and Astronomy, University of Leeds, Leeds, LS2 9JT, U.K.
[2]School of Electronic and Electrical Engineering, University of Leeds, Leeds, LS2 9JT, U.K.
[3]Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, U.K.

DOI: 10.1109/JPHOT.2013.2292337

Manuscript received November 3, 2013; accepted November 6, 2013. Date of publication November 21, 2013; date of current version January 22, 2014. Corresponding author: W. Tych (e-mail: w.tych@lancaster.ac.uk).

Abstract: A novel parametrically efficient approach to estimating the spectra of short transient signals is proposed and evaluated, with an application to estimating material properties, including complex refractive index and absorption coefficient, in the terahertz frequency range. This technique includes uncertainty analysis of the obtained spectral estimates, allowing rigorous statistical comparison between samples. In the proposed approach, a simple few-parameter continuous-time transfer function model explains over 99.9% of the measured signal. The problem, normally solved using poorly numerically defined Fourier transform deconvolution methods, is reformulated and cast as a time-domain dynamic-system estimation problem, thus providing a true time-domain spectroscopy tool.

Index Terms: Terahertz, terahertz sensing, novel methods, modeling.

1. Introduction

The terahertz (THz) frequency range is commonly defined to lie between 0.3 and 10 THz, and corresponds to a range of fundamental material properties including: low frequency chemical bond vibrations, crystalline phonon vibrations, hydrogen-bond stretches and torsion motions [1]. THz time-domain spectroscopy (THz-TDS) is becoming an established technique for substance identification and for obtaining material parameters in the form of the refractive index and absorption coefficient spectra of a broad range of samples, including small molecules [2]–[4], biological samples [5], [6], and semiconductors [7], [8]. THz-time domain data often takes the form of a single pulse, less than a picosecond in duration, followed by a series of attenuated pulses arising either from reflections at the interface of components within the TDS system, or from etalon reflections within the sample itself. Typically, two separate measurements are taken: one serves as a reference and consists of a free-space, or empty sample cell measurement, and the second comprises what is normally assumed to be a convolution of the reference and the sample itself, often referred to as the "cross spectrum" when considering the frequency domain interpretation. To extract spectral estimates purporting to the sample alone, the reference signal must, therefore, be deconvolved from the second signal.

The goal of spectral estimation, in this case, is to describe the distribution (over frequency) of the power contained in a signal (the power spectral density, PSD), based on a finite set of data. In terahertz time-domain spectroscopy, the PSD of finite time-domain THz pulses are commonly estimated using the Fast Fourier Transform (FFT) or some variation [2]. Fourier Transforms operate on the principle that the measured signal, which has been recorded for a finite time period, is repeated indefinitely outside that time window. While the Fourier transform is therefore suited to the frequency analysis of periodic signals, a pulsed (transient) signal is likely to end at a discontinuity which the FFT will then include in the transform. Sharp discontinuities in the time-domain contain a broad frequency spectrum, and will therefore cause the signal's frequency spectrum to be spread out (spectral leakage). Furthermore, any given frequency in the output spectrum will contain energy both from the signal and from noise arising from the rest of the spectrum. The raw periodogram produced by the Fourier Transform is therefore considered a poor spectral estimate, as its variance at a given frequency does not decrease with the number of samples used to perform the transfer function, and is indeed equal to the square of its expectation value [9]. Variance and spectral leakage may be reduced by, respectively, smoothing of the signal and windowing of the time-series - which effectively multiplies the finite time domain signal with a periodic envelope. The resultant convolution is then Fourier transformed. The choice of the smoothing and window functions introduces a trade-off between the resulting spectral resolution (i.e., the ability to distinguish closely-spaced features), and how efficiently the spectral leakage is reduced [10]. These choices, which rely heavily on the interpretation of the user, can therefore affect significantly the resulting PSD resolution, bias and variance, leading to differences in the positions and intensities of absorptions/resonances in the frequency domain, potentially even obscuring weaker features which are in close proximity to stronger signals [11].

The method of deconvolving the data, i.e., dividing the cross-spectrum by the reference spectrum to obtain the sample spectrum, carries with it two strong assumptions. The first is the treatment of both objects as dynamic Linear Time Invariant (LTI) systems, namely that the system is fully characterized by its amplitude and phase responses (**L**inear), and that these amplitude and phase responses do not change in time (**T**ime **I**nvariant). The second assumption is that both spectral estimates are good in terms of their bias and inherent uncertainty. Duvillaret et al. address this by presenting a method of calculating physical sample characteristics from their spectral response, including estimates of the associated uncertainty [12]. Separately, Dorney et al., and later Pupeza et al. describe methods for improving the accuracy of material parameter estimation from time-domain signals, indicating the need for regularization of the poorly defined deconvolution process and the numerical problems encountered [13], [14].

In this paper, we suggest that a more accurate and robust approach would be to avoid deconvolution in the frequency domain - and its associated uncertainties - altogether, and instead, at the estimation stage, to deal only with the time-domain data (thus remaining true to the term "Time Domain Spectroscopy"). We demonstrate that, by treating the spectral response in the time domain as a dynamic system's impulse response with transient components, it is possible to reliably and easily obtain accurate frequency domain characteristics, directly comparable to those published in the literature, without the need - and therefore without the associated error - for windowing or padding of the recorded data. Additional information, in the form of uncertainty bounds, provides objective measures of quality of spectrum and thus has the potential to improve substance identification, for example, in THz spectroscopy.

2. Methods

2.1. Experiment

The time-domain data were collected using the broadband THz-TDS system at the Institute of Microwaves and Photonics, University of Leeds, described in detail elsewhere [2]. Pressed-pellet samples of two materials well-characterized in the THz-frequency range were chosen for study: α-lactose monohydrate (12.5% in a PTFE matrix) and glucose (100%). These samples were selected

as both, particularly lactose, are commonly used as test substances, as reported in the literature [15]–[23]. For each measurement, averages of five time-domain measurements taken from different points on the pellet samples were used for analysis, to minimize error arising from imperfect mixing and sample-thickness measurements. Both the reference THz signal, and that after transmission through the sample, were collected and focused onto an electro-optic crystal, enabling electro-optic (EO) sampling of the THz radiation in the time domain.

The EO crystal is only birefringent for the time that the THz field is present. This makes it possible to reconstruct a complete time domain signal from independent measurements of adjacent portions of a train of (ideally) identical THz pulses. In order to improve accuracy, and to reduce the noise in the spectrum, signal averaging was used. The acquisition time for each time domain spectrum is therefore limited by both the number of single measurements and the minimum time limit to any sampling technique ($t_{\min} \geq N \cdot \delta t$), where N is the number of data points required to measure the whole THz pulse, and δt is the pulse spacing in the pulse train [24].

In order for the time-domain data to be Fourier transformed into the frequency domain, it must first be truncated before the reflection peak arising from the EO crystal ($\sim$4.7 ps after the main signal) to remove spurious oscillations from the frequency spectrum [2]. The peak originates from etalon reflections within the EO crystal, and therefore, its position is determined by the crystal thickness. This truncation reduces the amount of data used in the FFT calculation, thus limiting the number of discrete data points in the resulting frequency spectrum. A typical, truncated data set may consist of, for example, 600 data points which, for an efficient FFT, is often truncated or padded to the next power of 2 (i.e., 512 or 1024). The effect of zero padding is to interpolate between actual data points in the frequency spectra and often produces a smoother response. Additional padding (to subsequent powers of 2) may be undertaken at the user's discretion to increase further the number of data points in the frequency spectra though this does not overcome the spectral resolution limitations inherent to the measurement and analysis, and may therefore mislead a user into believing that all spectral features have been resolved. Furthermore, care must be taken to avoid padding-artifacts by the introduction of sharp discontinuities. Following FFT, the resultant data may then be used to extract the frequency dependent absorption coefficient and complex refractive index, using a combination of Fresnel equations and the Beer-Lambert law, as described in detail in the Appendix.

2.2. Proposed Analysis Technique

Transfer functions utilizing physics-based estimations [e.g., (1)] have been used previously to estimate physical characteristics of materials probed using THz radiation. Such functions relay directly the effect on the input THz signal, $E_{sam}(\omega)$ of the material's complex refractive index, $\hat{n}(\omega)$, extinction coefficient, $\kappa(\omega)$ and sample thickness, l, which result in the observed output signal, $E(\omega)$ [23], i.e.,

$$\frac{E_{sam}(\omega)}{E_{ref}\omega} = \frac{4\hat{n}(\omega)n_0}{\left[(\hat{n})(\omega)+n_0\right]^2}\exp\left\{-\kappa(\omega)\frac{\omega l}{x}\right\} \times \exp\left\{-j[n(\omega)-n_0]\frac{\omega l}{c}\right\}. \quad (1)$$

However, such physical transfer functions remain difficult to solve, and often require the introduction of approximations to simplify the calculations, thereby introducing a further source of error. An alternative approach is to use a data-based rational polynomial spectral estimation. The technique of obtaining frequency domain spectra from temporal waveforms using parametric spectral estimation is well established (see, e.g., [9] and [25]). Some methods include: univariate auto-regressive (AR)-based [26], [27] and auto-regressive moving average (ARMA)-based spectral estimates, and related cross-spectral estimates based on Laplace transfer functions [9]. A univariate approach is logical where there is no knowledge of the character or timing of the input signal. In this case, however, we are interested in cross-spectrum estimation, as we wish to deconvolve the spectral response of the *sample* from the *sample and free space* cross-spectrum and a transfer function estimation is therefore more appropriate.

Transfer functions describe the effect that changes to a system input have on the output in the Laplace domain and are direct, complex rational approximations of cross-spectra [9]. When a good model fit is obtained (i.e., one which explains over 99% of the output data variance) using a transfer function with a given number of parameters, it means that any further detail obtained by increasing the number of parameters is quite likely to consist primarily of noise and artifacts. It is worth noting at this point that for the data we analyzed, simple 12^{th} and 18^{th} order transfer function models with only 24 and 36 parameters, respectively, explained 99.98% of the output variance of the measured samples. This implies that even a smoothed Fourier spectrum would be "over-parameterized", estimating, for example, an equivalent of 1024 values from the same sample of 600 points. A further advantage of using this technique is that the 20+ parameters in the transfer function, can be used directly to numerically estimate vibrational and relaxational processes within the sample (seen as dynamic modes observable in the time domain data) in response to the THz pulse, effectively enabling the reverse-engineering of the dominant dynamic modes of the sample [28].

Most of the methods in the references cited are based on *discrete-time* transfer functions. The main disadvantage of discrete-time models is that they only work well for a narrow spectral range: limited by Popov's Theorem at the high end of the spectrum, and by the numerical problems of estimating oversampled, discrete systems at the low end. The low frequencies are compressed owing to mapping between the left, complex half-plane of the continuous system's poles (eigenvalues) and the unit circle of the discrete system's poles. This causes problems with effective estimation of systems with broad spectra. The alternative *continuous-time* form of transfer functions has gained popularity in recent years owing to the development of modern estimation methods for continuous time systems [29]–[32]. Continuous-time models, as used in this work, can be used for systems with a broad spectral range, and allow precise spectral estimation at the low-frequency end of the useful spectrum.

The transfer function, which is a data-based analog of (1), is estimated directly in the following form, using the continuous Laplace operator s of order m, n (where m is the order of the numerator and n the denominator), with a time delay τ between the reference and transmitted time-domain signals:

$$Y(s) = \frac{P(s)}{Q(s)} = \frac{\beta_0 s^m + \beta_1 s^{m-1} + \cdots + \beta_m}{s^n + \alpha_1 s^{n-1} \cdots + \alpha_n} e^{-s\tau} U(s) \tag{2}$$

where $U(s) = \mathcal{L}\{u(t)\}$ and $Y(s) = \mathcal{L}\{y(t)\}$ are, assuming zero initial conditions, the Laplace transforms of the system's respective input (in this case, the reference THz signal) and output (the THz pulse transmitted through the material under test).The spectrum is then calculated by substituting $s = j2\pi\nu$, where ν is the normalized frequency.

The problem remains of effective estimation of transfer function models which will not produce spurious spectral peaks (caused for instance by estimating dynamic modes not sufficiently represented in the data), but which, at the same time, will explain the data sufficiently well. This is addressed by identifying parametrically efficient structures from a range of transfer function models fitted to the input-output data set (i.e., data-based mechanistic methodology [28], [33]).

The method proposed in this work involves the use of efficiently parameterized, continuous-time transfer functions to describe the dynamic relationship between the temporal responses of the reference and sample. The result is a cross-spectrum between the pair of signals, effectively deconvolving the reference dynamics from the overall spectrum. When using a Fourier transform, material parameters, such as the absorption coefficient and refractive index, are calculated by non-linear transformations of the amplitude and phase estimates. The uncertainty of these FFT spectrum estimates is usually high as a result of sample truncation and FFT over-parameterization effects. By using considerably fewer parameters in a transfer function, we limit the uncertainty of the spectral estimates and therefore of the subsequently extracted material parameters. Furthermore, because the transfer function estimate *includes* the uncertainty of its parameters, we are able to apply a simple Monte Carlo approach to estimate how the parametric uncertainty propagates into uncertainty in the calculated material parameters. This provides a vital tool when comparing the

characteristics of different substances or samples, by allowing rigorous statistical comparisons between different measurements.

For each signal series, the proposed procedure consists of the following:

1) Forming a pair of input-output signals by selecting a meaningful temporal range of the measured reference spectrum and cross-spectrum time-domain signals (i.e., truncating the data before the first reflection peak, as described in the supplementary information).
2) Identifying the structure of a continuous-time transfer function. This is achieved by assessing the suitability of a wide range of models, each of which has the form:

$$\widehat{Y}(s) = \frac{\widehat{P}(s)}{\widehat{Q}(s)} U(s) = \frac{\widehat{\beta_0} s^m + \widehat{\beta_1} s^{m-1} + \cdots + \widehat{\beta_m}}{s^n + \widehat{\alpha_1} s^{n-1} + \cdots + \widehat{\alpha_n}} e^{-s\tau} U(s) \tag{3}$$

with different orders m and n, and time delays, τ. This transfer function is a rational, complex data-based analog of Equation (1), and allows estimation of the effect of the material on the THz signal. The final model selection uses an Information-based Criterion (IC), similar to Akaike's (AIC) [27], [28], which allows the selection of a parametrically efficient representation of the time-domain signal, by varying m, n, and τ, while maintaining a good fit based on a high R_t^2 value:

$$R_t^2 = 1 - \frac{\sigma_e^2}{\sigma_y^2} \tag{4}$$

σ_e^2 is the variance of model residuals $e(t) = y(t) - \hat{y}(t)$, and σ_y^2 is the variance of the observed output variable y.
3) Estimating the coefficients of the selected transfer function, and calculating the associated spectrum. The transfer function is estimated using the continuous-time Simplified Refined Instrumental Variable (SRIV) method [29]. The SRIV method is related to linear, least squares algorithms, but is designed specifically for continuous-time dynamic systems, and is implemented in the CAPTAIN Toolbox for MATLAB [32]. This yields the estimated coefficients $\hat{\beta}_0 \ldots \hat{\beta}_m, \hat{\alpha}_1 \ldots \hat{\alpha}_n$ in Equation (3). The complex-conjugate pairs of roots of the transfer function denominator determine the frequencies and damping of the dominant oscillatory modes of the material. The complex spectrum estimate is then calculated simply by substituting $s = j2\pi\nu$ into the expression for $\widehat{G}(s)$ in Equation [3]. The material characteristics (refractive index and absorption coefficient) as functions of frequency ν, are then calculated from the obtained complex spectrum estimate, using the same method as for FFT-based calculations (see Supplementary Information).
4) Obtaining a measure of the uncertainty in the estimated material parameters. As this is a data-based estimation (as are FFTs), the uncertainty in the material characteristics due to sources of error in the measurement (e.g., sample thickness, reflections etc.) as discussed in detail in [24] are still embedded within the signal. However, unlike FFTs, the proposed technique allows calculation of the uncertainty associated with the spectral estimation procedure and therefore provides some confidence in the relative peak intensities of the resultant spectrum. It is important to have a measure of this type of uncertainty, as this allows inter-sample comparisons with a level of scientific rigor. The proposed method provides an uncertainty estimate for the transfer function parameters in the form of their covariance matrix. These uncertainty estimates can subsequently be used in a simple Monte Carlo calculation of randomized transfer function parameters, and hence of randomized spectra. In this approach, a large number (1000 in this example) of possible transfer function realizations are generated from a multivariable Normal distribution, where the mean arises from the SRIV transfer function estimates of α and β, obtained in step (3) above, and the covariance matrix of these parameters as obtained from the same SRIV procedure. Each of these parametric realizations is then used to calculate a realization of the physical characteristics of the sample, thus providing distributions of these characteristics at each frequency.

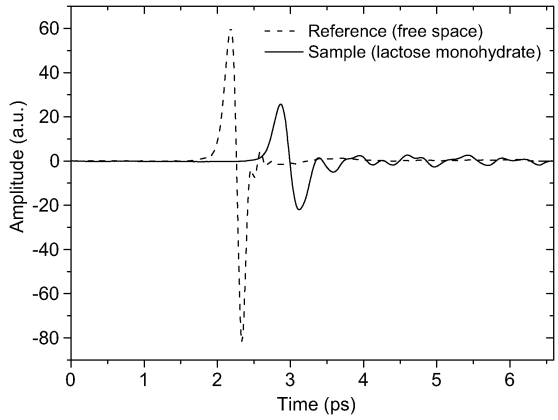


Fig. 1. Truncated THz time-domain measurements of the empty-frame-plus-free-space reference signal (dashed line) and the resultant pulse after transmission through a pressed pellet of 12.5% lactose (solid line) in a PTFE matrix, showing both the time-delay and reduction in amplitude of the transmitted THz pulse following interaction with the sample.

It should be noted that the transfer function estimate used to calculate the spectrum is obtained directly from the pair of time-domain signals (sample and reference), with no additional data (e.g., in the form of padding), or assumptions (e.g., windowing), other than the choice of truncation time [Step (1), above]. Steps (2)–(4) are easily automated in a Matlab script, an example of which can be obtained from the authors. Steps (3) and (4) take a few seconds to complete. However, initial identification of a suitable model order for each sample measured [Step (2)] is likely to take longer owing to the need to evaluate high numbers of potential models of different orders. The technique will be continually refined in future work, and the prospect of real-time data analysis in later iterations is therefore not precluded.

The technique presented here provides a reliable and well defined estimate of sample spectrum along with its uncertainty, as well as a secondary estimate of the refractive index and absorption coefficient, assuming a known sample thickness. In the case of an unknown sample thickness, a procedure for extraction of material parameters can be applied, where the thickness and refractive index are obtained using Fabry–Pérot oscillations from the frequency domain (which originate from etalon reflections within the sample) analogous to the Quasi-space [14] or Total Variation methods [34]. Unlike these techniques however, because of the lower uncertainty of the spectral estimate itself, the frequency domain material characteristics extracted using the method proposed here, will have significantly lower numerical uncertainty resulting in more accurate estimates.

3. Results

Fig. 1 shows the time-domain signals of a free-space reference (dashed line) and a pressed powder pellet containing 12.5% lactose monohydrate (solid line) in a PTFE matrix. Both time-domain signals exhibit a reflection (generated within the GaP detector) 5.4 ps after the main THz signal. To ensure that none of the reflection pulse was included in subsequent analysis, the data were truncated prior to this point (at 4.60 ps after the main pulse) resulting in 760 data points each separated by 6.7 fs.

The time-domain signals for each sample can be reconstructed using relatively low-order transfer function models; 18^{th} order for lactose [Fig. 2(a)], and 12^{th} order for glucose [Fig. 2(a)]. Both models are found to explain almost 99.99% of the output variance for the data sets analyzed. As over-parameterization can lead to (a) spurious dynamic modes and spectral peaks, and (b) poor definition (higher uncertainty) of the estimated parameters, the model orders are selected using a combination of model fitting and low-order parameterization. The fit is measured using an R_t^2 criterion [see (4)], obtained by running the input series through the transfer function model and comparing it with the observed output. The calculated R_t^2 typically lay between 0.99 and 0.999

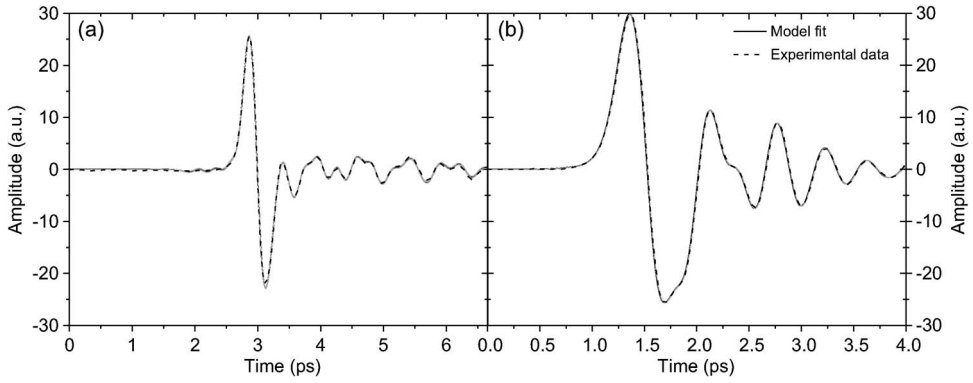


Fig. 2. (a) Lactose time-domain TF (transfer function) model fit and (b) glucose time-domain TF model fit. 95% uncertainty bands, obtained from the transfer function estimation, are shown as gray bands in both figures, though they are too narrow to resolve from the TF estimates.

(where 1.00 represents a 100% fit), indicating that all practically useful information has been encapsulated by the transfer function. This in turn implies that using spectral estimation techniques with more complicated parameterization, such as FFT based methods, will not produce a significantly better fit, and will more likely result in increased error and poorly defined results.

The material parameter uncertainty estimates, derived from the transfer function parameters and their covariance matrices as percentile based uncertainty bounds, are shown as gray bands in Fig. 2(a) and (b). The absolute error values are affected by a combination of the sample and experimental characteristics, including the subjective choice of data truncation point. It can be seen that the uncertainty for both samples is small (indeed they are hidden immediately behind the TF representation of the time-domain data) indicating a good model fit.

For comparison with FFT techniques, frequency-domain estimates were calculated by substitution of $s = j2\pi\nu$ into (3), for both the reference and sample measurements for each material. These were next analyzed as described in the supplementary information, using Fresnel equations and the Beer-Lambert law. Fig. 3(a) and (b) shows, respectively, the frequency-dependent absorption coefficient and refractive index, obtained between 0.4–4 THz for the lactose pellet. Equivalent results between 0.4–4 THz obtained from a pressed powder pellet containing 100% glucose are shown in Fig. 3(c) and (d). In each plot, FFT data are shown as dashed lines, and solid lines represent the Transfer Function estimates. A clear strength of our proposed technique is in the ability to calculate the percentage error associated with estimates of frequency-dependent material parameters, which are shown in the figures as gray bands. These errors translate directly from the time-domain signal estimates and give a clear indication of the spectral variance at each frequency. For both materials, good agreement is seen between our proposed technique and the Fourier Transforms. It should be noted, however, that the comparison shown here is between two approximations of material parameters: one obtained using a standard FFT of the time-domain data and one from data extracted using the proposed continuous-time transfer function. While similarities between the extracted parameters provides some degree of confidence in the proposed technique, perfect agreement is neither required nor expected. The frequency positions of the absorption peaks in the lactose sample are in good agreement with those published in the literature [22], [35].

4. Conclusion

We have demonstrated the use of continuous-time, Laplace domain transfer functions to reconstruct time-domain THz signals with high ($>$ 99.9%) agreement and low uncertainty. By avoiding the need for windowing or zero-padding of data, uncertainty arising from spectral variance and leakage are thus avoided. The constituent frequency parameters of the transfer functions can then be used to estimate the spectrum of a sample in a reliable and reproducible way and include

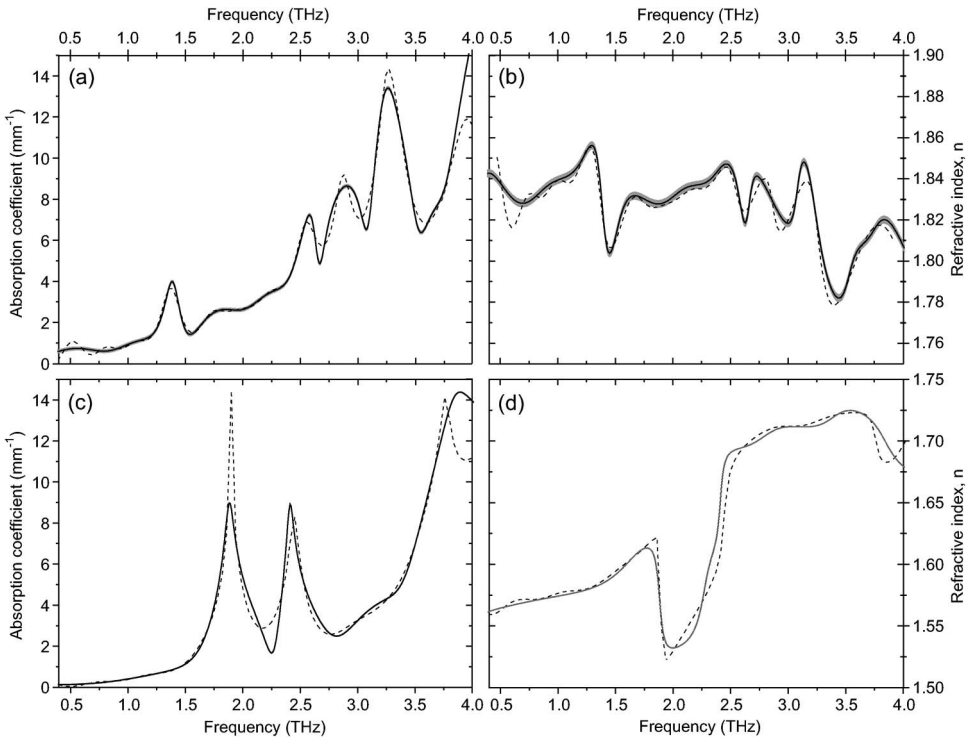


Fig. 3. Absorption coefficient and refractive index estimates showing the TF with 95% uncertainty band (solid line, uncertainty band shaded gray) and for the FFT-based approach (dotted line) for (a), (b) 12.5% lactose sample and (c), (d) the 100% glucose sample. As for the time-domain estimates, the error band associated with the glucose sample is very narrow and lies directly beneath the calculated sample characteristics in (c) and (d).

uncertainty measures of the calculated characteristics as well as offering the potential for identification of the dominant dynamic modes of a measured sample. This could potentially provide a unified technique for analyzing time-domain data from a range of terahertz spectroscopy systems, providing confidence bounds which are vital when comparing sample spectra, and avoiding the high numerical sensitivity of frequency based deconvolution.

5. Data and Software Availability

Both data and Matlab scripts used to prepare this Open Access publication can be obtained from the corresponding author. The CAPTAIN Toolbox for Matlab can be obtained from the URL below or by contacting the corresponding author (http://www.es.lancs.ac.uk/cres/captain/).

Appendix
Current Data Analysis Technique

The coherent detection of THz pulses in the time domain allows both the frequency-dependent absorption coefficient and refractive index of the sample studied to be calculated from the *attenuation* of the transmitted pulse, and the *time delay* between the reference pulse and the pulse transmitted through the sample, respectively.

In transmission THz-TDS, the THz frequency refractive index and absorption coefficient of a material are determined from the differences between a reference pulse $(E_{\text{reference}})$ and a sample pulse (E_{sample}) [2]. The reference pulse is taken in dry air, and the sample pulse defined as the signal transmitted through a sample placed at the focal spot of the THz beam (also in dry air). These

two pulses are truncated before the first reflection pulse, as described previously, then transformed into the frequency domain. The phase change between the two signals, $\phi_{\text{reference}}(\nu) - \phi_{\text{sample}}(\nu)$, where ν is frequency, is used to calculate the refractive index, n, of a sample of thickness d, directly [2]:

$$n(\nu) = 1 + \frac{\left(\phi_{\text{reference}}(\nu) - \phi_{\text{sample}}(\nu)\right)c}{2d\pi\nu}. \tag{5}$$

The refractive index of the material can then be used to calculate the Fresnel reflection coefficient R:

$$R(\nu) = \left|\frac{n_{\text{sample}}(\nu) - n_{\text{air}}}{n_{\text{sample}}(\nu) + n_{\text{air}}}\right|^2 \tag{6}$$

which can be re-written as the transmission coefficient T:

$$T(\nu) = 1 - R = \frac{4n_{\text{sample}}(\nu)n_{\text{air}}}{(n_{\text{sample}}(\nu) + n_{\text{air}})^2}. \tag{7}$$

If calculating the refractive index relative to free space, $n_{\text{air}} = 1$, this is simplified to

$$T(\nu) = 1 - R = \frac{4n_{sample}(\nu)}{(n_{sample}(\nu) + 1)^2}. \tag{8}$$

The ratio of amplitudes of the two spectra—$A(\nu)$ (sample spectrum) and $A_0(\nu)$ (reference spectrum) can be used to find the absorption coefficient, using the Beer-Lambert law

$$I(\nu) = I_0(\nu)\exp(-\alpha d) \tag{9}$$

where I(ν) is the intensity measured through the sample and $I_0(\nu)$ is the transmitted intensity measured through the reference. This can be re-arranged for α, giving

$$\alpha = -\frac{1}{d} \cdot \ln\left|\frac{I(\nu)}{I_0(\nu)}\right|. \tag{10}$$

As the intensity is proportional to the squared amplitude of the THz electric field, $A(\nu)^2$, we can write

$$\alpha = -\frac{2}{d} \cdot \ln\left|\frac{A}{A_0}\right|. \tag{11}$$

This calculation does not take into account any Fresnel reflections. In order to take these into account, we need to include the transmission coefficient

$$\alpha(\nu) = -\frac{2}{d}\ln\left\{\frac{A(\nu)}{A_0(\nu)T(\nu)}\right\}. \tag{12}$$

Substituting (8) into this gives

$$\alpha(\nu) = -\frac{2}{d}\ln\left\{\frac{A(\nu)}{A_0(\nu)}\frac{[n(\nu)+1]^2}{4n(\nu)}\right\}. \tag{13}$$

Owing to the shape of the generated THz pulse in the time domain, the power of the lower frequency components of the pulse is greater than that of the higher frequencies. At higher frequencies, the power diminishes until it reaches the noise floor of the experiment. The noise floor can be approximated by the spectrum recorded when the THz beam path is blocked [36], where the noise is a sum of contributions from all of the electronic components of the system. The bandwidth is defined as the free-space or reference signal (if an aperture is used) before the noise floor. The dynamic range (DR), $D(\nu)$, of the system can therefore be defined as the reference spectrum

normalized to this noise floor $A_0(\nu)$. As a result of this, the limit of the largest value of absorption coefficient which can be accurately measured can be derived from (13) [6] to be

$$\alpha_{\max} d = 2\ln\left[D(\nu)\frac{4n(\nu)}{(n(\nu)+1)^2}\right]. \tag{14}$$

Acknowledgment

The authors wish to thank Lancaster University for providing funding for this Open Access publication. The data were collected using the broad bandwidth THz-TDS system at the School of Electronic and Electrical Engineering at the University of Leeds, U.K. The authors wish to thank Dr. A. D. Burnett for help with data collection and Professors A. G. Davies, E. H. Linfield, and J. E. Cunningham for providing access to the laboratory where the THz data were obtained. The authors also wish to express their thanks to the anonymous reviewers who contributed their time and provided extremely useful critical comments, which served to improve the paper.

References

[1] C. J. Strachan, P. F. Taday, D. Newnham, K. Gordon, J. Zeitler, M. Pepper, and T. Rades, "Using terahertz pulsed spectroscopy to quantify pharmaceutical polymorphism and crystallinity," *J. Pharm. Sci.*, vol. 94, no. 4, pp. 837–846, Apr. 2005.
[2] W. H. Fan, A. Burnett, P. C. Upadhya, J. Cunningham, E. H. Linfield, and A. G. Davies, "Far-infrared spectroscopic characterization of explosives for security applications using broadband terahertz time-domain spectroscopy," *Appl. Spectrosc.*, vol. 61, no. 6, pp. 638–643, Jun. 2007.
[3] Y. C. Shen, T. Lo, P. Taday, B. Cole, W. Tribe, and M. Kemp, "Detection and identification of explosives using terahertz pulsed spectroscopic imaging," *Appl. Phys. Lett.*, vol. 86, no. 24, pp. 241116-1–241116-3, Jun. 2005.
[4] B. Fischer, M. Hoffmann, H. Helm, G. Modjesch, and P. Uhd Jepsen, "Chemical recognition in terahertz time-domain spectroscopy and imaging," *Semicond. Sci. Technol.*, vol. 20, no. 7, pp. S246–S253, Jul. 2005.
[5] K. Tych, A. Burnett, C. Wood, J. Cunningham, A. Pearson, A. Davies, and E. Linfield, "Applying broadband terahertz time-domain spectroscopy to the analysis of crystalline proteins: A dehydration study," *J. Appl. Crystallogr.*, vol. 44, no. 1, pp. 129–133, Feb. 2011.
[6] A. Markelz, A. Roitberg, and E. Heilweil, "Pulsed terahertz spectroscopy of DNA, bovine serum albumin and collagen between 0.1 and 2.0 THz," *Chem. Phys. Lett.*, vol. 320, no. 1/2, pp. 42–48, Mar. 2000.
[7] 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. Amer. B, Opt. Phys.*, vol. 7, no. 10, pp. 2006–2015, Oct. 1990.
[8] T.-I. Jeon and D. Grischkowsky, "Characterization of optically dense, doped semiconductors by reflection THz time-domain spectroscopy," *Appl. Phys. Lett.*, vol. 72, no. 23, pp. 3032–3034, Jun. 1998.
[9] P. Stoica and R. Moses, *Introduction to Spectral Analysis,* 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1997.
[10] J. M. Spyers, P. G. Bain, and S. J. Roberts, "A comparison of fast Fourier transform (FFT) and autoregressive (AR) spectral estimation techniques for the analysis of tremor data," *J. Neurosci. Methods*, vol. 83, no. 1, pp. 35–43, Aug. 1998.
[11] K. Aki and P. Richards, *Quantitative Seismology*. San Francisco, CA, USA: Freeman, 1980.
[12] L. Duvillaret, F. Garet, and J.-L. Coutaz, "Influence of noise on the characterization of materials by terahertz time-domain spectroscopy," *J. Opt. Soc. Amer. B, Opt. Phys.*, vol. 17, no. 3, pp. 452–461, Mar. 2000.
[13] T. D. Dorney, R. G. Baraniuk, and D. M. Mittleman, "Material parameter estimation with terahertz time-domain spectroscopy," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 18, no. 7, pp. 1562–1571, Jul. 2001.
[14] I. Pupeza, R. Wilk, and M. Koch, "Highly accurate optical material parameter determination with THz time-domain spectroscopy," *Opt. Exp.*, vol. 15, no. 7, pp. 4335–4350, Apr. 2007.
[15] E. R. Brown, J. E. Bjarnason, A. M. Fedor, and T. M. Korter, "On the strong and narrow absorption signature in lactose at 0.53 THz," *Appl. Phys. Lett.*, vol. 90, no. 6, pp. 061908-1–061908-3, Feb. 2007.
[16] M. Walther, M. Freeman, and F. A. Hegmann, "Metal-wire terahertz time-domain spectroscopy," *Appl. Phys. Lett.*, vol. 87, no. 26, pp. 261107-1–261107-3, Dec. 2005.
[17] D. Allis, A. Fedor, T. Korter, J. Bjarnason, and E. Brown, "Assignment of the lowest-lying THz absorption signatures in biotin and lactose monohydrate by solid-state density functional theory," *Chem. Phys. Lett.*, vol. 440, no. 4–6, pp. 203–209, Jun. 2007.
[18] H. Tuononen, E. Gornov, J. A. Zeitler, J. Aaltonen, and K.-E. Peiponen, "Using modified Kramers–Kronig relations to test transmission spectra of porous media in THz-TDS," *Opt. Lett.*, vol. 35, no. 5, pp. 631–633, Mar. 2010.
[19] X. Li, Z. Hong, J. He, and Y. Chen, "Precisely optical material parameter determination by time domain waveform rebuilding with THz time-domain spectroscopy," *Opt. Commun.*, vol. 283, no. 23, pp. 4701–4706, Dec. 2010.
[20] Y. Ueno and K. Ajito, "Analytical terahertz spectroscopy," *Anal. Sci.*, vol. 24, no. 2, pp. 185–192, Feb. 2008.
[21] M. B. Byrne, J. Cunningham, K. Tych, A. Burnett, M. Stringer, C. Wood, L. Dazhang, M. Lachab, E. Linfield, and A. Davies, "Terahertz vibrational absorption spectroscopy using microstrip-line waveguides," *Appl. Phys. Lett.*, vol. 93, no. 18, pp. 182904-1–182904-3, Nov. 2008.

[22] J. F. Federici, B. Schulkin, F. Huang, D. Gary, R. Barat, F. Oliveira, and D. Zimdars, "THz imaging and sensing for security applications—Explosives, weapons and drugs," *Semicond. Sci. Technol.*, vol. 20, no. 7, pp. S266–S280, Jul. 2005.
[23] W. Withayachumnankul, B. M. Fischer, H. Lin, and D. Abbott, "Uncertainty in terahertz time-domain spectroscopy measurement," *J. Opt. Soc. Amer. B, Opt. Phys.*, vol. 25, no. 6, pp. 1059–1072, Jun. 2008.
[24] W. Chan, J. Deibel, and D. Mittleman, "Imaging with terahertz radiation," *Rep. Progr. Phys.*, vol. 70, no. 8, pp. 1325–1333, Aug. 2007.
[25] S. Kay and S. J. Marple, "Spectrum analysis—A modern perspective," *Proc. IEEE*, vol. 69, no. 11, pp. 1380–1419, Nov. 1981.
[26] J. Burg, "Maximum entropy spectral analysis," in *Modern Spectrum Analysis*, D. Childers, Ed. Piscataway, NJ, USA: IEEE Press, 1978, pp. 34–41.
[27] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.
[28] P. Young, "Data-based mechanistic modelling, generalised sensitivity and dominant mode analysis," *Comput. Phys. Commun.*, vol. 117, no. 1/2, pp. 113–129, Mar. 1999.
[29] P. Young and H. Garnier, "Identification and estimation of continuous-time, data-based mechanistic (DBM) models for environmental systems," *Environ. Model. Softw.*, vol. 21, no. 8, pp. 1055–1072, Aug. 2006.
[30] P. Young, H. Garnier, and M. Gilson, "Refined instrumental variable identification of continuous-time hybrid Box–Jenkins models," in *Identification of Continuous-Time Models From Sampled Data*, H. Garnier and L. Wang, Eds. London, U.K.: Springer-Verlag, 2008, pp. 91–131.
[31] H. Garnier, M. Gilson, T. Bastogne, and M. Mensler, "The CONTSID toolbox: A software support for data-based continuous-time modelling," in *Identification of Continuous-Time Models From Sampled Data*, H. Garnier and L. Wang, Eds. London, U.K.: Springer-Verlag, 2008, pp. 249–290.
[32] C. J. Taylor, D. J. Pedregal, P. C. Young, and W. Tych, "Environmental time series analysis and forecasting with the captain toolbox," *Environ. Model. Softw.*, vol. 22, no. 6, pp. 797–814, Jun. 2007.
[33] P. Young, S. Parkinson, and M. Lees, "Simplicity out of complexity in environmental modelling: Occam's razor revisited," *J. Appl. Stat.*, vol. 23, no. 2/3, pp. 165–210, Jun. 1996.
[34] M. Scheller, C. Jansen, and M. Koch, "Analyzing sub-100-μm samples with transmission terahertz time domain spectroscopy," *Opt. Commun.*, vol. 282, no. 7, pp. 1304–1306, Apr. 2009.
[35] W. R. Tribe, D. A. Newnham, P. F. Taday, and M. C. Kemp, "Hidden object detection: Security applications of terahertz technology," in *Proc. SPIE*, 2004, vol. 5354, pp. 168–176.
[36] P. Uhd Jepsen and B. Fischer, "Dynamic range in terahertz time-domain transmission and reflection spectroscopy," *Opt. Lett.*, vol. 30, no. 1, pp. 29–31, 2005.