Research Article Vol. 31, No. 26 / 18 Dec 2023 / Optics Express 44575
Optics EXPRESS

Deep neural network ensembles for THz-TDS refractive index extraction exhibiting resilience to experimental and analytical errors

NICHOLAS KLOKKOU,[1,*] JON GORECKI,[2] BEN BEDDOES,[1] AND VASILIS APOSTOLOPOULOS[1]

[1]School of Physics and Astronomy, University of Southampton, UK
[2]Department of Bioengineering, Imperial College London, UK
[*]n.t.klokkou@soton.ac.uk

Abstract: Terahertz time-domain spectroscopy (THz-TDS) achieves excellent signal-to-noise ratios by measuring the amplitude of the electric field in the time-domain, resulting in the full, complex, frequency-domain information of materials' optical parameters, such as the refractive index. However the data extraction process is non-trivial and standardization of practices are still yet to be cemented in the field leading to significant variation in sample measurements. One such contribution is low frequency noise offsetting the phase reconstruction of the Fourier transformed signal. Additionally, experimental errors such as fluctuations in the power of the laser driving the spectrometer (laser drift) can heavily contribute to erroneous measurements if not accounted for. We show that ensembles of deep neural networks trained with synthetic data extract the frequency-dependent complex refractive index, whereby required fitting steps are automated and show resilience to phase unwrapping variations and laser drift. We show that training with synthetic data allows for flexibility in the functionality of networks yet the produced ensemble supersedes current extraction techniques.

1. Introduction

The excellent signal-to-noise ratio (SNR) inherent to terahertz time-domain spectroscopy (THz-TDS) coupled with the rich information present within the acquired time-domain data of picosecond pulses has been exploited by researchers for material parameter characterization [1], biomolecule analysis [2], imaging for security applications [3], probing conductivity transients on the picosecond timescale [4] and more [5]. The development of TDS systems is still an active area of research [6–8] but commercial spectrometers are now available and have greatly assisted in making the terahertz gap accessible. However, nuance and care is still required when analyzing the data since small changes in initial data processing can significantly affect the final results.

In THz-TDS, the synchronous emission and detection processes driven by ultrafast lasers suppresses background noise whilst also retaining the phase information of the frequency components which aids the extraction of complex material parameters, such as refractive index and conductivity. Iterative methods are typically used to fit the unknown parameters to a theoretical model of light propagation [9], however significant variation in the extracted parameters can arise during the parameter extraction process due to both the multiple processing steps as well as the lack of standard practices for experimental setups [10]. Furthermore, materials which exhibit resonances at low frequencies where the SNR is also low (e.g., lithium niobate) can lead to particularly pronounced inconsistencies in extracted refractive index as they do not offer an easy way to set the phase at low frequencies. Algorithms exploiting internal reflections retained by long delay lines can improve the accuracy of THz measurements but this is not always possible if

#507439 https://doi.org/10.1364/OE.507439
Journal © 2023 Received 3 Oct 2023; revised 17 Nov 2023; accepted 1 Dec 2023; published 14 Dec 2023

Optics EXPRESS

the sample thickness or substrate is not sufficient to separate pulses in the time-domain [11,12]. Additionally, these algorithms are typically more complicated and can fail to converge.

Temporal analysis of dehydration in sunflower leaves has been shown to be possible with reverse engineering techniques [13]. The minimization of a merit function was used to ascertain the complex refractive index of multiple layers of the leaf with varying thicknesses as it degrades, demonstrating how a minimization algorithm is able to extract a system with multiple variables. It is typical for many machine learning algorithms to have such a minimization step, and its application to terahertz spectroscopy and imaging has shown much interest in recent years [14]. Benefits include bio-molecule classification (a challenge for broad spectral signatures in terahertz spectroscopy) through use of convolutional neural networks (CNNs) [15], using single layer neural networks (perceptrons) for material parameter extraction [16], combining deep neural networks (DNNs) with THz-TDS to quantify rice freshness [17], as well as exploiting the computational speed of trained DNNs to decrease acquisition time of THz single-pixel imaging systems for image reconstruction [18].

Directly extracting spectroscopic information from a terahertz time-domain signal with neural networks has been shown to be possible in the case of real conductivity of thin-films, with models trained using augmented, real-world data [19]. Alternatively, training DNNs with simulated data as opposed to large, real-world data-sets has been shown to be effective in optical techniques that require complex fitting, such as for ellipsometry [20] and double electron-electron resonance spectroscopy [21]. With respect to THz-TDS, training neural networks with simulated data provides an additional level of control. Systematic errors such as noise, phase offset, and laser power drift can be inserted into the data sets manually with controllable magnitudes in an effort to properly characterize and understand the limits of the trained networks, with the aim to mitigate the black-box problem inherent to the training of neural networks. With this in mind we present a model that accounts for errors and inconsistencies in the THz-TDS extraction processes.

Our previous work has shown that the fitting of the theoretical transfer function can be achieved by a single hidden layer perceptron trained with simulated data [16]. In this case, the network accepted four inputs: a value of frequency, thickness, phase and amplitude of the transfer function obtained by the division of a sample and reference terahertz spectra at the given value of frequency. The network then outputs a prediction of the complex refractive index at the same inputted frequency. The trained network improved on the analytical solving of the transfer function with the required approximations, while still being more straightforward to implement and faster to run than an iterative root-finding method. Similarly, networks for parameter extraction have been extended to higher frequencies [22], but there has yet to be progress on creating networks that can compensate for error and variability between spectroscopists.

The work we present here aims to explore the use case of a deeper, multilayered architecture to predict the complex refractive index from the transfer function of THz-TDS transmission measurements. We show that the network can account and correct for preprocessing and experimental errors that are commonplace with THz-TDS in part by solving for a range of frequencies simultaneously, rather than extracting each frequency component independently. Furthermore, using large, simulated data sets provides the facility to train for multiple categories of errors which helps to increase the robustness and adaptability of the trained network. We show that by training DNNs with synthetic data with simulated noise, phase offset errors, and even laser drift, the trained network can compensate for imperfect data and supersede current extraction methods.

2. Complications with THz-TDS parameter extraction

A full description of the methods used to simulate terahertz time-domain pulses is outlined in the our previous work [16]. An example of a generated time-domain pulse is given in Fig. 1 whereby an analytical model [23] describing a broadband pulse generated from a photoconductive antenna

Optics EXPRESS

is propagated through a 500 μm thick slab sample with a non-dispersive complex refractive index of $\tilde{n}(\omega) = 3 + 0.05i$. Included in this figure are two examples of how errors can occur and propagate through the extraction process leading to an incorrect refractive index extraction. Specifically: (a) the laser power drifting between the reference and sample measurements, thereby modulating the terahertz time-domain pulses and (b) an erroneous phase fitting step where an offset is introduced, usually due to low SNR.

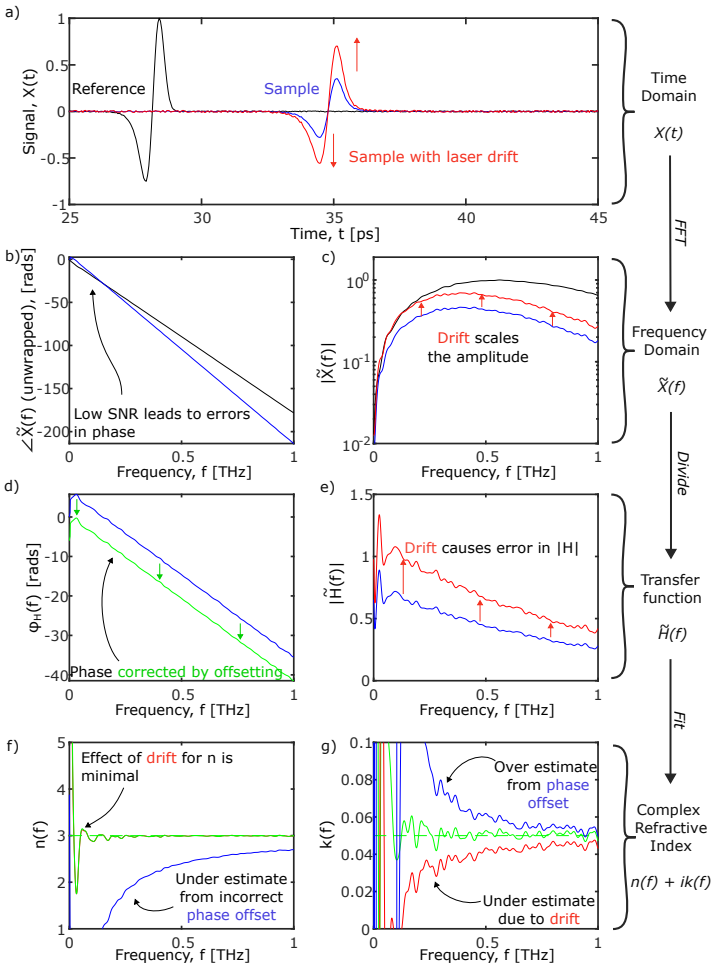

Fig. 1. Parameter extraction schematic with simulated data showing how laser drift and poor SNR at low frequencies results in erroneous extraction. In (a) the time-domain reference (black) and sample measurement (blue) is shown, where the effect of laser drift (red) after the reference can be seen to scale the sample pulse. A Fast Fourier Transform (FFT) is performed and the phase is shown in (b) where the effect of noise is causing an erroneous offset. In (c) the magnitude of the FFT is showing the reference and sample pulses, with drift causing a scaling off the amplitude. Division of the two spectra results in the complex transfer function, the phase of which shown in (d) is corrected (green) with an offset. Similarly, in (e) the laser drift causes a scaling of the magnitude of the transfer function. The extracted real refractive index (f) shows how the uncorrected phase produces a large error, but the effect of drift is minimal. The imaginary refractive index shown in (g) is affected both by the phase error and drift.

Optics EXPRESS

Extracting the complex refractive index from THz-TDS data is outlined in Fig. 1. In Fig. 1(a) the measured time-domain trace of the reference (air) and sample (slab of refractive index n = 3 + 0.05i) measurements are shown with black and blue lines respectively, with the effect of laser power drift over time scaling the pulse producing the red line. A Fast Fourier Transform (FFT) is performed on each set of time-domain data, which produces a complex spectrum with phase and magnitude shown in Fig. 1(b) and (c) respectively. The phases of the reference and sample measurements require unwrapping and are shown in Fig. 1(b). In Fig. 1(b) the drifted pulse's phase is not shown as it is identical to the non-erroneous case due to the modelling of the laser drift being a linear scaling of the amplitude only. In Fig. 1(c) the magnitude of the FFT shows the effect of laser drift across all frequencies and, in all cases, the signal falls off at low frequencies. The effect of this reduction in signal, and subsequently SNR, is what leads to the error in the phase reconstruction (an offset).

To fit the complex refractive index, first the experimental transfer function, $H(f)$, is calculated by dividing the sample spectrum by the reference spectrum. The phase of this resulting transfer function, $\Phi_H(f)$, is shown in Fig. 1(d), with the blue line showing an incorrect offset due to the low frequency noise. The phase must be corrected with an offset, resulting in a the green line, where such a correction requires prior assumptions of the dispersion of the sample.

Figure 1(e) shows the magnitude of the transfer function, $|H(f)|$, where the laser drift results in values that are greater than unity, which would incorrectly indicate the sample has gain. Figure 1(f) and (g) show how the phase error and laser drift effect the extracted real and imaginary refractive index respectively when a theoretical transfer function is fit to the data.

Extracting the parameters of the theoretical transfer function requires the fitting of two unknowns and cannot be solved analytically. The method used here is the Newton-Raphson method, an iterative root finding method whereby the correct refractive index is found when the difference between the experimental transfer function and theoretical transfer function is zero. The theoretical transfer function is given by modelling the transmission, reflection, and propagation of radiation through a sample of defined thickness, d, using Fresnel coefficients [11]. The Newton-Raphson method provides fast convergence as its iterations are informed by the gradient of the function it is trying to solve, but it cannot do so globally and is performed on a per frequency component basis.

The impact of the phase offset error on the real refractive index can be seen in Fig. 1(f) by the blue line, where there is a large error, particularly at lower frequencies. The laser drift, however, affects the refractive index extraction minimally here, due to the phase of the transfer function being primarily a consequence of the propagation of light through the medium rather than effects at the interfaces.

In the case of the imaginary part of the refractive index shown in Fig. 1(g), both the phase error and laser drift have a significant impact where the parameter is greatly underestimated. Both effects of the erroneous data are largest at lower frequency but converge to the correct solution as the frequency increases, although an error is still present.

3. Methods

3.1. Terahertz time-domain spectroscopy

In Fig. 2 we show a simple schematic of the THz-TDS used for the acquisition of experimental data in a transmission configuration. The spectrometer is typical but unlike most commercial systems the driving ultrafast laser is free space coupled to the photoconductive antennas (PCA). A reference measurement is made (air) and then the sample is measured by placing it in the focus of the lens system between the emitter and detector. The scan length is 50 ps and we average 120 scans in 120 seconds. The measured time-domain traces are corrected for any DC offset that can appear due to electrical noise and further windowed to remove reflections and retain just the

Optics EXPRESS

primary transmitted pulse. The subsequent processes of taking an FFT are the same as those discussed in section 2.

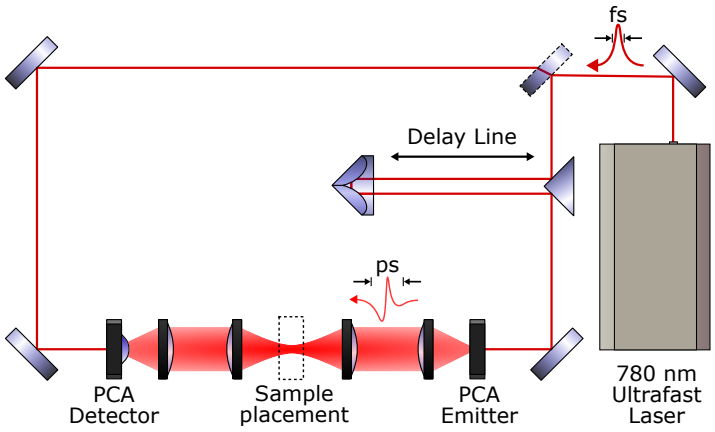

Fig. 2. A schematic of the THz-TDS system used for the experimental measurements. The photoconductive antenna (PCA) emitter and detector are free space coupled to the titanium:sapphire femtosecond laser driving it.

3.2. *Neural network ensembles for THz-TDS parameter extraction*

A schematic of the deep neural network architecture used in this work is shown in Fig. 3(a), where the networks comprise of five hidden layers with 64 fully connected neurons. Multiple neural networks with identical architectures are trained to form an *ensemble*, where the predictions of each network in the ensemble are averaged to increase accuracy and generality [24–26]. Each network in the ensemble is trained with identical data sets, but due to the random nature of the starting weights and biases in the network, as well as the random distribution of the data sets between training, testing and validation, each resultant network behaves slightly differently. Example predictions of the real refractive index of a simulated sample with a non-dispersive refractive index of n = 7 from each individual network as well as the averaged ensemble prediction is shown Fig. 3(b). It is clear that the averaged result is more stable. Figure 3(c) shows how the root mean squared error (RMSE) between the true value and that of an ensemble improves with ensemble size, following an expected inversely proportional trend. Increasing the number of neural networks in the ensemble reduces the error but with diminishing returns. Here we show the effect on the real refractive index, which is the first half of the output array. Predictions of the extinction coefficient are similar as they start with a percentage RMSE of 0.055 for one network, decaying to 0.032 for ten networks. Going forward we continue to train ensembles of 10 networks, but it should be noted that a significant improvement on single network results can be achieved with 3-5 networks.

In order to maintain generality of the model the propagation of terahertz pulses are simulated through samples with randomly generated dispersion curves. The material parameters are not restricted by the Kramers-Kronig relations, simplifying the data generation process, with efforts instead focusing on modelling the characteristics of physical experiments that can be responsible for the variability between THz-TDS extractions. The random, but still continuous, frequency-dependent refractive indices of the samples (all with a thickness of 500 μm) are achieved by summing a varying number of sine waves with random frequencies and phase offsets.

Optics EXPRESS

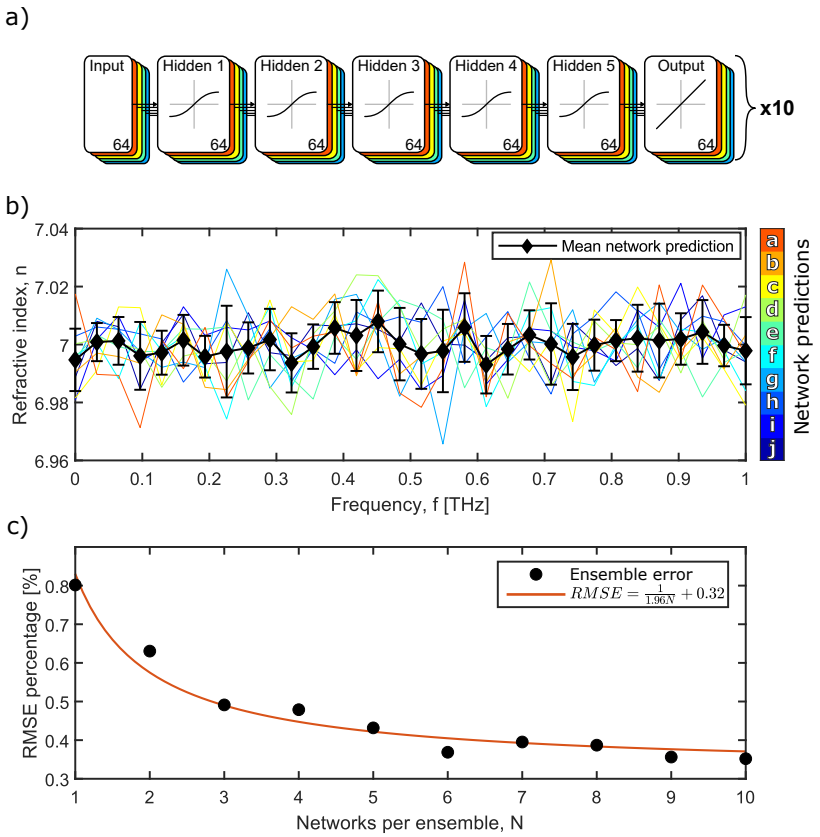

Fig. 3. Averaging neural network ensembles increases prediction accuracy and stability. The networks used have five hidden layers with 64 fully connected neurons and 10 of these are trained to produce an ensemble (a). Example extractions of a simulated material n = 7 (b) show the variance of each model and their mean and standard error (black diamonds). The performance (root mean squared error) of network ensembles with increasing number (c) broadly follows a 1/N trend, where N is the number of networks in the ensemble.

3.3. *Simulating noisy data*

The errors selected to be introduced into the training data to produce a more consistent extraction algorithm are white noise, phase offsetting and linear magnitude scaling of the simulated transfer function, all with randomly selected magnitudes. Applying just random noise to the training data has been shown to improve generality of networks [27]. The noise profile of a typical spectrometer varies with frequency, however in the case of improving the training of a neural network, a flat, white noise profile is sufficient as this acts to reduce the likelihood of the training algorithm to converge on a local minimum and halt training early to avoid over-fitting. In this work, the magnitude of the noise is defined as a signal-to-noise ratio between the intensity of the peak signal and that of the applied white noise s(t),

$$s^2(t) = R_s \cdot 10^{\frac{-SNR}{10}} \tag{1}$$

where a random number $R_s \sim U(-1, 1)$ (uniform distribution between -1 and 1) is selected for each data point in time.

Optics EXPRESS

3.4. Simulating phase errors due to low frequency noise

In the case of THz-TDS, algorithms are used to fit the phase information of the spectral data to reduce error in extracted parameters (predominantly n) [28]. The error is caused by the low frequency data having a diminished SNR as frequency decreases and results in an offset of the phase of the transfer function. This can only be corrected properly if one knows the expected dispersion curve of the material. Figure 1(d) shows the phase before and after correction and the subsequent effect on the extracted real refractive index is shown in Fig. 1(f) and the effect on the imaginary refractive index is shown in Fig. 1(g). The effect of the erroneous phase is not constant as a function of frequency and can mask the presence of low frequency resonances or introduce artificial oscillations. Simulating this error is achieved by simply applying a random offset to the simulated phases of the transfer function. In order to normalize the phase training data, an estimate for the maximum possible phase delay in the training set is calculated as,

$$\Phi_{f_{max}}^{n_{max}} = 2\pi \cdot f_{max} \cdot n_{max} \cdot \frac{d}{c}, \tag{2}$$

where f_{max} and n_{max} are the upper bounds of frequency and real refractive index defined in the training data, d is the sample thickness and c the speed of light. The value calculated here acts as a normalization factor and allows the random phase offset applied to the simulated data to be limited by the fractional maximum, ϕ_{max}. Therefore, the resulting phase offset, ϕ_0, is given as

$$\phi_0 = R_\phi \cdot \phi_{max} \cdot \Phi_{f_{max}}^{n_{max}}, \tag{3}$$

where the random number $R_\phi \sim U(-1, 1)$ is generated for each simulated experiment.

3.5. Simulating laser drift

Typically, a reference measurement is made before the sample is introduced, and during this time period the experimental environment can change. A common effect is the laser power changing between measurements. It is assumed that modest laser drift or small changes in alignment will affect all frequency components equally and so modelling this effect can be achieved by simply scaling magnitude of the terahertz pulses. Figure 1(c) shows clearly the scaling of the amplitude of spectrum and then the resultant effect on the extraction of the materials extinction coefficient in (d). Due to a small but measurable contribution the imaginary part of the refractive index has on the pulse's reflection and transmission through the material, the effect of the laser drift does not affect all frequencies equally on the final result. To motivate our networks to correct for laser drift, the fluctuations in power are simulated by scaling the transfer function, whereby fractional scalar is generated δ_{max} resulting in a scaling value,

$$\delta = R_\delta \cdot \delta_{max}, \tag{4}$$

where the random number $R_\delta \sim U(-1, 1)$ is generated for each simulated measurements. The transfer function should encapsulate the linear scaling present in laser drift leading to the ability of a network to distinguish between the impact of a change in real absorption of the sample: an exponential decay as a function of frequency.

The combination of the previously described effects result in the following transfer function,

$$\tilde{H}_{simulated}(f) = \frac{s_1^2(f) + (1 + \delta_1) \cdot \tilde{t_{01}}(f) \cdot \tilde{p_1}(f, d) \cdot \tilde{t_{10}}(f)}{s_0^2(f) + (1 + \delta_0) \cdot \tilde{p_0}(f, d)} \cdot \exp(i\phi_0) \tag{5}$$

where subscripts 0 and 1 correspond to the simulated reference and sample measurement respectively. $\tilde{p}(f)$ and $\tilde{t}(f)$ are the propagation and transmission Fresnel coefficients. The magnitude of the Fourier transform of white noise has the same power spectral density profile, therefore the expression given in Eq. (3.5) holds true in both the frequency and time-domain for the purposes of this experiment.

Optics EXPRESS

3.6. Training methods

The neural network ensembles were trained with the Matlab Deep Learning Toolbox. The simulated input spectra and output parameters were restricted to a 32 point resolution in frequency, resulting in an input and output layers of 64 neurons. The input parameters are the log of the magnitude and the unwrapped phase of the transfer function. The log of the magnitude is used as to not bias the values towards lower frequencies where attenuation is lowest. Both are normalized by taking the theoretical maximum and minimums and scaling between 0 and 1, although the introduction of random noise will create data that is outside these bounds. The architecture comprises of 5 hidden, fully connected layers of 64 neurons and up to 10 networks were trained with simulated data with a variation of parameters. The predictions of the resulting networks are averaged to produce an *ensemble prediction*. A *tanh* activation function was used for each hidden layer, as opposed to a ReLU (rectified linear unit) function for example, as this has resulted in the best performing networks so far, possibly due to the handling of negative numbers when noisy data is introduced. As this is a regression problem, the output layer uses a linear activation function although in principle a ReLU function could be used here. Each network was trained using the *fitnet* function with the builtin *scaled conjugate gradient decent* back-propagation algorithm and a mean squared error (MSE) loss function. 70 % of the simulated data is used for training, with 15 % used for testing and 15 % for validation. The stopping criteria was 6 sequential failures of a validation check where the networks performance regressed with the validation data to prevent over fitting. Using a stopping criterion of 6 failed validation steps gave us better results than setting a fixed number of epochs. For each set of parameters 200,000 materials were simulated within the range of $2 < n < 8$ and $0 < k < 0.3$. Thickness in this study was fixed at 500 μm, with only the training, testing and validation subsets differing due to random selection.

Networks of this size, trained with the aforementioned stopping criterion and data set size took no more than two hours to train on an Nvidia RTX 3090. The time, and therefore number of epochs, depends on the training data set and how quickly the stopping criteria is reached. Furthermore, for a single network in an ensemble, that trains on the same but randomly shuffled data set the variation in time and epochs before halting the training for each varies by ± 30 %. Similarly, the resulting losses had a large variation, although not as much as the training time, by 21.5 % for a given ensemble. The runtime of a single trained network is approximately 32 ms on the CPU.

4. Results

4.1. Network ensembles trained with white noise

The behaviour of the trained networks is consistent for a variety of samples but here we highlight measurements of lithium niobate, due to the difficulty in extracting a consistent refractive index of this material because of the existences of resonances in the low frequency range. The importance of adding noise to the data for correct extraction is demonstrated in Fig. 4, where the difference in extraction between the networks trained with and without noise is shown, which include noise levels resulting in 60, 70 and 80 dB peak SNR. It is clear that the addition of noise is essential for training stable networks to predict experimental data. As can be seen with the experimental extractions in Fig. 4(b) and (d) of a sample of lithium niobate, noiseless models fail to provide a reasonable extraction. The addition of noise improves the extraction, but the magnitude should be carefully chosen as the networks trained with 70 dB peak SNR show the most accurate fitting as compared to 60 or 80 dB SNR. Although introducing noise can assist in training, too much noise will reduce the proportion of usable signal in the data sets being trained on. Furthermore, it is not clear if this choice of SNR is also a function of the noise profile of the specific spectrometer

Optics EXPRESS

used for the acquisition of this data. This result demonstrates that using only simulated data without noise to quantify the predictive capability of trained networks is insufficient.

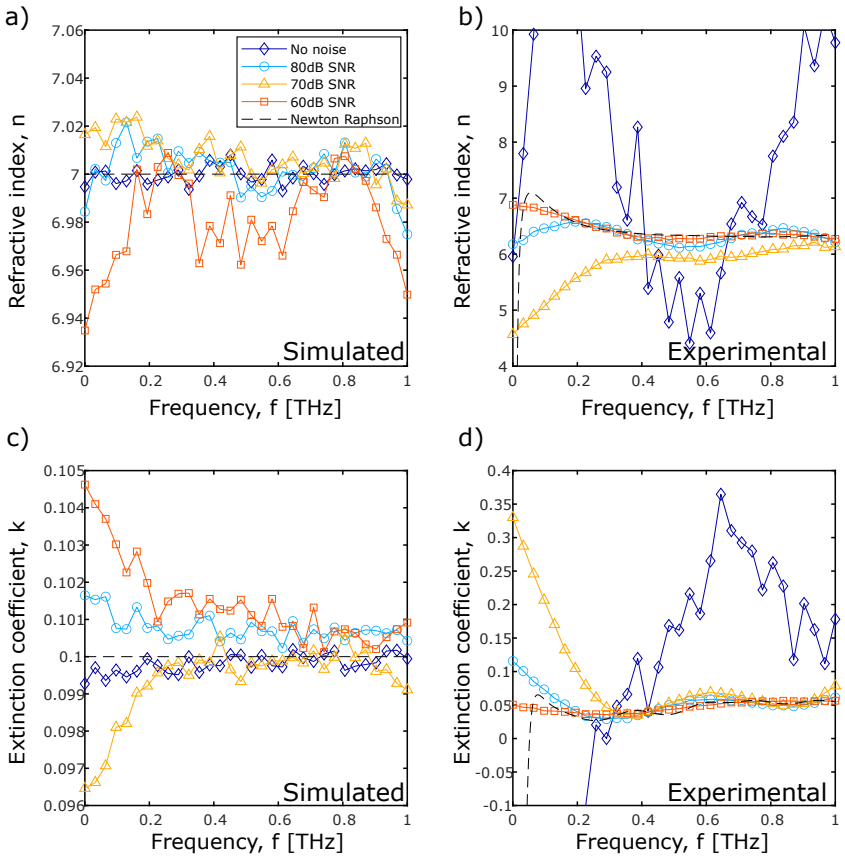

Fig. 4. Prediction of the real refractive index of a simulated material of n = 7 + 0.1i (a) and experimental data of a sample of lithium niobate (b) with the same network ensembles trained with data containing a range of noise. Additionally the respective imaginary index predictions are shown (c) and (d). Characterizing the networks with simulated data alone is not sufficient to validate the network's real world performance as training with noisy data gives rise to better experimental data fitting at the cost of simulated data accuracy.

4.2. Network ensembles trained with phase offset errors

The addition of white noise in the time-domain will induce a small amount of noise to the phase of the Fourier transformed signal. However the further addition of a forced, random phase offset is needed to further increase the accuracy of extracted parameters. Figure 5 shows the extracted complex refractive index of the lithium niobate sample, where a range of artificial phase offsets are introduced into the experimental data as much as -0.5π. Shown in the figure are extractions using a Newton-Raphson fit (dashed), networks trained with white noise (triangles) and networks trained with white noise and phase noise (diamonds). The networks trained with phase noise were trained with data that is offset with a random percentage between ± 10 % of the phase value of the transfer functions at 1 THz.

The phase error propagates into the extracted parameters as expected with the Newton-Raphson fit, however networks trained on noisy data produce a less varying result as the phase offsets

Optics EXPRESS

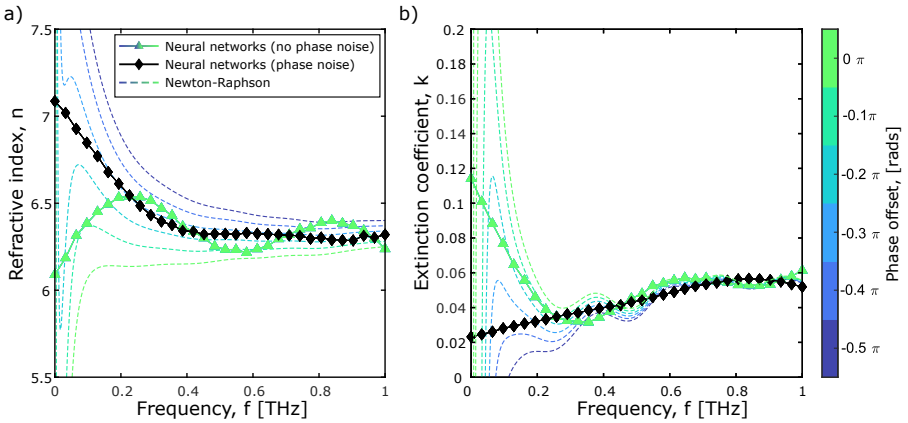

Fig. 5. Real (a) and imaginary (b) refractive index extracted from an experimentally measured sample of lithium niobate with a range of forced phase offset errors. Neural network ensembles trained with data containing white noise (triangles) and data also containing phase noise (diamonds) show more consistent, overlapping extractions regardless of the phase offset applied. The direct Newton-Raphson fittings (dashed lines) are erroneous proportionally to the forced phase offset.

change. The white noise trained networks show oscillations in frequency in the real and imaginary refractive index that are clearly erroneous. Introducing forced offsets into the training data produces networks that both have the characteristic dispersion curves extracted with the Newton-Raphson method, while also being resilient phase offsetting.

4.3. Network ensembles trained with laser drift

An unstable laser or a change in experimental conditions will interfere with a THz-TDS result particularly when the measurement scheme relies on separate scans of the reference and sample. Figure 6 shows how the extraction of the real (a) and imaginary (b) refractive index of lithium niobate is affected when laser drift is simulated by scaling the amplitude of the sample scan with factors between 0.8 and 1.2. The Newton-Raphson fits are shown as dashed curves with their colour corresponding to the drift factor. Ensemble extractions trained with phase noise and white noise in the data set are shown by triangles and multicoloured lines and ensembles trained with laser drift in the training data are shown by the black diamonds.

When attempting to fit the refractive index with this erroneous data, the imaginary part of the calculated refractive index is strongly affected: an artificial increase in magnitude of a pulse that has propagated through a material would correspond to that material being less absorptive than in reality. The neural networks trained with data that has undergone a random scaling is able to consistently produce the same output regardless of the drifted signal for both the real and imaginary components. Note that networks trained with only noise are worse performing than the Newton-Raphson method with respect to the real refractive index (Fig. 6(a)). Correct fitting here is expected for the Newton-Raphson method due to the large contribution the phase information makes for the real refractive index in the theoretical model, of which is not affected by the scaling of the signal amplitude. The network trained with only noise fails in this case, highlighting the importance of a wide variety of simulation parameters in order to have more performant and reliable networks than the state of the art methods. The resilience the drift models retain are however very encouraging.

Optics EXPRESS

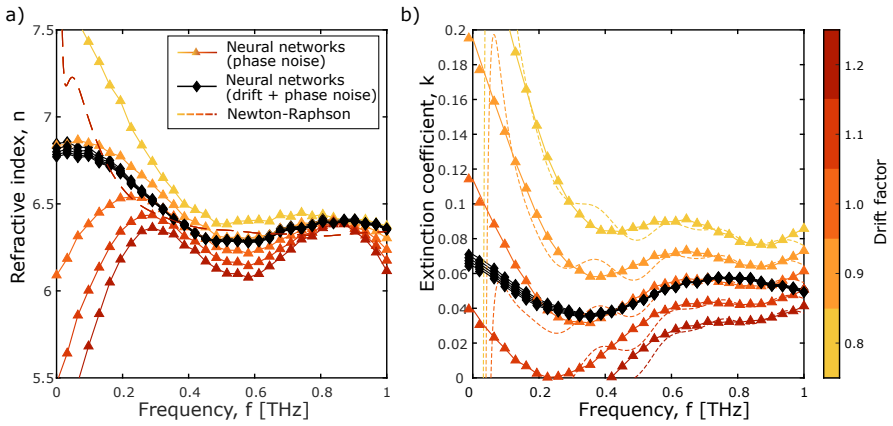

Fig. 6. Real (a) and imaginary (b) refractive index extracted from an experimentally measured sample of lithium niobate with a range of simulated laser drifts, modelled by scaling the transfer function ±20 %. Direct Newton-Raphson fits (dashed lines) show little error in the real refractive index but significant error in the extinction coefficient. Neural network ensembles trained with data containing phase noise (triangles) show errors in both parts of the refractive index. The network ensemble trained with laser drift in the training data (diamonds) remains stable for both the real and imaginary refractive index extractions, as there is very little deviation present.

5. Discussion

Training DNN ensembles shows improvement over a single model approach, as a way to avoid over fitting while still improving accuracy. Further, introducing relatively simple modelling of noise, experimental and analytical errors produce models that can interpret experimental data and provide consistent results, something that remains an issue in the THz community. The results on the phase offset error correction shows that the neural network can outperform the traditional Newton-Raphson method, which requires phase correction prior to extraction. The performance of the models trained with drift are not as strong as those with the phase offset error, however, the fact that the network can learn to mitigate this experimental artefact at all is useful as this is not something that can be implemented easily with traditional algorithms.

In future the demonstrated techniques could be applied to other sources of error such as those from the optical delay line, where the case of misalignment or motor inaccuracies [29] causes a scaling in the frequency domain along the frequency axis. In addition, we could in future correct for more complex phenomena which exist due to periodic sampling errors that reduce the usability of higher frequency components [30].

We believe these techniques are not limited to mitigating errors but also exhibit potential for solving complex problems that are possible to model but difficult to reverse engineer, such as multiple films and porous samples. Another example is in the measurement of protein hydration dynamics [31], which is challenging experimentally due to water's high attenuation of THz waves. In this case, neural networks can be used to extract the radius of the protein's hydration shell. To further improve these network ensembles, possible enhancements could be obtained through both optimizing the training strategy or by developing more complex architectures. One such example in which the resolution could be greatly improved is with the use of autoencoders in combination with the fully connected layers we see here [32], which would act as a more sophisticated alternative to linear interpolation. Simulation of the training data for the presented networks is fast, but a general model would also require acceptance of sample thickness in addition. A challenge associated with this is that the thickness changes the relative contribution

Optics EXPRESS

of propagation, transmission, and reflection coefficients. Therefore, this may require larger data sets containing samples with a range of thicknesses and correspondingly longer training times, but would result in a network that no longer requires retraining for different samples.

6. Conclusion

We demonstrate that DNN ensembles offer a solution to THz-TDS refractive index extraction that compensates for experimental deficiencies. The strength of DNNs is highlighted by their ability to correct for phase offsets and account for any laser drift that occurs during the experiment. By training up to ten neural networks with synthetic data to form an ensemble, we find that the stability of the extracted refractive index is improved due to the reduction of overall error. Surprisingly, simply introducing white noise into the synthetic training data yields ensembles that perform well with experimental measurements despite the significant divergence seen in the simulated noiseless case, highlighting the need for characterization with experimental, real world data to truly evaluate the performance of the models. Beyond introducing noise, we artificially model phase noise and laser drift, which further advances the networks' capability. Unlike the iterative, root finding methods, these networks are able to analyse multiple points in a frequency spectra simultaneously, which we believe is what enables the more robust extractions, and is a promising result given the modest degree of simulation required.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are available from the authors upon reasonable request.

References

1. M. Naftaly and R. E. Miles, "Terahertz time-domain spectroscopy for material characterization," Proc. IEEE **95**(8), 1658–1665 (2007).
2. R. J. Falconer and A. G. Markelz, "Terahertz spectroscopic analysis of peptides and proteins," J. Infrared Milli. Terahz Waves **33**(10), 973–988 (2012).
3. J. F. Federici, B. Schulkin, F. Huang, *et al.*, "Thz imaging and sensing for security applications—explosives, weapons and drugs," Semicond. Sci. Technol. **20**(7), S266–S280 (2005).
4. G. Jnawali, Y. Rao, H. Yan, *et al.*, "Observation of a transient decrease in terahertz conductivity of single-layer graphene induced by ultrafast optical excitation," Nano Lett. **13**(2), 524–530 (2013).
5. P. Jepsen, D. Cooke, and M. Koch, "Terahertz spectroscopy and imaging – modern techniques and applications," Laser Photonics Rev. **5**(1), 124–166 (2011).
6. R. B. Kohlhaas, S. Breuer, S. Mutschall, *et al.*, "Ultrabroadband terahertz time-domain spectroscopy using iii-v photoconductive membranes on silicon," Opt. Express **30**(13), 23896–23908 (2022).
7. V. Cherniak, T. Kubiczek, K. Kolpatzeck, *et al.*, "Laser diode based THz-TDS system with 133 db peak signal-to-noise ratio at 100 GHz," Sci. Rep. **13**(1), 13476 (2023).
8. U. Nandi, K. Dutzi, A. Deninger, *et al.*, "Eras:in(al)gaas photoconductor-based time domain system with 4.5 thz single shot bandwidth and emitted terahertz power of 164 μw," Opt. Lett. **45**(10), 2812–2815 (2020).
9. L. Duvillaret, F. Garet, and J. Coutaz, "A reliable method for extraction of material parameters in terahertz time-domain spectroscopy," IEEE J. Sel. Top. Quantum Electron. **2**(3), 739–746 (1996).
10. M. Naftaly, "An international intercomparison of THz time-domain spectrometers," *International Conference on Infrared, Millimeter, and Terahertz Waves*, 0–1 (2016).
11. J. Gorecki, N. Klokkou, L. Piper, *et al.*, "High-precision THz-TDS via self-referenced transmission echo method," Appl. Opt. **59**(22), 6744 (2020).
12. P. R. Whelan, K. Iwaszczuk, R. Wang, *et al.*, "Robust mapping of electrical properties of graphene from terahertz time-domain spectroscopy with timing jitter correction," Opt. Express **25**(3), 2725–2732 (2017).
13. Y. Abautret, D. Coquillat, M. Lequime, *et al.*, "Analysis of the multilayer organization of a sunflower leaf during dehydration with terahertz time-domain spectroscopy," Opt. Express **30**(21), 37971–37979 (2022).
14. Y. Jiang, G. Li, H. Ge, *et al.*, "Machine learning and application in terahertz technology: A review on achievements and future challenges," IEEE Access **10**, 53761–53776 (2022).
15. J. Liao, B. Wang, Z. Wang, *et al.*, "Amino-acid classification based on terahertz absorption spectroscopy with gaussian process and maximum likelihood," Sens. Actuators, B **388**, 133806 (2023).
16. N. Klokkou, J. Gorecki, J. S. Wilkinson, *et al.*, "Artificial neural networks for material parameter extraction in terahertz time-domain spectroscopy," Opt. Express **30**(9), 15583–15595 (2022).
17. Q. Wang, Y. Zhang, H. Ge, *et al.*, "Identification of rice freshness using terahertz imaging and deep learning," Photonics **10**(5), 547 (2023).

Optics EXPRESS

18. Y.-L. Zhu, R.-B. She, W.-Q. Liu, *et al.*, "Deep learning optimized terahertz single-pixel imaging," IEEE Trans. Terahertz Sci. Technol. **12**(2), 165–172 (2022).
19. M. Z. Güngördü, P. Kung, and S. M. Kim, "Non-destructive evaluation and fast conductivity calculation of various nanowire-based thin films with artificial neural network aided thz time-domain spectroscopy," Opt. Express **31**(6), 10657–10672 (2023).
20. J. Liu, D. Zhang, D. Yu, *et al.*, "Machine learning powered ellipsometry," Light: Sci. Appl. **10**(1), 55 (2021).
21. S. G. Worswick, J. A. Spencer, G. Jeschke, *et al.*, "Deep neural network processing of deer data," Sci. Adv. **4**(8), eaat5218 (2018).
22. Z. Zhou, S. Jia, and L. Cao, "A general neural network model for complex refractive index extraction of low-loss materials in the transmission-mode thz-tds," Sensors **22**(20), 7877 (2022).
23. L. Duvillaret, F. Garet, J.-F. Roux, *et al.*, "Analytical modeling and optimization of terahertz time-domain spectroscopy experiments, using photoswitches as antennas," IEEE J. Sel. Top. Quantum Electron. **7**(4), 615–623 (2001).
24. Z. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: Many could be better than all," Artificial Intelligen. **137**(1-2), 239–263 (2002). IJCAI 2001 Conference, SEATTLE, WA, 2001.
25. L. Hansen and P. Salamon, "Neural network ensembles," IEEE Trans. Pattern Anal. Machine Intell. **12**(10), 993–1001 (1990).
26. S. Hashem, "Optimal linear combinations of neural networks," Neural Networks **10**(4), 599–614 (1997).
27. G. P. Zhang, "A neural network ensemble method with jittered training data for time series forecasting," Inf. Sci. **177**(23), 5329–5346 (2007). Including: Mathematics of Uncertainty.
28. P. U. Jepsen, "Phase Retrieval in Terahertz Time-Domain Measurements: a "how to" Tutorial," J. Infrared Milli. Terahz Waves **40**(4), 395–411 (2019).
29. A. Mamrashev, F. Minakov, L. Maximov, *et al.*, "Correction of optical delay line errors in terahertz time-domain spectroscopy," Electronics **8**(12), 1408 (2019).
30. A. Rehn, D. Jahn, J. C. Balzer, *et al.*, "Periodic sampling errors in terahertz time-domain measurements," Opt. Express **25**(6), 6712–6724 (2017).
31. S. Ebbinghaus, S. J. Kim, M. Heyden, *et al.*, "An extended dynamical hydration shell around proteins," Proc. Natl. Acad. Sci. **104**(52), 20749–20752 (2007).
32. M. Ali, A. K. M. N. Haque, N. Sadik, *et al.*, "Predicting strongly localized resonant modes of light in disordered arrays of dielectric scatterers: a machine learning approach," Opt. Express **31**(2), 826–842 (2023).