Terahertz fingerprinting in presence of quasi-ballistic scattering

Mayank Kaushik,[1,a)] Brian W.-H. Ng,[1] Bernd M. Fischer,[1,2] and Derek Abbott[1]
[1]*Centre for Biomedical Engineering (CBME) and School of Electrical and Electronic Engineering, The University of Adelaide, SA 5005, Australia*
[2]*Institut Franco-Allemand de Recherches de Saint Louis, 68301 Saint Louis Cedex, France*

(Received 8 March 2012; accepted 27 July 2012; published online 8 August 2012)

Recent years have seen significant advances in material diagnostics and analysis using terahertz (THz) time domain spectroscopy (TDS) and imaging. Despite its widespread application, the interaction between THz radiation and materials with random structure is not yet fully studied. Separation of absorption and scattering is required to extract the true absorption spectra, thus enabling direct comparison with pure samples in a spectral data base for automated recognition. Here, we present a discrete wavelet transform based iterative reconstruction technique that reduces the scattering contribution in THz-TDS measurements, in composites with absorbing constituents that exhibit sharp absorption features. [http://dx.doi.org/10.1063/1.4745182]

Over the past decade, terahertz spectroscopy and imaging have become popular and important tools for a variety of applications such as pharmaceutical material characterization, food quality control, investigation of explosive materials, and biomedical sensing.[1–4] Many materials exhibit characteristic molecular vibrational and rotational modes in the terahertz spectral range (0.3–3 THz), enabling them to absorb terahertz (THz) radiation at specific frequencies. These absorption features are unique to every material that absorbs in the THz spectral range and therefore may be used as spectral fingerprints for their classification and identification.[5] However, variations in refractive index within the sample, caused by sample granularity and impurities, cause the THz radiation to scatter, which can significantly alter or obscure the spectral fingerprints of the material under study.

Usually in the case of the solids, the material of interest is quite dense and causes multiple scattering of THz radiation within the sample. The response of a dense medium, as a consequence of multiple scattering, can be classified into three regimes: ballistic, quasi-ballistic, and diffusive transport.[6] While the THz time domain spectroscopy (TDS) technique is sensitive to both quasi-ballistic and diffusive scattering, the criteria to determine which scattering regime is dominant, depends on the scattering (λ_{sc}) and transport mean free path lengths (λ_{tr}) in the medium, given by,

$$\begin{aligned} \lambda_{sc} &= c/2n_{\mathrm{i}}\omega, \\ \lambda_{tr} &= \lambda_{sc}/(1-\langle\cos(\theta)\rangle, \end{aligned} \tag{1}$$

where n_{i} is imaginary part of the complex refractive index of the medium and $\langle\cos(\theta)\rangle$ is the average of cosine of the scattering angle.[8,9] The diffusive component is only significant and needs consideration when the medium thickness l is greater than 7 to 10 times the transport mean free path length, λ_{tr}.[9–11] Generally in transmission mode measurements, the setup collects signal only in the forward narrow cone. The contribution of the diffusively scattered signal to the measured signal depends on the grain size, concentration, and thickness of the sample[7] and can be determined using the equation $\rho = \rho_B + (d\Omega/4\pi)p_D$, where ρ represents the total power captured by the detector, ρ_B represents the sum of contributions from the ballistic and the quasi-ballistic transport and ρ_D represents the diffusive component of the transmitted signal. In our measurements, the setup collects signal only in forward narrow cone ($d\Omega \approx 6^\circ$ off the normal axis). For such narrow angles the transport mean free path lengths, given by Eq. (1), are several orders of magnitude larger than the scattering mean free path. Under such conditions, given that the medium thickness, l is smaller than 7 to 10 times the transport mean free path length, λ_{tr}, only the information carrying ballistic and quasiballistic photons are collected by the detector and the diffuse photons that are the source of noise are mostly rejected.[12] It must be noted that, in principle, the sample thickness l can be extended in which the quasi-ballistic light can be measured by reducing the angle $d\Omega$.[13]

As the detected radiation suffers from both absorption and scattering, it becomes essential to separate scattering from the detected response in order to reveal the true spectral fingerprint of the material under study. Several attempts have been made in order to reduce, mitigate and separate the effects of scattering from the detected signal.[14–17,19,20] Most of these techniques either require special sample preparation or custom measurement apparatus. Recently, Arbab *et al.*[17] showed the implementation of a wavelet based technique to retrieve the true THz spectroscopic signature in presence of surface scattering using a reflection mode THz-TDS setup. They carried out a multiresolution analysis using the maximum overlap discrete wavelet transform (MODWT) of the measured extinction spectrum of lactose samples. Based on visual inspection of the results, they argued that the resonant feature of the lactose sample (at 0.54 THz) can be extracted from the details subband at certain levels of the decomposition. The selection was performed manually by researchers based on visual, instead of, physical reasons. They did not propose any reconstruction technique following their decomposition, and directly relied on the details subband for the scattering invariant material identification. This makes their technique unsuitable for automated material identification and classification in real-world applications.

a)mayank@eleceng.adelaide.edu.au.

In this letter, we show that a simple reconstruction based only on the details subband can lead to loss or misinterpretation of significant spectral features in the extinction spectrum. Thereafter, we propose an iterative reconstruction technique using only the details subband to estimate the scattering baseline. This can be then subtracted from the original extinction spectrum to obtain a scattering reduced extinction spectrum, and avoid issues involved in the previous approach. The proposed method assists in enabling direct comparison with spectra of pure samples in a spectral data base for automated recognition. It must be noted that the proposed technique and the technique proposed by Arbab *et al.*[17] are only applicable for materials that exhibit sharp and narrow absorption features (such as α-lactose monohydrate, α-D-glucose, sucrose) and not to those that exhibit slowly varying and very broad absorption features (such as morphine). However, we argue that the proposed technique is useful as there is a broad range of materials that exhibit sharp and narrow absorption features in the THz range.

We selected two materials namely, α-monohydrate lactose and granular glucose, for this study. Both the materials are common ingredients in many pharmaceutical tablets and food products and show distinctive and sharp absorption features in the THz spectral range, making them suitable for this study. Two sample pellets were made by mixing α-monohydrate lactose and granular glucose with PE powder in a ratio of 1:1 (material:PE) and pressing them using a hydraulic press at $10\,\text{tons}/\text{cm}^2$. Note that α-monohydrate lactose has characteristic absorption fingerprint frequencies at 0.54, 1.2, and 1.39 THz, while glucose has its characteristic absorption fingerprint frequencies at 1.25, 1.89, 2.39, and 2.57 THz.

The extinction spectra of such materials are always represented as a discrete data series in the Fourier domain. We carry out a multilevel discrete wavelet decomposition (DWT) of this data series (the extinction spectrum) in order to represent it in a set of localized contributions (details and approximations). At each level, the contributions are identified by a scale and position parameter and represent the information of the different frequency contents in the original data series. A given spectrum $X(f)$ can be represented by a set of approximation $c_{j,k}$ and details $d_{j,k}$ coefficients, which correspond to a wavelet series expansion of $X(f)$,

$$X(f) = \sum_{j=0}^{J} \sum_{k=0}^{N_j} d_{j,k}\psi_{j,k}(f) + \sum_{k=0}^{N_J} c_{J,k}\phi_{J,k}(f), \qquad (2)$$

where $\phi(f)$ and $\psi(f)$ are the chosen pair of scaling and wavelet functions, respectively. For discrete cases, the conventional technique it to consider $X(k)$ as the finest level approximation subband and the subsequent decomposition into coarser approximation and detail coefficients is known as the discrete wavelet transform (DWT).

In our implementation, Mallat's multiresolution analysis[21] is used to decompose the discretized measured extinction spectrum $X(k)$ into several sets of approximations and details coefficients, cA and cD, respectively. This process is depicted in Fig. 1(a). First, the data series is decomposed into cD_1 and cA_1. The sequence cD_1 is associated with the high frequency component of the data series, while, cA_1 is associated with the low frequency components. Subsequently, cA_1 becomes the input for the next level and the process is repeated until the desired level J is reached. For materials exhibiting sharp and sparse absorption features, the extinction spectrum consists of a smooth, slowly varying part (scattering contribution), some sharp, high order polynomials (absorption peaks) and other background contributions such as systematic errors due to laser drift, etc.

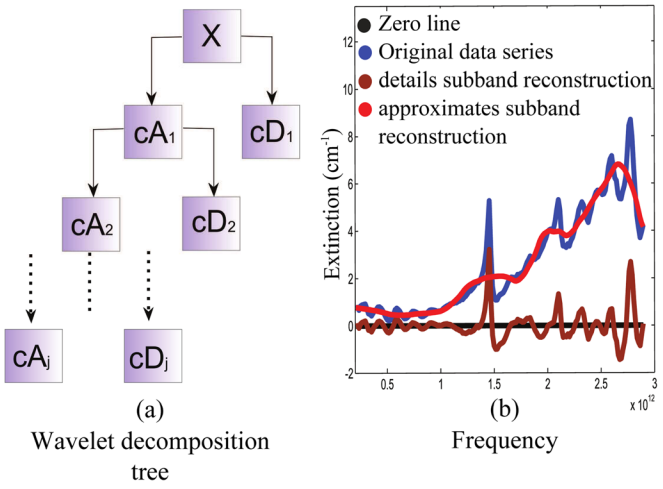

FIG. 1. (a) Wavelet decomposition tree: Signal X id decomposed into several sets of approximations and details coefficients, cA and cD respectively. The process is repeated by decomposing the approximate coefficients cA at each level upto the desired level. (b) Extinction spectrum of glucose (blue) with reconstructions based on approximations only (red) and details only (brown).

Generally, scattering attenuation can be expressed as a linear or quadratic function over the given frequency range, depending on the medium of propagation.[18,19,22–24] Using the DWT decomposition we want to estimate and thereafter separate the smooth, slowly varying component (scattering contribution) from the original data series (extinction spectrum). To do so, we have selected the "db4" wavelet of the Daubechies family as the mother wavelet for our decomposition process. Here, "db4" is an orthogonal wavelet and has 8 taps and 4 vanishing moments, which guarantees removal of most of the scattering contributions, as the scaling functions, or just the approximation coefficients alone, can perfectly represent all polynomials up to order 3.

Fig. 1(b) illustrates a wavelet based decomposition example. Here, we consider the extinction spectrum of a α-D-glucose and PE pellet obtained using a standard transmission mode THz-TDS setup. After the multilevel DWT based decomposition, the data series is reconstructed first from the details only and then from approximations only. It can be clearly seen in Fig. 1(b) that the reconstruction carried out from the details subband has negative values at various frequencies, which clearly is a physically impossible artifact and can cause loss or misinterpretation of the data when used for automated classification and identification. The same effect can be seen in the approximations only reconstruction, as it overdetermined the spectrum background at various frequencies, especially around the absorption peaks. In order to eliminate these errors and to obtain the true absorption fingerprints of the material under study, we propose the following iterative reconstruction scheme:

1. **Decompose and selective reconstruction**: Using DWT, first decompose the original data series E (experimentally obtained extinction spectrum) into details and approximations and then reconstruct the data series from details only (D).
2. Modify, subtract and iterate,

$$\begin{aligned} D_i'(n) &= \max\{D_i(n), 0\} \quad \forall n, \\ E_1'(n) &= E(n) - D_1'(n), \\ E_i'(n) &= E_{i-1}'(n) - D_i'(n) \quad \forall i > 1, \end{aligned} \tag{3}$$

where i represents the iteration number. This process is illustrated in Fig. 2.

3. **Convergence condition**: As we want to avoid any data, in the details only reconstructed data series D from going negative, we propose the following convergence criteria: Let $N_i(m)$ represent all the negative data values in $D_i(n)$, then

$$V_i = \sum |N_i(m)|^2 \quad \forall m, \tag{4}$$

represents the *error energy* and convergence is met when the *error energy* becomes less than 0.01% of the overall extinction spectrum energy. This can be expressed as

$$V_i < 0.0001 \times \sum |E(n)|^2 \quad \forall n. \tag{5}$$

This process is illustrated in Fig. 3. It can be seen from Fig. 3(d) that at the optimum number of iterations, the input data series E', obtained by subtracting the D' from E', of the previous iteration, closely matches the slowly varying spectrum background and hence serves as a reasonable estimate of the frequency dependent scattering contribution in the measured spectrum.

In order to test our algorithm, we carry out THz-TDS measurements of four samples: (i) granular α-D-glucose + PE pellet (thickness $l = 0.18$ cm), (ii) coarsely ground α-D-glucose + PE pellet (thickness $l = 0.12$ cm), (iii) α-lactose monohydrate + PE (particle diameter 200−400 μm) pellet (thickness $l = 0.45$ cm), and (iv) α-lactose monohydrate + PE (particle diameter 60−80 μm) pellet (thickness $l = 0.45$ cm). Using the measured complex refractive index and assuming a forward narrow cone of 6°, we calculated the transport mean free path lengths for all the four samples according to the Eq. (1) and found them to be several orders higher than the sample thickness. Thus confirming that the measurements contain no or negligible amount of contribution from the diffusive scattered radiation. We apply the above described algorithm to the extinction spectra of these four samples to obtain an estimate of frequency dependent scattering baseline for each sample. Thereafter, to reduce the effects of scattering from the measurements, we simply subtract the estimated scattering data series from the measurements. The results are shown in Fig. 4. In order to validate our results, we compare the scattering eliminated spectrum with a scattering free/limited measurement for each sample. A completely scattering free extinction can be obtained by THz-TDS of a single crystal of the given material. However, due to the very small size of a crystal, this measurement is not possible for a number of materials. One may attempt to grow a bigger crystal from many small crystals by dissolving in an appropriate solvent, but this process can take days and still there is a risk of introducing impurities to the crystal. Thus in the absence of a true scattering free measurement, one can resort to a measurement of a sample prepared by milling the material into a very fine powder and then pressing it into a pellet of very high density (volume fraction ≥90%). This sample preparation technique ensures very

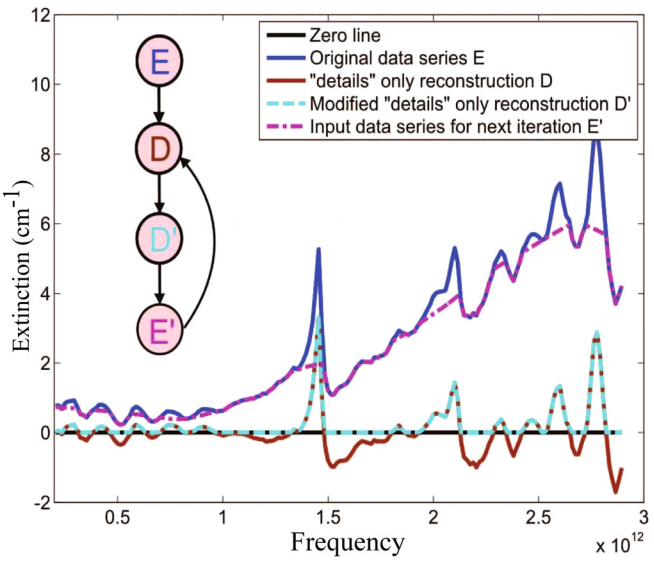

FIG. 2. Iteration scheme: Original data series (blue) with reconstruction based on details only (brown), and modified details (dotted cyan) and input data for next iteration.

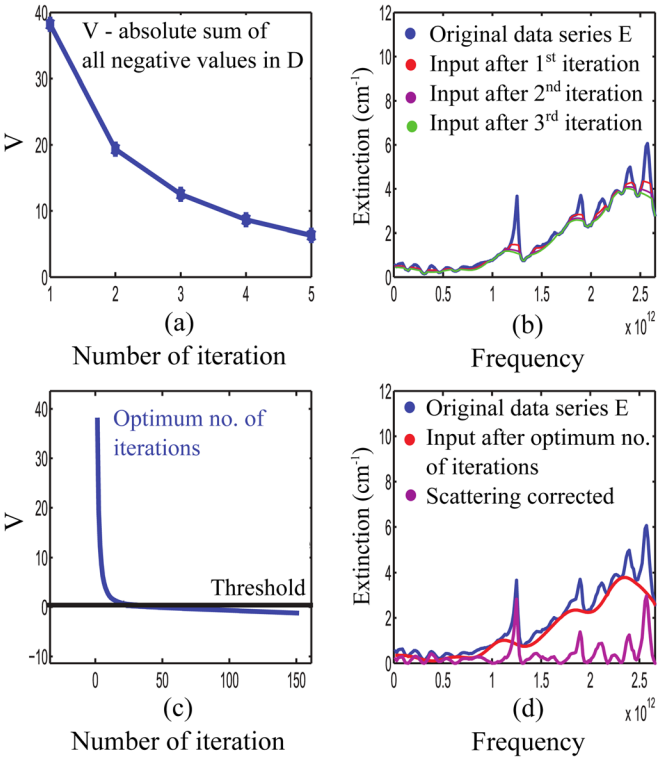

FIG. 3. (a) Value of V for first four iterations, (b) input data series for first four iterations, (c) convergence condition, and (d) scattering corrected spectrum after optimum number of iterations.

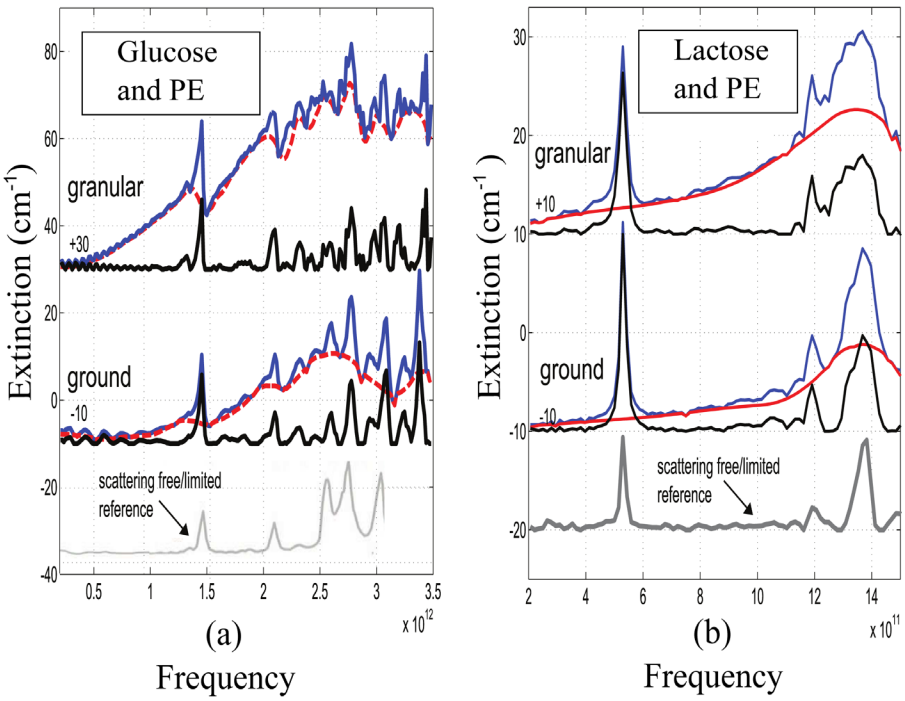

FIG. 4. Scattering baseline estimation and reduction—(a) for glucose and PE sample and (b) for lactose and PE sample, with measured spectrum in blue, scattering baseline estimate in red, scattering reduced spectrum in black, scattering free/limited reference for glucose (after Walter *et al.*[25] scaled to 1/45) and scattering free/limited reference for lactose (scaled to 1/3) in grey.

small scatterer (generally the air-voids) size and volume fraction, thus limiting the contribution of scattering in the measured extinction spectrum. Walter *et al.*[25] carried out such a measurement for α-D-glucose, and here we have used their measurement as the scattering free/limited reference for comparison with our results. For the scattering free/limited measurement of the lactose sample, we carry out THz-TDS of a sample prepared by pressing (pressure $\approx 10\,\text{ton/cm}^2$) a mixture (1:1) of very fine α-lactose monohydrate (particle size $\approx 35\,\mu$m) and very fine PE powder (particle size $\approx 30-40\,\mu$m), to form a dense (volume fraction $>90\%$) pellet. It must be noted that the measurement obtained from such samples is not completely scattering free.

It must be noted that the proposed iterative reconstruction technique provides with separate estimates of scattering and absorption spectra for the sample under study as opposed to the method proposed by Arbab *et al.*[17] For the four samples, considered in this letter, convergence was achieved after 12 iterations for the granular α-D-glucose-PE sample, 11 iterations for the coarsely ground α-D-glucose-PE sample, with 4-level DWT, and 15 iterations for the α-lactose monohydrate-PE (diameter $60-80\,\mu$m) sample, and 20 iterations for the α-lactose monohydrate-PE (diameter $200-400\,\mu$m) sample with 5-level DWT.

In conclusion, we present an iterative multilevel DWT based technique for estimating frequency dependent scattering baseline for transmission mode THZ-TDS measurements. The method is tested on four sample pellets, two comprising α-lactose monohydrate and PE (with different granularity), while the other two comprising α-D-glucose (with different granularity) and PE. From the comparison of the scattering reduced spectrum with the scattering free/limited spectrum obtained from well prepared samples, shown in Fig. 4, it is clear that the technique reasonably estimates most of the general scattering attenuation profile. The proposed method helps in cleaning the measurements for scattering and other background contributions such as systematic errors due to laser drift to enable direct comparison with spectra of pure samples in a spectral data base for automated recognition. It must be noted that the technique requires the material under study to exhibit sharp absorption features and uses no other *a priori* information of the physical characteristics of the sample or the material, which makes it potentially useful for material analysis in real-world applications such as in stand-off measurements and quality control.

[1]C. Jansen, S. Wietzke, O. Peters, M. Scheller, N. Vieweg, M. Salhi, N. Krumbholz, C. Jördens, T. Hochrein, and M. Koch, Appl. Opt. **49**, 48–57 (2010).
[2]P. F. Taday, I. V. Bradley, D. D. Arnone, and M. Pepper, J. Pharm. Sci. **92**, 831–838 (2003).
[3]E. M. Pickwell, S. Huang, K. W. C. Kan, Y. Sun, and Y. T. Zhang, J. Innovative Opt. Health Sci. **1**, 29–44 (2008).
[4]B. Fischer, M. Hoffmann, H. Helm, R. Wilk, F. Rutz, T. Kleine-Ostmann, M. Koch, and P. Jepsen, Opt. Express **13**, 5205–5215 (2005).
[5]B. Fischer, H. Helm, and P. Jepsen, Proc. IEEE **95**, 1592–1604 (2007).
[6]J. Pearce, and D. M. Mittleman, Phys. Med. Biol. **47**, 3823 (2002).
[7]K. M. Nam, L. M. Zurk, and S. Schecklman, Prog. Electromag. Res. B **38**, 205–223 (2012).
[8]K. Schokichi and T. Hamada, Astron. Soc. Jpn. **27**, 545–552 (1975).
[9]J. Pearce and D. M. Mittleman, Opt. Lett. **26**, 2002–2004 (2001).
[10]B. B. Das, F. Liu, and R. R. Alfano, Rep. Prog. Phys. **60**, 227 (1997).
[11]K. Yoo, and R. Alfano, Opt. Lett. **15**, 320–322 (1990).
[12]S. Mujumdar, G. D. Dice, and A. Y. Elezzabi, Opt.Comm. **247**, 19–27 (2005).
[13]Y. Chen, J. Peter, W. Semmler, and R. Schulz, in *Biomedical Optics (BIOMED), OSA Technical Digest (CD), St. Petersburg, FL, 16 March 2008* (Optical Society of America, 2008), paper BSuE50.
[14]Y. C. Shen, P. F. Taday, and M. Pepper, Appl. Phys. Lett. **92**, 051103 (2008).
[15]A. Bandyopadhyay, A. Sengupta, R. Barat, D. Gary, J. Federici, M. Chen, and D. Tanner, Int. J. Infrared Millim. Waves **28**, 969–978 (2007).
[16]M. Franz, B. M. Fischer, and M. Walther, Appl. Phys. Lett. **92**, 021107 (2008).
[17]M. H. Arbab, D. P. Winebrenner, E. I. Thorsos, and A. Chen, Appl. Phys. Lett. **97**, 181903 (2010).
[18]J. Ophir, T. H. Shawker, N. F. Maklad, J. G. Miller, S. W. Flax, P. A. Narayana, and J. P. Joness, Ultrason. Imaging **6**, 349–395 (1984).

[19]M. Kaushik, B. W.-H. Ng, B. M. Fischer, and D. Abbott, IEEE Photon. Technol. Lett. **24**(2), 155–157 (2012).
[20]M. Kaushik, B. Ng, B. M. Fischer, and D. Abbott, in *Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)* (IEEE, 2011), pp. 33–36.
[21]S. G. Mallat, IEEE Trans. Pattern Anal. Mach. Intell. **11**, 674–693 (1989).
[22]P. A. Narayana and J. Ophir, Ultrason. Imaging **5**, 17–21 (1983).
[23]S. X. Bao, J. S. Liu, L. M. Rong, Z. H. Wang, and J. Guo, Spectrochim. Acta Part B **55**, 379–382 (2000).
[24]H. Martens and E. Stark, J. Pharm. Biomed. Anal. **9**, 625–635 (1991).
[25]M. Walther, B. M. Fischer, and P. U. Jepsen, Chem. Phys. **288**, 2–3 (2003).