

Objective and efficient terahertz signal denoising by transfer function reconstruction

Cite as: APL Photonics 5, 056104 (2020); <https://doi.org/10.1063/5.0002968>

Submitted: 21 February 2020 . Accepted: 24 April 2020 . Published Online: 12 May 2020

Xuequan Chen , Qiushuo Sun, Rayko I. Stantchev , and Emma Pickwell-MacPherson 



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[Ultra-wideband tri-layer transmissive linear polarization converter for terahertz waves](#)

APL Photonics 5, 046101 (2020); <https://doi.org/10.1063/1.5144115>

[High-resolution sub-sampling incoherent x-ray imaging with a single-pixel detector](#)

APL Photonics 5, 056102 (2020); <https://doi.org/10.1063/1.5140322>

[One-dimensional, surface emitting, disordered Terahertz lasers](#)

APL Photonics 5, 036102 (2020); <https://doi.org/10.1063/1.5131253>

APL Photonics The Future Luminary Award

Journal
Impact Factor
4.383

LEARN MORE!

Objective and efficient terahertz signal denoising by transfer function reconstruction

Cite as: APL Photon. 5, 056104 (2020); doi: 10.1063/5.0002968

Submitted: 21 February 2020 • Accepted: 24 April 2020 •

Published Online: 12 May 2020



View Online



Export Citation



CrossMark

Xuequan Chen,¹  Qiushuo Sun,² Rayko I. Stantchev,¹  and Emma Pickwell-MacPherson^{1,3,a)} 

AFFILIATIONS

¹Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong 999077, China

²Department of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, United Kingdom

³Physics Department, Warwick University, Coventry CV4 7AL, United Kingdom

^{a)}Author to whom correspondence should be addressed: E.MacPherson@warwick.ac.uk

ABSTRACT

As an essential processing step in many disciplines, signal denoising efficiently improves data quality without extra cost. However, it is relatively under-utilized for terahertz spectroscopy. The major technique reported uses wavelet denoising in the time-domain, which has a fuzzy physical meaning and limited performance in low-frequency and water-vapor regions. Here, we work from a new perspective by reconstructing the transfer function to remove noise-induced oscillations. The method is fully objective without a need for defining a threshold. Both reflection imaging and transmission imaging were conducted. The experimental results show that both low- and high-frequency noise and the water-vapor influence were efficiently removed. The spectrum accuracy was also improved, and the image contrast was significantly enhanced. The signal-to-noise ratio of the leaf image was increased up to 10 dB, with the 6 dB bandwidth being extended by over 0.5 THz.

© 2020 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0002968>

I. INTRODUCTION

Spectroscopy at terahertz (THz) frequencies has been extensively studied in the last three decades since the invention of THz time-domain spectroscopy (TDS). The straightforward extraction of material properties enables applications in physics,¹ chemistry,² and biology.³ To adapt to these applications, system developments have been focused on improving the acquisition speed,⁴ operational bandwidth,⁵ and functional devices⁶ and reducing the cost.⁷ Nowadays, a typical TDS system can provide a peak signal-to-noise ratio (SNR) of 60–80 dB. Photoconductive antenna based systems typically provide a usable bandwidth below 4 THz, but this can be greatly extended to over 30 THz using optical rectification or air-plasma, reaching the mid-infrared regime.^{8,9} However, the picosecond-short time-domain pulse dictates that the SNR must exponentially decrease toward the lower and upper bandwidth.^{10,11} The actual effective bandwidth in a specific experiment could be a lot shorter than the system's spectral range, depending mainly on the signal attenuation and the available integration time. A longer integration time is essential to improve the SNR by suppressing the

white noise. However, this is inefficient in TDS measurements as the longer time-constant has to be applied for all sampling points in the time-domain, instead of only the low SNR region. The SNR and the effective bandwidth are especially limited in applications where a fast acquisition is favorable or necessary, such as raster-scan imaging or monitoring a fast dynamic process. In many other studies, numerical signal denoising could be an efficient approach to improve the SNR.^{12,13} Unfortunately, it has been left as a relatively blank space for THz spectroscopy.

The simplest denoising techniques for TDS signals are time-domain windowing and Fourier filtering. The former only removes noise outside the main pulse region by smoothing the wings of the signal to zero.¹⁴ The main pulse is not denoised and can even be mis-shaped. The window function is subjectively selected and restricts many applications, of which multiple reflections are considered. Fourier filtering usually applies a low-pass or band-pass filter to remove noise outside the effective bandwidth.¹⁵ It improves the time-domain signal quality but has no improvement to the useful spectrum at all. Wavelet denoising is almost the only useful technique for THz spectroscopy. Unlike Fourier transform that

represents a time-domain signal by infinitely long sinusoidal waves, wavelet transform uses time-finite wavelets. The wavelets are shifted and scaled to decompose a THz signal. The resulting wavelet coefficients are distributed in a 2D wavelet-domain, representing the frequency compositions varying with time. To denoise a signal, a threshold value needs to be defined to set coefficients below the threshold to zero. Using wavelet denoising for THz signals was first proposed by Mittleman *et al.* for gas sensing.¹⁶ After that, successful applications in tomography and imaging were also reported.^{17,18} Apart from denoising, it has also been applied for THz deconvolution.^{19,20} The limitations of the wavelet technique, however, have seldom been discussed. An obvious point is the abstract physical meaning. Unlike Fourier coefficients, which stand for the physical properties of an electromagnetic wave, the processing of the wavelet coefficients is mainly numerically based with little physical meaning. Furthermore, it also has a limited performance at low frequencies and in the regions with dense water-vapor absorptions. Low-frequency wavelet components are usually greater than the threshold and are not filtered. Water vapors generate echoes that weakly oscillate in the time-domain. Some of them could be mistakenly classified as redundant noise in the wavelet coefficients and removed, altering the spectrum profile near the water-vapor lines. These can be seen from the examples in the work by Kim *et al.*²¹ and will also be demonstrated in our later comparison. From this point, wavelet denoising works better in the time-domain, such as the ECG (electrocardiography) signals.¹³ Last, the denoising performance also relies on the proper selection of the wavelet function, decomposition level, and filtering threshold. They should be carefully optimized to avoid shaping the signal or suppressing peaks in the spectrum, which makes the processing very subjective.²²

Apart from direct denoising, Pupeza *et al.* proposed a Spatially Variant Moving Average Filter (SVMAF) to reduce the influence of noise. The material properties are smoothed with the restriction that the transfer function falls within the confidence interval.²² Noise-induced oscillations on the characterized sample properties were efficiently removed. This also limits its versatility as it is not applicable when the sample properties are not extracted in a measurement. Another technique of autoregressive extrapolation reported by Dong *et al.* extends the effective bandwidth by predicting values of the transfer function from a high-SNR region.²³ Although the work is demonstrated for depth-resolution enhancement, it can be viewed as a denoising approach to improve the data quality in the low SNR region. The major limitation is that a continuous high-SNR interval should be provided. A signal may need to be measured in low humidity as noisy fluctuations induced by the water vapor, if included in the utilized high-SNR bandwidth, will not be denoised and may affect the prediction accuracy.

In this work, we propose a denoising method from a totally different perspective. Instead of directly processing the time-domain or frequency-domain of a signal, we denoise by objectively reconstructing the transfer function (i.e., the complex ratio between the sample and reference signals in the frequency-domain). We remove noise-induced spikes and oscillations in the transfer function by using a genetic algorithm (GA). Both reflection imaging and transmission imaging were conducted to verify the denoising performance. Water-IPA (isopropanol) mixtures were imaged to show a greatly enhanced contrast and reduced spatial variation, as well

as the improved characterization accuracy. A fresh leaf was also scanned by transmission to demonstrate the improvement in the image quality and the effective bandwidth.

II. METHOD

A typical THz-TDS measurement computes the transfer function M by comparing the sample signal to the reference signal in the frequency-domain as the first step for further data processing. Theoretically, this is expressed as

$$M_{theory}(\omega) = \frac{E_{sample}(\omega)}{E_{reference}(\omega)} = \frac{\prod_i R_i(\omega) R_{sample}(\omega)}{\prod_i R_i(\omega) R_{reference}(\omega)} = \frac{R_{sample}(\omega)}{R_{reference}(\omega)} = g(n_{sample}, \kappa_{sample}, \dots), \quad (1)$$

where $E_{sample}(\omega)$ and $E_{reference}(\omega)$ are the detected sample and reference signals in the frequency-domain. R_i may represent any systematic response that occurs in both the sample and reference measurements, for example, the emitted signal, the detector response, the reflection/transmission of any intermediate optical components, and the water-vapor absorptions. $R_{sample}(\omega)$ and $R_{reference}(\omega)$ are the response from the sample and the reference medium, respectively, with the latter already known for its optical properties. The division removes all the common terms, and $M_{theory}(\omega)$ only contains the sample response normalized to the reference. By establishing a proper optical model, $M_{theory}(\omega)$ can be expressed by a series of Fourier transmission and/or reflection coefficients, represented as g , a function of the sample optical properties (e.g., refractive index n and absorption coefficient α). In the case of no resonance arising from the sample, $M_{theory}(\omega)$ should be smooth in the frequency-domain without oscillations.^{22,24} Fabry-Perot resonance from a thick substrate may generate periodic variations in $M_{theory}(\omega)$. However, this can be mostly avoided in TDS by temporally removing the secondary reflection pulses. Thin-film samples only produce mild variations in the spectrum and do not affect the “smoothness.” In a practical measurement, however, considerable oscillations would occur to $M(\omega)$ due to the existence of noise, especially in low SNR regions. This can be seen from the following equation:

$$M_{raw}(\omega) = \frac{E_{sample}^{raw}(\omega)}{E_{reference}^{raw}(\omega)} = \frac{E_{sample}(\omega) + N_{sample}(\omega)}{E_{reference}(\omega) + N_{reference}(\omega)} \approx \frac{E_{sample}(\omega) + N_{sample}(\omega)}{E_{reference}(\omega)}, \quad (2)$$

where $M_{raw}(\omega)$ represents the sample/reference ratio from the raw data without denoising processing. $N_{sample}(\omega)$ and $N_{reference}(\omega)$ are the noise in the sample and reference signals. In most cases, a long integration time can be practically achieved for the reference signal, and thus, the equation can be simplified to the lower form by neglecting the reference noise. Compared to Eq. (1), as $N_{sample}(\omega)$ randomly changes for every frequency, it makes $M_{raw}(\omega)$ oscillate around $M_{theory}(\omega)$, especially when $N_{sample}(\omega)$ is comparable to $E_{sample}(\omega)$. Our denoising is based on the expectation that $M_{raw}(\omega)$ should be smooth in the absence of noise. The algorithm reconstructs a smooth transfer function $M_{smooth}(\omega)$ around $M_{raw}(\omega)$. Of course, not every $M_{smooth}(\omega)$ approximates $M_{theory}(\omega)$. To ensure

this, we compare the time-domain sample signal reconstructed from $M_{smooth}(\omega)$ and the detected sample signal,

$$E_{diff}(t) = E_{sample}^{raw}(t) - E_{sample}^{reconstructed}(t) = E_{sample}^{raw}(t) - iFFT(M_{smooth}(\omega) \cdot E_{reference}^{raw}(\omega)). \quad (3)$$

Theoretically, if $M_{smooth} = M_{theory}$, we have $E_{diff}(t) = N_{sample}(t)$. In contrast, if any point of M_{smooth} deviates away from M_{theory} , the smooth profile will drag the neighboring $\pm span$ region to be all away from M_{theory} , resulting in a large $E_{diff}(t)$. Therefore, the evaluation function f_{eval} can be set to minimize the standard deviation of the absolute difference,

$$f_{eval} = std(E_{diff}(t)). \quad (4)$$

We optimize M_{smooth} until f_{eval} is below the noise level $N_{level} = std(N_{sample}(t))$. N_{sample} can be estimated by many approaches, such as by measuring the signal with the light being blocked. The above optimization filters solutions M that are smooth, close to M_{raw} , and result in a reconstructed sample signal having difference from the detected sample signal in a noise level. A solution satisfying these criteria would have a high probability to approximate M_{theory} , which is then regarded as the denoised version of M_{raw} . We point out that a noiseless reference is ideal to support the above theory, but it is not compulsory. In case the reference is noisy, $E_{diff}(t)$ by Eq. (3) (with $M_{smooth} = M_{theory}$) is changed to

$$E_{diff}(t) = N_{sample}(t) - iFFT(M_{theory}(\omega) \cdot N_{reference}(\omega)) \leq 2 \times N_{sample}(t), \quad \text{if } |M_{theory}(\omega)| \leq 1.$$

In this case, the stopping criterion should be changed to $f_{eval} < std(2N_{sample}(t))$. M_{smooth} can still be found theoretically, while the accuracy could be reduced due to the weaker stopping criterion. In addition, the reconstructed sample signal in the time-domain will be as noisy as the reference signal.

As $M(\omega)$ usually contains hundreds of complex data points to be optimized, denoising is realized by using the GA. The GA is powerful in multiple-domain optimization problems and can provide a fast convergence to approach the global minimum.²⁵ It mimics the evolution in nature by following a major iteration process of initialization–evaluation–crossover–mutation. The detailed steps are shown in Algorithm 1.

In step 1, a band-pass filter is applied to the sample and reference signals to remove the useless noise outside the effective bandwidth. M_{raw} is calculated according to Eq. (2) and decomposed into its magnitude $|M_{raw}|$ and phase $Arg(M_{raw})$ for later processing.

ALGORITHM 1. Genetic algorithm steps.

1. Band-pass filter pre-processing. Calculate $M_{raw}(\omega)$
2. **Initialization.** Randomly initialize $M_i(\omega)$, $0 \leq i \leq NumPop$
3. **Evaluation.** Calculate f_{eval} by Eq. (4) and rank
4. **Crossover.** Select the top half to generate new populations
5. **Mutation.** Mutate P_{mut} percent of the crossover population
6. If $f_{eval_1} < N_{level}$ or Iteration $> NumItp$
Go to step 7 if yes. Go to step 3 if no.
7. Output the top ranked M_i in the last iteration

Step 2 initializes M_i . To build M_i with a higher chance of approximating M_{theory} , the key issue is to keep M_i close to M_{raw} and smooth. The simplest example is directly applying a smooth function to M_{raw} . Mathematically, this is represented by

$$Smooth(M_{raw}, span) = \begin{cases} \frac{\sum_1^{2p-1} M_{raw}(p)}{2p-1}, & p - span < 1 \\ \frac{\sum_{p-span}^{p+span} M_{raw}(p)}{2span+1}, & 1 + span < p < p_{max} - span \\ \frac{\sum_{2p-p_{max}}^{p_{max}} M_{raw}(p)}{2p_{max}-2p+1}, & p + span > p_{max}. \end{cases} \quad (5)$$

This *Smooth* function is the same as the one used in *Matlab*. Here, p is the index and p_{max} denotes the length of M_{raw} . The *Smooth* function averages M_{raw} in every $\pm span$. The span at the start and the end of the M_{raw} array was reduced to adapt to the available data points. The length should be selected such that the smoothed curve can maintain the dispersion profile. An example is shown in Fig. 1. Here, M_{raw} is the ratio between a quartz-water (sample) and a quartz-air (reference) reflection, with the magnitude $|M_{raw}|$ shown as the gray curve. The dispersive properties of water result in a frequency-dependent curve. Noise-induced oscillations are mostly distributed in the high-frequency region where the SNR is low. The *span* is set to be the length of 0.05 THz, and the curve is smoothed for every 0.1 THz range, as shown in the black curve. The oscillations have been greatly suppressed, but the dispersion is maintained. However, it may not approximate M_{theory} . We initialize a large number of M_i and find out the best one. This is done by randomly sampling every 0.1 THz ($=2span$). At every sampled point, a value was randomly assigned in-between $|M_{raw}|$ and $|Smooth(M_{raw})|$. Two examples are shown as the red dots and blue triangles in Fig. 1.

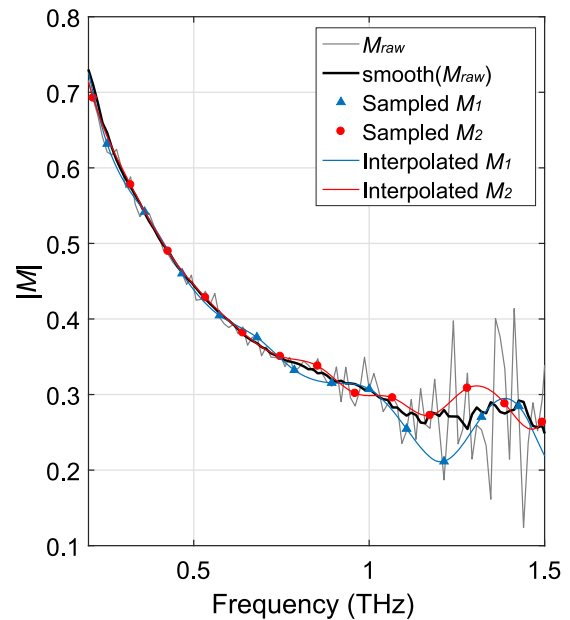


FIG. 1. An example of initializing $|M_i|$ from $|M_{raw}|$ and $|smooth(M_{raw})|$.

A spline interpolation is applied to the assigned points to obtain values at all frequencies, shown as the red and blue solid curves. Such initialization ensures curves have good smoothness and are close to M_{raw} . The values in the high-SNR regions are very close to M_{raw} because they have a narrow assigning range, while they have large reconstructive flexibilities where noisy oscillations occur. Similar initialization was also applied to $\text{Arg}(M_{raw})$ to get the complex arrays of $M_i = |M_{raw}| \exp(i \text{Arg}(M_{raw}))$. The number of populations (i.e., the total number of M_i) was set as $NumPop$.

The initialized M_i were substituted into Eqs. (3) and (4) to calculate f_{eval} , and they were ranked from low to high in step 3. The top half represents a higher chance to approximate M_{theory} . They were kept for crossover in step 4. The lower half is abandoned.

The crossover in step 4 calculates the average of every two adjacent populations. As the remaining populations could be close to M_{theory} , their averages may inherit some features from their parent population and could also have a high chance to approximate M_{theory} .

To avoid being trapped in local minima, the mutation process in step 5 is set to increase the optimization divergency. In detail, P_{mut} percentage of the crossover populations was selected. A frequency point in the effective bandwidth region of the selected population was randomly picked, and the corresponding value was randomly changed by less than 5%. Directly doing this generates a spike to break the smoothness. Therefore, the values at a $span$ distance [same as the $span$ in Eq. (5)] from the mutation point were abandoned, with a spline interpolation used to reassign their values. This can be expressed as

$$M_i^{mut} = y_i = \text{Interpolate}(x, y, x'_i, \text{spline}'), \quad (6)$$

where

$$\begin{aligned} x &= (1 : p - span), p, (p + span), \\ y &= M_i(x), \\ x_i &= 1 : p_{max}, \end{aligned}$$

where p denotes the index of the point being mutated. The smoothness was maintained by the spline interpolation. The mutation in step 5 creates new genes to improve the optimization divergency and accuracy.

Step 6 checks the iteration condition. f_{eval_1} represents the top ranked evaluation value. If $f_{eval_1} < N_{level}$, or the iteration has reached the maximum iteration number $NumItp$ specified, the program stops and outputs the top ranked M_i in the last generation in step 7. Otherwise, steps 3–5 iterate to keep optimizing M_i .

In our experiments, the GA parameters were set as $NumPop = 200$, $NumItp = 50$, and $P_{mut} = 0.1$. These parameters are different from the thresholds in wavelet denoising that conclusively affect the denoising results. They are adjusted only to balance the processing speed and accuracy without affecting the objective evaluation principle.

III. EXPERIMENTS

Two measurements were conducted to verify the algorithm performance. First, water-IPA mixture solutions of different volume fractions were imaged in reflection under 70% relative humidity. The

measurement was performed using the Menlo K15 reflection THz-TDS system as described in our previous work,²⁶ with a quartz window above the scanning THz optics and samples being placed on top of the window. In the second experiment, a fresh leaf was scanned in a free-space transmission TDS system with a 4.5% relative humidity.

A. Water-IPA imaging

In the first imaging experiment, we raster scanned a square area of $30 \times 30 \text{ mm}^2$ using a step of 0.5 mm in both directions. The THz waveforms were acquired at a rate of 4 Hz, which in the frequency-domain give an SNR = [40, 24, 12] dB at [0.4, 1.0, 1.4] THz, respectively. Measurements were conducted in a relative humidity of 70%, which introduces many water-vapor-absorption lines. Liquid samples were placed on the quartz window and contained by a rubber ring, indicated by the circle in Fig. 2. At first, 1.5 ml pure water was in the circular area. The THz optics started at the original point $(x, y) = (0, 0)$ and scanned along the y-direction to $(0, 30)$. It then moved to the next line to make a Z-route scan. Right before scanning the line of $x = 7.5 \text{ mm}$, 0.27 ml IPA was added and mixed with the water in the circle, resulting in a solution with an IPA volume fraction of 15%. Similarly, right before the lines of $x = 15 \text{ mm}$ and $x = 22.5 \text{ mm}$, 0.36 ml and 0.6 ml of IPA were mixed with the previous solution and increased the IPA volume fraction to be 30% and 45%, respectively. By doing this, we manually create three sharp boundaries in-between the four mixtures, as shown in the four separated regions inside the

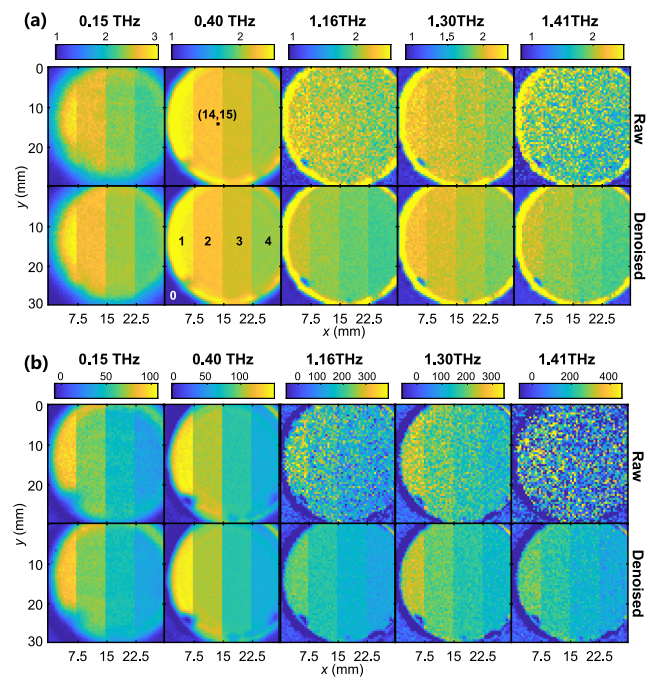


FIG. 2. (a) Refractive index and (b) absorption coefficient images from the raw and denoised data at 0.15 THz, 0.4 THz, 1.16 THz, 1.3 THz, and 1.41 THz. The labels 0, 1, 2, 3, and 4 in (a) indicate the region of air and IPA-water mixtures with IPA volume fractions of 0%, 15%, 30%, and 45%, respectively. The dot in (a) indicates the data at (14, 15), which are shown as an example in Fig. 3.

circle, labeled as 1, 2, 3, and 4 in Fig. 2(a). The area outside the circle is ambient air and labeled as 0. n and α at different frequencies with and without denoising processing are shown in the upper and lower parts of Figs. 2(a) and 2(b), respectively.

The presented five frequencies have different SNRs. 0.15 THz locates at the lowest effective bandwidth with a relatively low SNR. The raw data have a certain level of noise but do not obviously affect the image quality because the contrast among different volume fractions was very large at low frequencies. We can observe the improvement by the algorithm from the more uniform color distribution. 0.4 THz has the best SNR of the spectrum. The contrast of the denoised images is only slightly improved. It is more important that consistent colors are found for the raw and denoised images, verifying that the accuracy was well maintained after denoising. 1.16 THz locates at a deep water-vapor line and thus is strongly attenuated.^{27,28} Both n and α of the raw data are very noisy and can barely display the boundaries. Promisingly, denoising was able to retrieve the noiseless values, and the four regions are clearly separated. 1.3 THz was less affected by water vapor but was closer to the upper bandwidth limit. The boundaries of α are barely noticeable, while they can still be clearly seen after denoising. 1.41 THz has a very weak SNR due to both the system bandwidth and the water-vapor. In addition, the difference among different volume fractions at high frequencies is much smaller. The raw images are extremely noisy, and α can even be negative. Even in this case, one can still easily distinguish the four regions in the denoised images. The comparison demonstrates the efficient denoising in the full bandwidth. The successful application to the four mixtures and the outside air regions shows the good adaptability to various types of signals.

In Fig. 3, we show the denoising detail of the middle point at (14, 15). The red curves in Fig. 3(a) show the magnitude and phase of

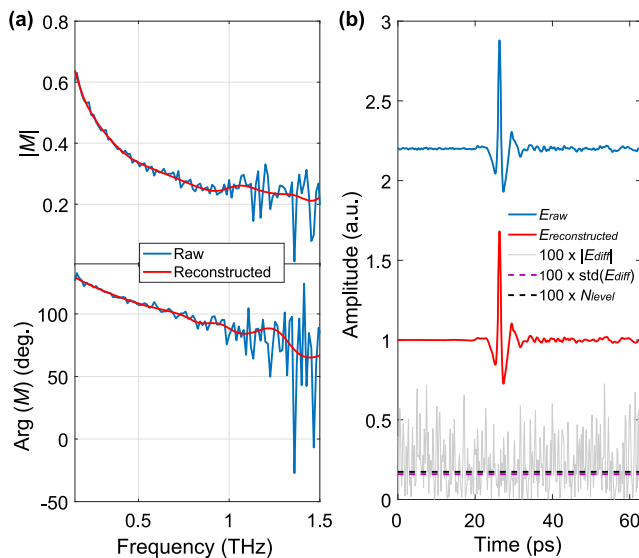


FIG. 3. Denoising example of the image point at (14, 15). (a) Magnitude and phase of the reconstructed M_{smooth} and the detected M_{raw} . (b) Raw and reconstructed sample signals in the time-domain (offset) and their absolute difference $|E_{\text{diff}}|$ multiplied by 100. The standard deviation of E_{diff} is given to be compared with the noise level.

the reconstructed M_{smooth} output from the algorithm. Compared to M_{raw} shown in the blue curves, the noise-induced oscillations were obviously removed. To evaluate the accuracy, Fig. 3(b) shows the corresponding reconstructed and raw sample signals (they are offset for clarity). The greatly reduced noise can be clearly observed from the flat region before the main pulse. Note that the raw signal has been processed with the band-pass filter to remove the meaningless spectrum range, and the improvement here is solely from denoising the effective bandwidth. Their absolute difference $|E_{\text{diff}}|$ multiplied by a factor of 100 is shown as the gray curve. The dashed purple line calculates its standard deviation, which is slightly lower than the noise level shown as the dashed black line. This indicates that the optimization of this point was stopped by the condition of $f_{\text{eval}_1} < N_{\text{level}}$.

To quantitatively compare the contrast enhancement, we plot n and α at 1.16 THz and 1.41 THz along the x-direction with and without denoising in Fig. 4. Each symbol represents the average of the data in every three lines ($\Delta x = 1.5$ mm and $y = 10\text{--}20$ mm). The error bars are given by their standard deviation. Figure 4(a) shows that, at 1.16 THz, a clear stair-shape is resolved in the denoised data, with the error bars of different regions separated from each other. In contrast, the boundaries in the raw data cannot be recognized, and the error bars of different regions are all vastly overlapped. We can even observe a small increasing trend in each region from the denoised results. This is due to slight IPA evaporation during the image scanning, which can also be verified from the tiny color variation in Fig. 2. This tiny variation is not observable in the raw data. Figure 4(b) gives the results at 1.41 THz. Due to the lower SNR and sample contrast, we can only observe an overall decreasing trend in the raw data. After denoising, the standard deviation was significantly reduced by a factor of about 7, providing a stair-shape

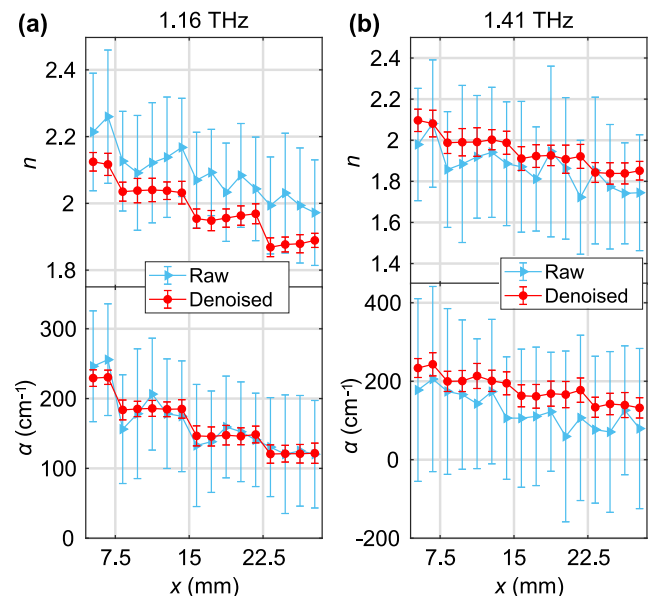


FIG. 4. n and α at (a) 1.16 THz and (b) 1.41 THz from the raw and denoised data varying in the x-direction. The error bars are calculated from the standard deviation.

distribution. For both frequencies, obvious changes in both n and α are noticed after denoising. This is due to the inaccurate raw data at these two frequencies affected by the intense water-vapor attenuation. The accuracy is greatly improved after denoising. This is verified by the following full spectrum comparison.

We take n and α of the 30% IPA–water mixture (i.e., region 3 in Fig. 2) as an example. The data in $x = 15\text{--}22$ mm and $y = 5\text{--}25$ mm were averaged and are shown as a function of frequency in Fig. 5. Figures 5(a) and 5(b) compare the raw data with the reported wavelet denoising and the proposed denoising method, respectively. The error bars stand for the standard deviation, and the bright green, orange, and purple lines on the background indicate the water-vapor absorption lines at 0.56 THz, 1.16 THz, and 1.41 THz, respectively.²⁸ The raw data show increasing errors toward higher frequencies. However, the errors in the water-vapor regions are much higher than the neighboring values, indicating an extra influence by the water vapor. This results in an accuracy that is greatly affected, as evidenced by the abrupt spikes that are not expected as both water and IPA contain no resonance in the studied frequency range.²⁴ The wavelet denoising obviously reduces the standard deviations; however, the accuracy is also significantly affected. This is in line with our discussion in the Introduction. In contrast, the denoised results are all within the error bars of the raw data. The curves present reasonable continuity without spikes. This is especially noticeable in water-vapor regions, showing a consistent variation and unchanged error bars with the neighboring values. The significantly improved accuracy by denoising well explains the change in n and α in Fig. 4. It is an outstanding ability of being able to remove the influence of water-vapor absorptions in the THz range, as this has barely been realized

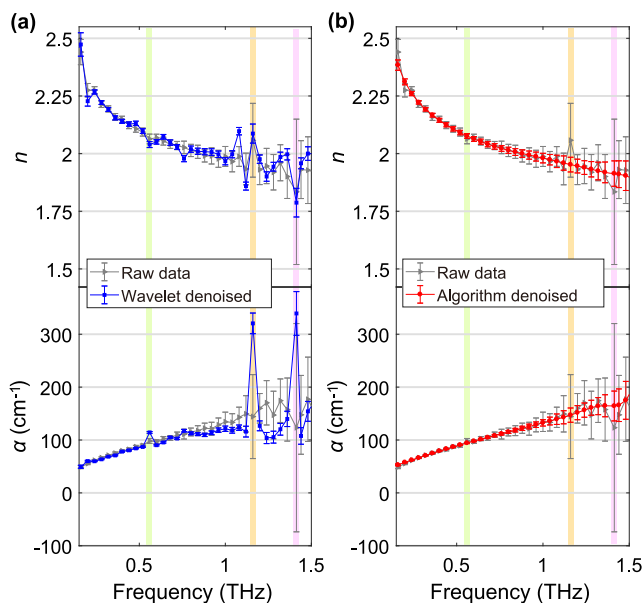


FIG. 5. Average n and α of the 30% IPA mixture in the frequency-domain, with the error bars given by the standard deviation. The raw data are compared with (a) wavelet denoising and (b) the proposed algorithm. The bright green, orange, and purple lines on the background refer to the water-vapor-absorption lines at 0.56 THz, 1.16 THz, and 1.41 THz.

in an effective way. The only reported work on removing water-vapor effects by using deconvolution²⁹ was done by Withayachumnankul *et al.* However, the method relies on an accurate water-vapor model under the specific humidity, temperature, and pressure and requires a high frequency resolution and SNR to resolve the water-vapor lines. Here, the proposed denoising algorithm does not depend on any physical circumstance or spectrum resolution, providing a much better versatility.

B. Leaf transmission imaging

To demonstrate the practical application of the algorithm, we scanned a fresh leaf in transmission under a low relative humidity of 4.5%. As one of the first THz imaging examples, Hu and Nuss demonstrated an impressive THz leaf image in 1995 showing the capability of using THz to investigate the water-content in biological samples.³⁰ Since then, many water-sensitive studies were reported, such as skin and breast cancer detection.³¹ The study of leaves was also further extended to more practical applications such as monitoring the water dynamics and exploring the leaf permittivity model.^{32,33} Here, we show how our algorithm can efficiently improve the image SNR and extend the effective bandwidth to assist these applications.

The leaf transmittance at 0.6 THz, 1.45 THz, 1.67 THz, 2.1 THz, and 2.5 THz is shown in Fig. 6(a), with the raw and denoised images shown in the upper and lower parts, respectively. A homogeneous area of the leaf and air regions was labeled with red and blue rectangles, respectively. The SNRs of these two areas were calculated by the ratio between their average E-field and their standard deviation, represented by $SNR = 20 \log_{10}[\text{mean}(E(\omega))/\text{std}(E(\omega))]$. The results of the raw and denoised data are given in Fig. 6(b). Due to the large water content, the signals of the leaf region at high frequencies are strongly attenuated, reflected by the darker color in Fig. 6(a) and the low SNR of the red curves in Fig. 6(b). Both images at 0.6 THz provide a relatively high 20 dB SNR in the leaf region. The algorithm only makes a little improvement in the SNR. The image resolution was poor due to the diffraction limit at this long wavelength. The raw images at 1.45 THz and 1.67 THz show obvious noise due to the water-vapor absorption, as can be seen from the dashed red curves in Fig. 6(b). The denoising algorithm fully eliminated the water-vapor influence and reconstructed clear images to resolve the veins on the leaf. This is also reflected by the red solid curve in Fig. 6(b) where the SNR was improved by 6.5 dB and is smooth over these water-vapor regions. These two examples show that water vapor still affects the SNR at the humidity as low as 4.5%, and the ability to remove water-vapor effects is important even in a dry-air environment. The raw image at 2.1 THz only has a 6 dB SNR in the leaf region. The veins are barely noticeable. The denoising algorithm reduces the noise to 4.4 dB and can clearly resolve the veins with a high resolution. The last raw image at 2.5 THz has the SNR less than 0 dB in the leaf region. Even the highly opaque stem was fully merged in noise. The denoised image clearly displays the stem and some blurry features of the veins by increasing the SNR to over 10 dB. We can also observe the noise reduced in the air region, as the SNR was also obviously increased from 17.3 dB to 24.2 dB. One may also notice the SNR improvement outside of the effective bandwidth in Fig. 6(b) (e.g., >3 THz of the red curves). This is because the GA initializes $M_{\text{smooth}}(\omega)$ in between $M_{\text{raw}}(\omega)$ and $\text{Smooth}(M_{\text{raw}}(\omega))$, making $\text{std}(M_{\text{smooth}})$

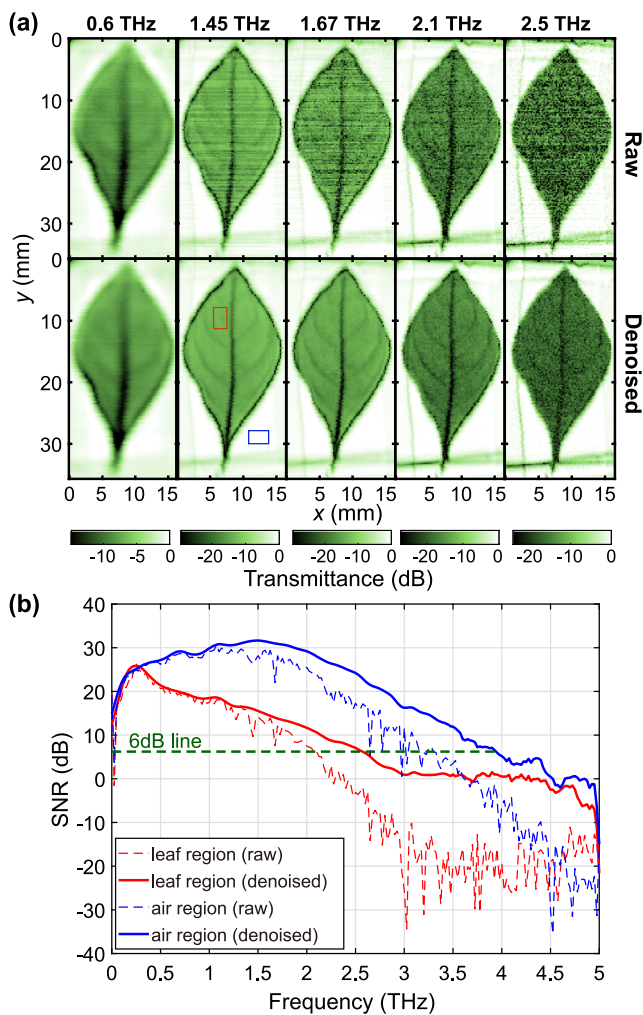


FIG. 6. (a) Transmittance images of the leaf without and with denoising at 0.6 THz, 1.45 THz, 1.67 THz, 2.1 THz, and 2.5 THz. (b) SNR of the leaf and air regions labeled with the red and blue rectangles in (a), calculated by $SNR = 20 \log_{10} |mean(E(\omega))/std(E(\omega))|$.

smaller than $std(M_{raw})$. With $E = M \cdot E_{reference}$, $std(E_{denoised})$ is smaller than $std(E_{raw})$ to result in a higher SNR. However, the 0 dB SNR indicates no useful information was provided. This is decided by the denoising principle as even if M_{smooth} has deviated from M_{theory} far outside of the effective bandwidth, the reconstructed signal is not affected as the amplitudes in this region are extremely small. Therefore, M_{smooth} cannot be optimized with great certainty. Here, we set the 6 dB threshold to define the effective SNR, as indicated in Fig. 6(b).

The cross points between the 6 dB line and curves also demonstrate a broadened effective bandwidth. The upper frequency limit of the leaf and air regions was extended from 2.075 THz to 2.6 THz and from 3.275 THz to 3.95 THz, corresponding to a bandwidth extension of 0.525 THz and 0.675 THz, respectively.

The broadened bandwidth effectively enriches the spectrum information for spectroscopic studies.

IV. DISCUSSION

In this article, we propose a denoising method by reconstructing the transfer function using the GA. The method is built on the expectation that the transfer function should be smooth over the effective bandwidth in the absence of noise and resonance. The GA objectively optimizes the smooth transfer function by minimizing the difference between the reconstructed signal and the detected signal. Two experiments were conducted to verify the performance and demonstrate the application capabilities. The water-IPA reflection imaging shows the greatly enhanced contrast and improved image quality. Further analysis shows that both low- and high-frequency noise and the water-vapor effect were successfully removed by the algorithm. The data accuracy in both the spatial domain and frequency-domain was obviously improved. The leaf imaging demonstrates the application in transmission and low humidity. The SNR of the images was clearly improved up to 10 dB. The water-vapor influence is still obvious at the very low humidity and was fully eliminated by the algorithm. The 6 dB bandwidth was extended by over 0.5 THz. The demonstrated performance would be very helpful in THz imaging, especially in biomedical studies, such as cancer,³⁴ scar,³⁵ and diabetic diagnosis.³⁶ It would also be very useful in many studies demanding a rapid data acquisition that cannot provide a high SNR, for example, the degradation of perovskite,³⁷ the skin occlusion,³⁸ and the fast phase transition of vanadium dioxide.³⁹ Similar to the autoregressive technique,²³ the functionality of extending the effective bandwidth can also enhance the temporal resolution in THz deconvolution, which may further benefit biomedical or industrial applications that demand a good depth-resolving ability.^{40,41}

The denoising method is effective for white noise as it removes the random oscillations in the frequency-domain induced by white-noise. Statistical errors introduced by laser instability and temperature or humidity variations cannot be eliminated by this approach because these errors modulate the whole spectrum without producing random oscillations. However, such statistical errors can be accounted for if the sample is placed on an imaging window rather than in free space, as the initial reflection of the imaging window can be used for calibration.²⁶ Despite this, the proposed method objectively and efficiently eliminates the noise in the whole effective bandwidth. The ability of removing low-frequency and water-vapor noise is superior than the existing technique of wavelet denoising. The proposed method improves the SNR and extends the bandwidth, which also allows a faster acquisition rate to be used. For example, it may accelerate the imaging speed or increase the sampling rate in monitoring a fast variation process. The processing of the transfer function is independent of the system types and signal profile, making it widely adaptive to all types of THz-TDS systems such as those based on air-plasma that have bandwidths up to 30 THz. It may also be applied in other interference based techniques such as Fourier-transform infrared (FTIR) spectroscopy. This denoising algorithm can therefore serve as a powerful tool in various photonics based spectroscopy applications.

ACKNOWLEDGMENTS

This work was partially supported by the Research Grants Council of Hong Kong (Project No. 14206717), the Hong Kong Innovation and Technology Fund (Project No. ITS/371/16), and the Royal Society Wolfson Merit Award (E.P.-M.)

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- ¹S. Zhang, Y. Shik Park, J. Li, X. Lu, W. Zhang, and X. Zhang, *Phys. Rev. Lett.* **102**, 023901 (2009).
- ²B. Fischer, M. Hoffmann, H. Helm, G. Modjesch, and P. U. Jepsen, *Semicond. Sci. Technol.* **20**, S246 (2005).
- ³Q. Sun, Y. He, K. Liu, S. Fan, E. P. J. Parrott, and E. Pickwell-MacPherson, *Quant. Imaging Med. Surg.* **7**, 345 (2017).
- ⁴C. Janke, M. Först, M. Nagel, H. Kurz, and A. Bartels, *Opt. Lett.* **30**, 1405 (2005).
- ⁵L. L. Zhang, W. M. Wang, T. Wu, R. Zhang, S. J. Zhang, C. L. Zhang, Y. Zhang, Z. M. Sheng, and X. C. Zhang, *Phys. Rev. Lett.* **119**, 235001 (2017).
- ⁶X. Chen, K. Li, R. Zhang, S. K. Gupta, A. K. Srivastava, and E. Pickwell-MacPherson, *Adv. Opt. Mater.* **7**, 1901321 (2019).
- ⁷H. Roehle, R. J. B. Dietz, H. J. Hensel, J. Böttcher, H. Künzel, D. Stanze, M. Schell, and B. Sartorius, *Opt. Express* **18**, 2296 (2010).
- ⁸A. Bonvalet, M. Joffre, J. L. Martin, and A. Migus, *Appl. Phys. Lett.* **67**, 2907 (1995).
- ⁹B. Clough, J. Dai, and X.-C. Zhang, *Mater. Today* **15**, 50 (2012).
- ¹⁰M. Tonouchi, *Nat. Photonics* **1**, 97 (2007).
- ¹¹A. Singh, A. Pashkin, S. Winnerl, M. Welsch, C. Beckh, P. Sulzer, A. Leitenstorfer, M. Helm, and H. Schneider, *Light Sci. Appl.* **9**, 1 (2020).
- ¹²A. Buades, B. Coll, and J. M. Morel, in *Proceedings of 2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005)* (IEEE, 2005), Vol. II, p. 60.
- ¹³M. Alfaouri and K. Daqrouq, *Am. J. Appl. Sci.* **5**, 276 (2008).
- ¹⁴W. Withayachumnankul, B. Ferguson, T. Rainsford, S. P. Micken, and D. Abbott, *Proc. SPIE* **5840**, 221 (2005).
- ¹⁵J. Choi, S. Y. Ryu, W. S. Kwon, K.-S. Kim, and S. Kim, *J. Opt. Soc. Korea* **17**, 454 (2013).
- ¹⁶D. M. Mittleman, R. H. Jacobsen, R. Neelamani, R. G. Baraniuk, and M. C. Nuss, *Appl. Phys. B: Lasers Opt.* **67**, 379 (1998).
- ¹⁷B. Ferguson and D. Abbott, *Microelectron. J.* **32**, 943 (2001).
- ¹⁸D. M. Mittleman, M. Gupta, R. Neelamani, R. G. Baraniuk, J. V. Rudd, and M. Koch, *Appl. Phys. B: Lasers Opt.* **68**, 1085 (1999).
- ¹⁹E. P. J. Parrott, Y. Sun, and E. Pickwell-MacPherson, *J. Mol. Struct.* **1006**, 66 (2011).
- ²⁰Y. Chen, S. Huang, and E. Pickwell-MacPherson, *Opt. Express* **18**, 1177 (2010).
- ²¹Y.-C. Kim, K.-H. Jin, J.-C. Ye, J.-W. Ahn, and D.-S. Yee, *J. Opt. Soc. Korea* **15**, 103 (2011).
- ²²I. Pupeza, R. Wilk, and M. Koch, *Opt. Express* **15**, 4335 (2007).
- ²³J. Dong, A. Locquet, and D. S. Citrin, *Opt. Lett.* **42**, 1828 (2017).
- ²⁴X. Chen and E. Pickwell-MacPherson, *Sensors* **19**, 4118 (2019).
- ²⁵K. Li, X. Chen, S. Shen, R. Zhang, and E. Pickwell-MacPherson, *IEEE Trans. Terahertz Sci. Technol.* **9**, 675 (2019).
- ²⁶X. Chen, E. P. J. Parrott, B. S.-Y. Ung, and E. Pickwell-MacPherson, *IEEE Trans. Terahertz Sci. Technol.* **7**, 493 (2017).
- ²⁷M. van Exter, C. Fattinger, and D. Grischkowsky, *Opt. Lett.* **14**, 1128 (1989).
- ²⁸X. Xin, H. Altan, A. Saint, D. Matten, and R. R. Alfano, *J. Appl. Phys.* **100**, 094905 (2006).
- ²⁹W. Withayachumnankul, B. M. Fischer, and D. Abbott, *Proc. R. Soc. A* **464**, 2435 (2008).
- ³⁰B. B. Hu and M. C. Nuss, *Opt. Lett.* **20**, 1716 (1995).
- ³¹P. C. Ashworth, E. Pickwell-MacPherson, E. Provenzano, S. E. Pinder, A. D. Purushotham, M. Pepper, and V. P. Wallace, *Opt. Express* **17**, 12444 (2009).
- ³²E. Castro-Camus, M. Palomar, and A. A. Covarrubias, *Sci. Rep.* **3**, 2910 (2013).
- ³³C. Jördens, M. Scheller, B. Breitenstein, D. Selmar, and M. Koch, *J. Biol. Phys.* **35**, 255 (2009).
- ³⁴E. Pickwell-MacPherson and V. P. Wallace, *Photodiagn. Photodyn. Ther.* **6**, 128 (2009).
- ³⁵S. Fan, B. S. Y. Ung, E. P. J. Parrott, V. P. Wallace, and E. Pickwell-MacPherson, *J. Biophotonics* **10**, 1143 (2016).
- ³⁶G. G. Hernandez-Cardoso, M. Alfaro-Gomez, S. C. Rojas-Landeros, I. Salas-Gutierrez, and E. Castro-Camus, *J. Infrared, Millimeter, Terahertz Waves* **39**, 879 (2018).
- ³⁷Q. Sun, X. Liu, J. Cao, R. I. Stantchev, Y. Zhou, X. Chen, E. P. J. Parrott, J. Lloyd-Hughes, N. Zhao, and E. Pickwell-MacPherson, *J. Phys. Chem. C* **122**, 17552 (2018).
- ³⁸Q. Sun, E. P. J. Parrott, Y. He, and E. Pickwell-MacPherson, *J. Biophotonics* **11**, e201700111 (2018).
- ³⁹E. P. J. Parrott, C. Han, F. Yan, G. Humbert, A. Bessaudou, A. Crunteanu, and E. Pickwell-MacPherson, *Nanotechnology* **27**, 205206 (2016).
- ⁴⁰E. Pickwell, B. E. Cole, A. J. Fitzgerald, M. Pepper, and V. P. Wallace, *Phys. Med. Biol.* **49**, 1595 (2004).
- ⁴¹J. Dong, B. Kim, A. Locquet, P. McKeon, N. Declercq, and D. S. Citrin, *Composites, Part B* **79**, 667 (2015).