Terahertz deconvolution

Article

Published Version

Walker, G. C., Bowen, J. W., Labaune, J., Jackson, J.-B., Hadjiloucas, S. ORCID: https://orcid.org/0000-0003-2380-6114, Roberts, J., Mourou, G. and Menu, M. (2012) Terahertz deconvolution. Optics Express, 20 (25). pp. 27230-27241. ISSN 1094-4087 doi: 10.1364/OE.20.027230 Available at https://reading-clone.eprints-hosting.org/31533/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1364/OE.20.027230

www.reading.ac.uk/centaur

Terahertz deconvolution

Gillian C. Walker,[1,*] **John W. Bowen,**[1] **Julien Labaune,**[2] **J-Bianca Jackson,**[2] **Sillas Hadjiloucas,**[1] **John Roberts,**[1] **Gerard Mourou,**[2] **and Michel Menu**[3]

[1]School of Systems Engineering, University of Reading, Reading, UK
[2]ILE, École Polytechnique, Palaiseau, France
[3]C2RMF, Palais du Louvre, Paris, France
[*]g.c.walker@reading.ac.uk

Abstract: The ability to retrieve information from different layers within a stratified sample using terahertz pulsed reflection imaging and spectroscopy has traditionally been resolution limited by the pulse width available. In this paper, a deconvolution algorithm is presented which circumvents this resolution limit, enabling deep sub-wavelength and sub-pulse width depth resolution. The algorithm is explained through theoretical investigation, and demonstrated by reconstructing signals reflected from boundaries in stratified materials that cannot be resolved directly from the unprocessed time-domain reflection signal. Furthermore, the deconvolution technique has been used to recreate sub-surface images from a stratified sample: imaging the reverse side of a piece of paper.

©2012 Optical Society of America

OCIS codes: (100.0100) Image processing; (070.0070) Fourier optics and signal processing.

References and links

1.   J. Takayanagi, H. Jinno, S. Ichino, K. Suizu, M. Yamashita, T. Ouchi, S. Kasai, H. Ohtake, H. Uchida, N. Nishizawa, and K. Kawase, "High-resolution time-of-flight terahertz tomography using a femtosecond fiber laser," Opt. Express **17**(9), 7533–7539 (2009).
2.   Y. Chen, S. Huang, and E. Pickwell-MacPherson, "Frequency-wavelet domain deconvolution for terahertz reflection imaging and spectroscopy," Opt. Express **18**(2), 1177–1190 (2010).
3.   P. R. Griffiths and J. A. de Haseth, *Fourier Transform Infrared Spectrometry* (New York, Wiley 1986), Chap. 3.
4.   R. N. Bracewell, *The Fourier Transform and its Applications* (New York: London: McGraw-Hill 1978) Chap. 18.
5.   G. C. Walker, J. B. Jackson, W. Matthews, J. W. Bowen, J. Labaune, G. A. Mourou, J. F. Whitaker, M. Menu, and I. Hodder, "Seeing Through Walls: Sub-surface imaging at Çatalhöyük", Presented at the Raman and Infrared User Group Meeting, Barcelona, March 2012.
6.   E. Berry, R. D. Boyle, A. J. Fitzgerald, and J. W. Handley, *Computer Vision Beyond the Visible Spectrum*, (Springer 2004), Chap. 9.
7.   R. K. Galvão, S. Hadjiloucas, A. Zafiropoulos, G. C. Walker, J. W. Bowen, and R. Dudley, "Optimization of Apodization Functions in THz Transient Spectrometry," Opt. Lett. **32**(20), 3008–3010 (2007).
8.   J. T. Kindt and C. A. Schmuttenmaer, "Far-infrared dielectric properties of polar liquids probed by femtosecond pulsed terahertz spectroscopy," J. Phys. Chem. **100**(24), 10373–10379 (1996).

1. Introduction

Pulsed terahertz (THz) spectroscopy and imaging, by design and necessity often operates in the reflection geometry, resulting in signals which comprise reflections from multiple interfaces, including the surface of interest. This technique is used in a range of application areas including: medical imaging, cultural heritage and non-destructive testing. Deconvolution is a mathematical technique which, theoretically, enables the deconstruction of signals reflected from stratified materials to recreate the internal structure of those samples. For example, a reflection measurement of a liquid sample inside a fluid cell contains information about the fluid and the cell window. Correct deconvolution will extract from this signal the reflection from the fluid, providing direct information about the sample. Techniques have been described in the literature which present, or use a deconvolution method, but concentrate on technological advances to reduce the terahertz pulse width to gain enhanced resolution [1], or use commercial algorithms in conjunction with additional signal processing techniques to extract information from the signal [2]. Unlike these earlier attempts,

this paper shows that, when correctly applied, deconvolution can yield deep sub-wavelength and sub-pulse width depth resolution in terahertz images. Moreover, the resulting impulse response function provides a new imaging parameter, which is entirely sample dependent.

The deconvolution algorithm presented in this paper has been designed specifically for pulsed terahertz radiation. It innovatively uses padding in the time domain to enhance the signal to noise of the impulse response function, a process which is essential to extract this feature, and furthermore it demonstrates that the technique is not dependent on the width of the incident terahertz pulse. The standard approach to deconvolution is enhanced through filtering in the frequency domain and, when necessary, the inclusion of de-noising techniques to improve signal to noise ratio. The deconvolution technique is described through theoretical modeling, and demonstrated in a range of practical examples which indicate the range of impact of the technique, both for reflection spectroscopy, and imaging stratified material, chronologically at depth.

2. Theory

The reflected signal in THz spectroscopy, $P_o(t)$, is the convolution of the incident THz signal, $P_i(t)$, represented in the time domain, with the impulse response function, $IRF(t)$, which represents the reflected signal that would be obtained if the incident signal were an infinitesimally thin impulse. The impulse response function will therefore consist of a series of time domain signatures, obtained with infinite time resolution, corresponding to reflections from the boundaries within a stratified sample (Eq. (1)).

$$P_o(t) = P_i(t) \otimes IRF(t) \qquad (1)$$

Using the convolution theorem, Eq. (1) can be transformed into the frequency domain, where the reflected signal spectrum, $S_o(v)$, equals the incident THz signal spectrum, $S_i(v)$, multiplied by the transfer function of the sample, $T(v)$. This is expressed mathematically in Eq. (2).

$$S_o(v) = S_i(v) \times T(v) \qquad (2)$$

Deconvolution retrieves the impulse response function from the inverse Fourier transform of the transfer function, which can be expressed, via Eq. (2), as the ratio of the reflected and incident THz signal spectra (Eq. (3)).

$$IRF(t) = iFT \left(\frac{S_o(v)}{S_i(v)} \right) \qquad (3)$$

Deconvolution is a black box technique. The incident (sometimes referred to as the "reference") THz pulse is a signal collected in transmission geometry, i.e. the radiation incident on the stratified sample. By instrument design or experimental practicality it may not always be possible to directly measure this data. In this scenario a reference pulse can be recorded through reflection from a metallic reflector and the frequency dependent phase shift induced by the mirror can be corrected through application of the Drude model. While it is acceptable just to multiply by a factor of −1 to correct for the reflection induced phase shift, calculations using the Drude model indicate a frequency dependent trend in reflectance (reflectance decreases with increasing frequency) for gold mirrors, which were used exclusively in this paper. However, the assumptions made to derive the Drude model become increasingly less appropriate at higher frequencies and further work is required to determine an acceptable frequency range for this theoretical description of the reflection of electromagnetic radiation from a metallic surface.

Blindly applying Eq. (3) to THz data proves unsuccessful. A series of concepts and modifications are necessary for the technique to work. The sequence of these adaptations is as follows: a) ensure that the time domain signals are sampled with a sufficiently short sampling interval to enable the retrieval of sample features with the desired resolution; b) pad the

incident and reflected data sets in the time domain; c) denoise the data when necessary; d) employ thresholding in the frequency domain. These four key elements are illustrated in the following section.

2.1 Data sampling

As indicated by Eq. (1), the reflected time domain signal, $P_o(t)$, contains information about the internal boundaries within a stratified sample in the form of its impulse response function. Moreover, this information will be contained within $P_o(t)$ whatever the form or duration of the incident signal. The only limitation to the information contained within $P_o(t)$ is that it is representative of the optical behaviour of the sample over the range of frequencies present in the incident signal and not at any other frequency.

However, although both the incident and reflected pulses are continuous signals, all terahertz pulsed imaging and spectroscopy systems sample the reflected signal at discrete time intervals. Therefore, in order to resolve features of a given width in the impulse response function, the Nyquist sampling theorem indicates that the time domain sampling interval needs to be less than or equal to half that width. This has been verified through numerical simulation.

In the simulation, the incident signal and impulse response function were both represented by 10 000 point data series. The simulated incident pulse had positive and negative going parts and spanned an overall width of 1100 points. The simulated impulse response function consisted of two peaks, each of width 5 points, with a centre-to-centre spacing of 10 points (see Fig. 1). The incident pulse and impulse response function were convolved to give the simulated reflected signal, another 10 000 point data series. These incident and reflected signals were then sampled with different sampling intervals by omitting intervening points within each data series, and the resulting impulse response function recovered by taking the inverse Fourier transform of the ratio of their spectra, as in Eq. (3). The transformation of the signals from the time to frequency domain and the subsequent inverse transform of their ratio to retrieve the impulse response function were carried out using complex fast Fourier transforms. In Fig. 1, the resulting recovered impulse response functions corresponding to sampling every second point (1:2), every fifth point (1:5) and every tenth point (1:10) in the incident and reflected signals are compared with the original, "unsampled" impulse response function.

Sampling 1:2 data points results in recovery of the features in the impulse response function. Sampling 1:5 data points, which corresponds to the Nyquist limit for the impulse response function peaks spaced at 10 data points in the original data series, just resolves those two peaks. Sampling 1:10 data points does not recover the two peak feature in the impulse response function.

This simulation has verified that it is the relative size of the sampling interval to the feature to be reconstructed that determines the theoretical resolution limit of the technique, not the pulse width. Other authors have spent considerable effort on trying to reduce the pulse width to improve the depth resolution, however, this is unnecessary if the techniques in this paper are used. These results were repeated with a variety of pulse widths and shapes, including a sine wave, and with a variety of impulse response functions. It is worth noting that in the case of a periodic incident waveform, the reconstruction technique works equally well for wavelengths both shorter and longer than the distance between the peaks in the impulse response function.

2.2 Time domain padding

It is well known that padding the time domain signal with zeroes prior to fast Fourier transformation, i.e. extending the time domain waveform by stringing additional zero amplitude points onto the end of the recorded data series, produces a similar increase in the number of points in the frequency domain, but with the new points interpolated between the existing spectral data points. The spectral bandwidth remains unchanged through this process, but the energy is spread across more spectral "bins". While this process is sometimes used to

produce a spectrum that appears more pleasing to the eye through interpolation, it can be shown that the additional points introduced into the spectrum contain no extra spectral information above and beyond that contained in the spectrum obtained without padding [3,4]. In the context of deconvolution, however, padding serves a different purpose: it improves the signal-to-noise ratio in the recovered impulse response function.
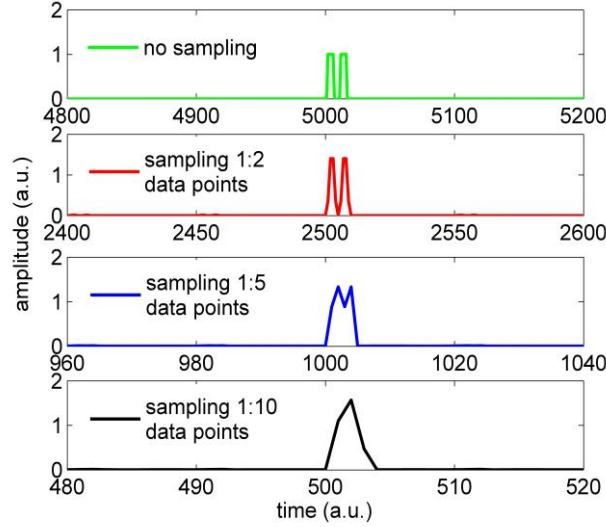

Fig. 1. Illustration of the Nyquist sampling theorem in the recovery of a two peak feature in the impulse response function following deconvolution.

In considering the role that padding plays in the deconvolution process, it is helpful to firstly set out the equations for the discrete Fourier transform and associated inverse. One convention for the discrete Fourier transform as computed by some fast Fourier transform algorithms is

$$X_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x_n \cdot e^{i\,2\pi\frac{k}{N}n} \tag{4}$$

where X_k is the transformed data array (complex), N is the number of points in the data series, x_n is the initial data series and k is related to frequency, the quantity k/N being analogous to frequency measured in cycles per sampling interval.

The corresponding inverse transform is:

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k \cdot e^{-i\,2\pi\frac{n}{N}k} \tag{5}$$

Now, if the number of points in the time domain data series is increased by a factor p through padding with zeroes, each of the resulting pN points in the frequency domain will be computed from the summation in Eq. (4), which now extends over the pN time domain points and will have its amplitude modified by the pre-multiplying factor which becomes $(pN)^{-1/2}$. Therefore, if both the incident and reflected time domain signals are padded to the same degree, their corresponding spectra will accrue the same number of bins in the frequency domain. Equally, the factor $p^{-1/2}$ that reduces the amplitude is the same for each data series and therefore the transfer function (see Eq. (2)), although distributed over more bins, has the same amplitude irrespective of the amount of padding that has been applied.

The contribution at each frequency in the transfer function can be considered to be the sum of a signal component S_k and a noise component E_k:

$$X_k = S_k \pm E_k \tag{6}$$

Upon performing an inverse fast Fourier transform of the transfer function to retrieve the impulse response function (Eq. (3)), each point in the impulse response function can also be written in terms of a signal component s_n and a noise component e_n:

$$x_n = s_n \pm e_n = \frac{1}{\sqrt{pN}}\left(s_n' \pm e_n'\right) \tag{7}$$

Each point in the impulse response function is, in essence, the result of a summation over all pN component frequencies, each with an appropriate phase factor (Eq. (5), with N replaced by pN). The phase factors in Eq. (5) adjust with p in order to ensure that the phase evolves with frequency at a rate which keeps step with the number of points in the interpolated transfer function, maintaining the overall shape of the impulse response function regardless of the padding factor that has been used. As the padding spreads the transfer function over p times more bins in the frequency domain, but without altering its shape or overall amplitude, the signal contribution to the impulse response function from the summation part of Eq. (5) will be directly proportional to the number of bins:

$$s_n' \propto pN \tag{8}$$

However, as the noise is random, it has to be added in quadrature, i.e.

$$e_n' \propto \left(E_0^2 + E_1^2 + E_2^2 + \ldots E_{pN-1}^2\right)^{1/2} \tag{9}$$

Now, if all the noise components are of equal amplitude but random phase (i.e. white noise), this becomes

$$e_n' \propto \sqrt{pN} \cdot E_k \tag{10}$$

Thus, taking account of the $(pN)^{-1/2}$ factor in the inverse Fourier transform, the total contribution from the signal to the impulse response function is seen to be proportional to the square root of the number of bins:

$$s_n \propto \frac{1}{\sqrt{pN}} \cdot s_n' \propto \sqrt{pN} \tag{11}$$

whereas the noise contribution becomes independent of the number of bins:

$$e_n \propto \frac{1}{\sqrt{pN}} \cdot e_n' \propto E_k \tag{12}$$

Therefore, as a result of padding, the signal is expected to increase as the square root of the padding factor p while the noise stays more or less constant. This behaviour has been observed by extending the simulation discussed in Section 2.1, adding normally distributed noise to the sampled signals and including padding. The effect of different padding factors on the recovered impulse response function is illustrated using the simulation in Fig. 2.
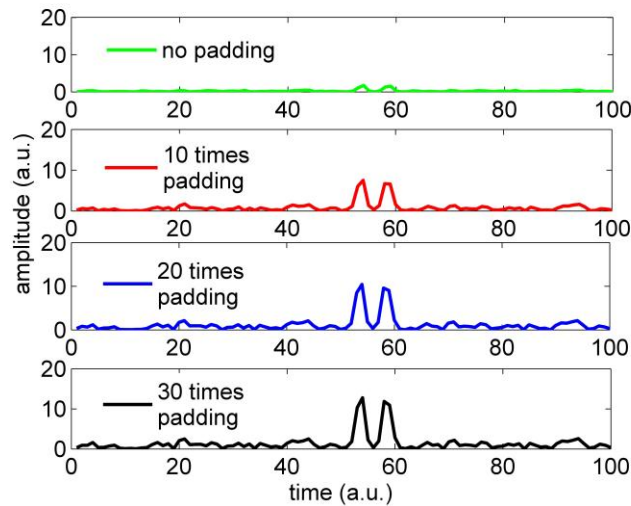
Fig. 2. Illustration of the effect of padding on the signal to noise ratio of the recovered impulse response function following deconvolution.

The signal level shows the expected $p^{1/2}$ increase with padding factor. However, rather than staying constant, the noise also shows a slight increase with padding factor. This small discrepancy may be due to the fact that not all of the noise terms in the transfer function are strictly independent as has been assumed in the simple treatment above, as a consequence of the interpolation process that is inherent in transforming the padded data series to the frequency domain. Indeed, in real data the noise is unlikely to be distributed equally across all frequencies and will tend to be system dependent, so the exact behaviour of the noise component may show some deviation from that predicted above. Nevertheless, both the simulation and real data show a significant enhancement of the signal-to-noise ratio of the recovered impulse response function through the use of padding.

It should be noted that some fast Fourier transform routines use alternative conventions for the discrete Fourier transform and its inverse, instead of those given in Eq. (4) and Eq. (5). In one in common use, the factors of $N^{1/2}$ that appear in front of both transforms are replaced with a factor of $1/N$ in the forward transform and factor of 1 in the inverse transform. This is usually accompanied by a reversal in sign of the exponents. Following the same arguments as above, in this case padding would be expected to result in an increase in the signal level in the recovered impulse response function by a factor of p, while the noise should increase by $p^{1/2}$. However, the predicted increase in the signal-to-noise ratio remains the same, regardless of the Fourier transform convention used.

2.3 Denoising

Signal denoising is not always necessary, but the authors have found, particularly on data reflected from highly absorbing/scattering stratified samples, that it makes reconstruction of internal structure from even poor signals possible. None of the results presented in this paper required the use of denoising, however, the authors have used this technique to identify and locate the presence of sub-surface paint layers in inhomogeneous plaster [5]. Due to the low signal to noise ratio in the recorded data the following denoising routine was used to enhance the quality of the results. Denoising routines are applied after padding, using motherfunctions from the Daubechies family (04 and 09) for the wavelet denoising of incident and reflected signals respectively, however the interested reader can fine more detail in the following references [2,6]. Following denoising, apodisation of the signal is necessary to prevent signal ringing [7], a tapered cosine apodisation function having been found to work well.

2.4 Thresholding

Thresholding in the frequency domain is a necessary consequence of Eq. (3). The division of noise by noise in the higher frequency region of the signal results in division by very small numbers, causing spikes in the higher frequency region of the transfer function which dominate the signal of interest (the channel spectrum) in the lower frequency part of the signal. If not removed, these spikes are distributed across the impulse response function following application of the inverse fast Fourier transform, and dominate the information of interest. Locations within the incident data where the signal falls below a set percentage of the maximum value (5% was used to achieve the experimental results presented [8]) are identified and set to zero in the reflected data set. This suppresses dominant spikes in noisy regions of the data (Eq. (3)) and results in the calculation of the impulse response function of the stratified sample of interest.

2.5 Summary

A methodology has been identified for the effective deconvolution of THz reflection data which includes the following sequence of tasks. This methodology is presented schematically in Fig. 3.
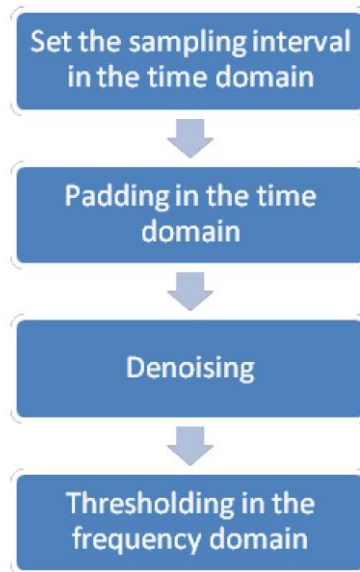

Fig. 3. Schematic diagram illustrating the deconvolution process.

3. Experimental results

All experimental results were recorded on a Picometrix T-Ray 4000 system. Reference signals, to be used as incident signal data, were collected from reflections from a metallic surface, and the Drude model applied to correct for the induced phase change between the incident and reflected radiation. All reflection measurements were recorded at normal incidence. A 320 ps measurement window was used and 4096 data points collected with a sampling interval of 0.078 ps. This corresponds to a theoretical resolution limit of 13 μm.

For all data analysis using the proposed deconvolution routine a padding factor of 20 was used with a 5% thresholding percentage. Padding below a factor of 20 compromised results by reducing the benefits of the enhanced signal to noise as a result of padding (as indicated in Fig. 2), while a factor greater than 20 significantly increased computation time for only a

marginal increase in signal to noise of the impulse response function (again deducible from Fig. 2).

3.1 Reflection spectroscopy

Reflection data were recorded from two sheets of modern papyrus, 150 µm thick, spaced 2 mm apart. The reflected signal can be seen in Fig. 4(a), the two peaks representing reflections from the first and second layer of papyrus respectively. The individual contributions from the front and back of each sheet of papyrus cannot be distinguished from the overall reflected radiation in this signal. The absolute value of the impulse response function calculated from the deconvolution algorithm is shown in Fig. 4(b). The impulse response function indicates four peaks, corresponding to reflection interfaces in the sample: the front and back of each of the papyrus sheets. These peaks were isolated from the data set, by setting all data to zero except the peak in question, for each of the four main peaks, and each were individually reconvolved with the incident data set. The sum of these four individual signals matches the original measured signal. Furthermore, the four individual signals show all the hallmarks of reflections from each of the papyrus surfaces: for example, a 180° phase inversion for reflections from a low to high refractive index boundary. These results are seen in Fig. 5, while Fig. 6 shows the summation of these four signals compared to the measured reflection data, the slight discrepancy in peaks three and four of these two signals is due to the ambiguity of the selection of the end of the second deconvolved peak in Fig. 4(b) and the start of the third.
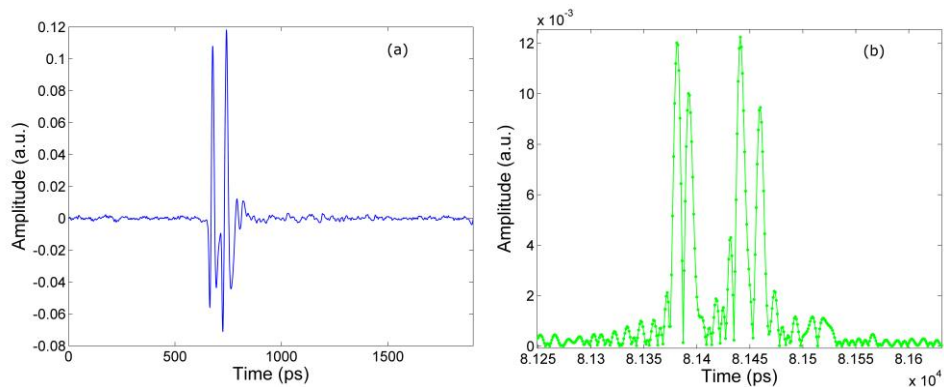

Fig. 4. (a) shows a reflection signal from two sheets of modern papyrus, each 150 µm thick, separated by 2 mm. Figure 4(b) shows the resulting impulse response function calculated using the deconvolution routine.

The resolution limit of the technique was tested by recording reflections from a series of optically transparent plastic sheets. The thinnest of these was 57 ± 0.1 µm, (measured by a Mitutoyo Litematic) and was easily resolved by the technique. The measured time domain trace is shown in Fig. 7(a) and the deconvolved signal in Fig. 7(b).
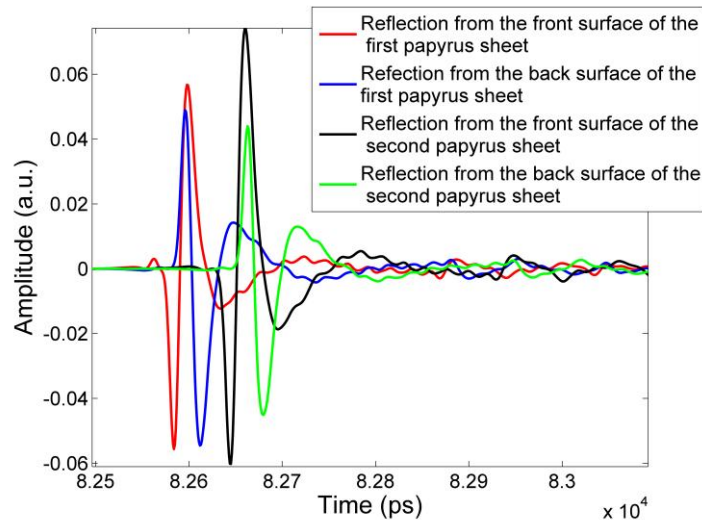
Fig. 5. shows reconvolved signals representing reflections from the front and back surfaces of two sheets of papyrus separated by 2 mm. These add to match the measured signal reflected from the two sheets.
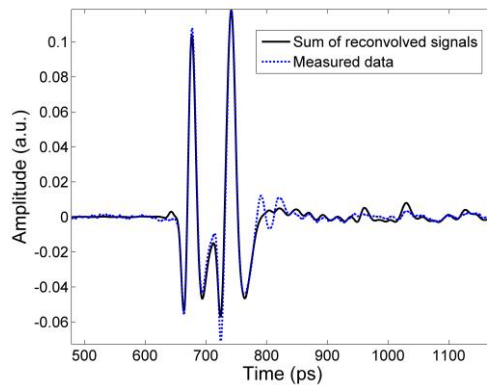

Fig. 6. shows the sum of the reconvolved signals in Fig. 5 compared to the measured signal reflected from two sheets of papyrus.

3.2 Reflection imaging

While the deconvolution technique presented in this paper will prove a powerful tool for the analysis of reflection spectroscopy data for a variety of applications, including the analysis of biological spectroscopy data, it also has potential to open up new application areas in THz reflection imaging. While the reconvolved signals, independently representing surfaces that were not resolved in the measured reflection signal, can be used in their own right through traditional THz imaging parameters to form images, there is a compromise in signal to noise. The impulse response function, however, proves an interesting source of imaging parameters.

Each peak of the impulse response function corresponds to a reflection from a boundary or refractive index mismatch within the sample. The amplitude of the peak is an indication of the reflection coefficient of this boundary, although second and subsequent peaks are adjusted in light of absorption/reflection at previous surfaces. This can be corrected for in post-processing, using a simple addition-based algorithm.

Reflection data were recorded from a stack of three pieces of paper, each 100 μm
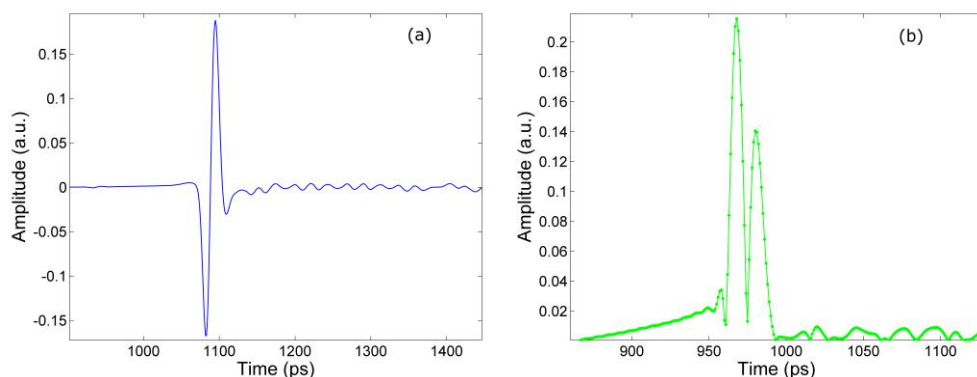
Fig. 7. (a) Time domain pulse measured following reflection from a 57 μm plastic sheet.
Figure 7 (b) Impulse response function for a 57 μm plastic sheet.

thick, with a unique symbol drawn on each side of each sheet. The ink used was created from a 2:3 ratio of carbon black to gum Arabic. Figure 8 shows a typical signal from the data set, while Fig. 9 shows the THz images of the front and back of the first layer of paper generated using the maximum amplitude of the first and second peaks of the impulse response function, compared to photographs of the same.
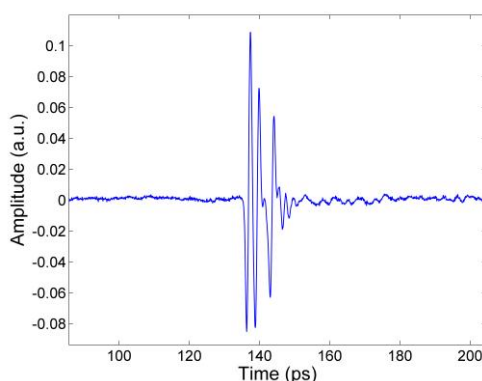

Fig. 8. shows a typical data series from the reflection image from a stack of three sheets of paper, each 100 μm thick.

It is evident from Fig. 8 than more information is contained within the recorded signal about the further obscured pages. Regardless of this, a clear image of the reverse side of the top page was achieved, without indication of the images beneath, using the proposed deconvolution routine.

4. Discussion

A robust deconvolution algorithm has been described through theoretical simulation for pulsed THz radiation. It is achieved through an innovative use of data padding that the authors have never seen demonstrated elsewhere. The potential of the algorithm for reflection spectroscopy and imaging is presented in this paper. This is an important tool for a number of THz applications which by design, or necessity, use reflection spectroscopy. The reflection signal of interest from within a stratified sample can now be identified by simply identifying the correct peak in the impulse response function, and recreated by reconvolving this peak with the incident signal. Furthermore, the impulse response function provides a powerful imaging parameter, allowing the reconstruction of sub-surface images by using the peak representing the layer of interest. This gives the potential to move, layer-by-layer,

chronologically, through a stratified sample using an imaging parameter based on the peak of the impulse response function features, which theoretically (from Eq. (3)) contain only information about the stratified sample and the refractive index contrast at boundaries within it.
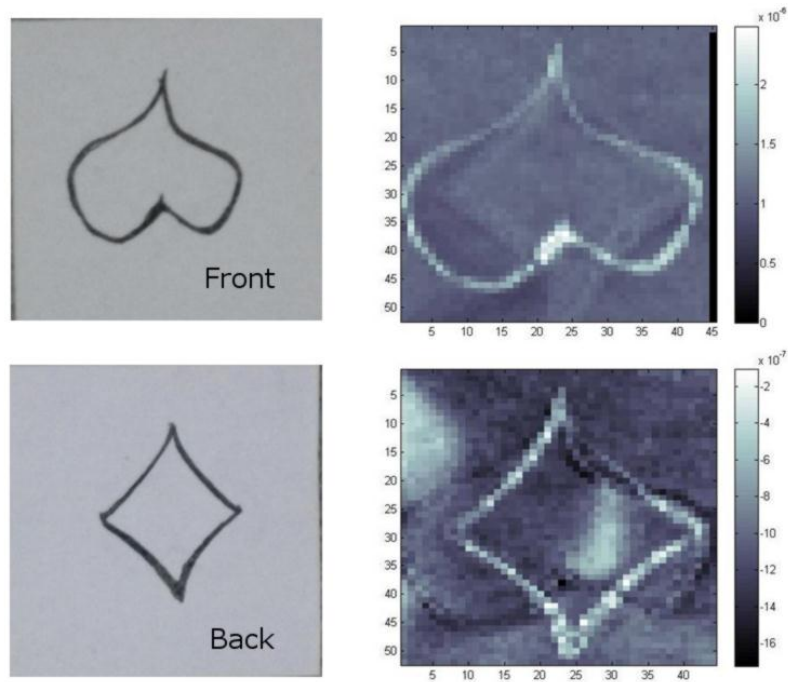

Fig. 9. shows THz images of the front and back of the first sheet of paper.

Other image parameters however become available to the user as a result of the technique. As demonstrated in Fig. 4, it is possible to reconstruct reflections from layers that are not resolvable using the unprocessed time domain measurement. Standard THz imaging parameters in both the time and frequency domain can be used to construct images from this new data. There are however two main drawbacks to this approach. Firstly, the use of padding in the deconvolution routine significantly improves the signal to noise ratio of the impulse response function. Any subsequent reconvolution reintroduces noise from the original data and reduces this advantage. Secondly the image parameters derived from the time and frequency domain of the reconvolved data intrinsically also contain pulse and system dependent information removed from the impulse response function.

5. Conclusion

A deconvolution algorithm has been presented, which has been specifically designed for THz spectroscopy and imaging and incorporates a series of novel features. The resolution of the technique is not a feature of the pulse shape or width used, but rather a consequence of the sampling period. Padding the time domain data series by a factor p as part of the algorithm increases the signal of the impulse response function by a factor of $p^{1/2}$ while the noise remains more or less constant. Thresholding the signals in the frequency domain, before calculating the transfer function reduces high frequency noise spikes which dominate the lower frequency channel spectrum following inverse Fourier transformation. For a 0.078 ps time step a theoretical resolution of 13 μm was calculated and in practice a 57 ± 0.1 μm sheet of transparent plastic was resolved.

Finally, the use of the peaks of the impulse response function to recreate a sub-surface image was demonstrated by imaging the reverse side of a 100 μm sheet of paper. Work continues with the aim of reconstructing images from specific sides of a sheaf of pages.

Acknowledgments