

Terahertz imaging applied to cancer diagnosis

To cite this article: M-A Brun *et al* 2010 *Phys. Med. Biol.* **55** 4615

View the [article online](#) for updates and enhancements.

You may also like

- [The progress and perspectives of terahertz technology for diagnosis of neoplasms: a review](#)
K I Zaytsev, I N Dolganova, N V Chernomyrdin *et al.*
- [Multi-band passive detection and imaging system for concealed weapon with dual assessment method](#)
Ahmet Ünal
- [Pulsed terahertz reflection imaging of tumors in a spontaneous model of breast cancer](#)
Nagma Vohra, Tyler Bowman, Paola M Diaz *et al.*

JOIN US | ESTRO 2024
**In-Booth Talks, Demos,
& Lunch Symposium**

[Browse talk schedule >](#)

SUN NUCLEAR
A MEDTRON MEDICAL COMPANY

Terahertz imaging applied to cancer diagnosis

M-A Brun¹, F Formanek¹, A Yasuda¹, M Sekine², N Ando² and Y Eishii²

¹ Life Science Laboratory, Advanced Materials Laboratories, Sony Corporation, Tokyo Medical and Dental University, 1-5-45 Yushima Bunkyo-ku, Tokyo, 113-8510, Japan

²Department of Human Pathology, Tokyo Medical and Dental University, 1-5-45 Yushima Bunkyo-ku, Tokyo, 113-8510, Japan

E-mail: florian.formanek@jp.sony.com

Received 6 May 2010, in final form 25 June 2010

Published 30 July 2010

Online at stacks.iop.org/PMB/55/4615

Abstract

We report on terahertz (THz) time-domain spectroscopy imaging of 10 μm thick histological sections. The sections are prepared according to standard pathological procedures and deposited on a quartz window for measurements in reflection geometry. Simultaneous acquisition of visible images enables registration of THz images and thus the use of digital pathology tools to investigate the links between the underlying cellular structure and specific THz information. An analytic model taking into account the polarization of the THz beam, its incidence angle, the beam shift between the reference and sample pulses as well as multiple reflections within the sample is employed to determine the frequency-dependent complex refractive index. Spectral images are produced through segmentation of the extracted refractive index data using clustering methods. Comparisons of visible and THz images demonstrate spectral differences not only between tumor and healthy tissues but also within tumors. Further visualization using principal component analysis suggests different mechanisms as to the origin of image contrast.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

With the recent development of convenient sources and detectors, an increasing number of terahertz (THz) applications are being investigated, among which are chemical spectroscopy, materials science, biomedicine and homeland security (Chan *et al* 2007, Ferguson and Zhang 2002, Kawase 2004). Medical imaging is also an interesting and promising application (Pickwell and Wallace 2006). Currently, cancer diagnosis relies on the availability of qualified pathologists. Thus, the diagnosis workflow could benefit from the introduction of automated visualization techniques such as THz spectroscopic imaging.

First published results on cancer tissue imaging using THz pulsed radiation (Woodward *et al* 2003), later confirmed by studies on various cancer types and organs (Fitzgerald *et al* 2006), suggest that THz imaging can be used for macroscopic visualization of tumor margins in fresh tissues. However, the reliability of these measurements is limited by the difficulty to control the water content in a raw sample and to ensure appropriate contact between the sample and the imaging window. Moreover, construction of the images from arbitrarily chosen features in the time-domain spectrum prevents from interpreting the origin of the contrast. The possibility of producing segmented images from physically significant parameters such as the frequency-dependent complex refractive index was demonstrated using several millimeter thick slices of paraffin-embedded tissue samples in transmission geometry (Berry *et al* 2004, Nakajima *et al* 2007). However, as the signal collected from such samples includes contribution from the entire thickness, precise comparative investigation with much thinner stained slices used for pathological diagnosis is not feasible. Furthermore, the limited availability of raw tissue and the lack of equipment to prepare thick tissue slices hinder the development of THz imaging through large-scale studies.

To overcome these shortcomings, we image 10 μm thick tissue slices commonly used for pathology diagnosis and available from the paraffin-embedded tissue archives of any pathology department. Such samples can be stained in various ways and high-magnification microscope images reveal the representative cellular structure that actually interacts with THz radiation. To ensure constant water content and reproducible conditions, the samples are kept in water during measurement. The data obtained in reflection geometry are used to extract the complex refractive index at each sample position by means of a calculation method similar to the one described by Jepsen *et al* (2007), additionally taking into account the etalon effect within the thin sample. Segmented images are produced by applying clustering algorithms to the complex refractive index values. Principal component analysis (PCA) provides an insight into the possible physical origin of contrast. The link between THz images and the cellular arrangement is investigated using digital pathology tools (Molnar *et al* 2003), which automatically acquire visible images of high magnification and store them as digital data so that pathologists can ‘virtually’ investigate tissues slices at various magnifications.

Details on the experimental setup and data analysis procedures will be followed by a discussion on the results obtained on two different tissue samples for which THz images positively identifying cancer tissues were produced.

2. Experimental procedure

2.1. Imaging system

Our experiments are performed using a homemade THz imaging system working in reflection geometry, as depicted in figure 1. THz radiation is generated and detected using low-temperature grown gallium arsenide photoconductive antennas gated by femtosecond pulses from a Ti:sapphire laser. The THz beam is focused on a sample through a pair of off-axis parabolic mirrors, and the reflected beam is collected back to a detector. To extract a weak THz signal from the ambient noise, a 60 kHz modulated bias voltage is applied to the emitter, and the output signal measured at the detector is sent to a current pre-amplifier connected to a lock-in amplifier working at the modulation frequency. The sample is raster scanned through the focused THz beam.

The sample is mounted on a Z-cut quartz window. At each position on the sample, we acquire a time-domain spectrum including the reflection from the first air–quartz interface and the reflection from the quartz–sample interface. The former is used as a reference to remove

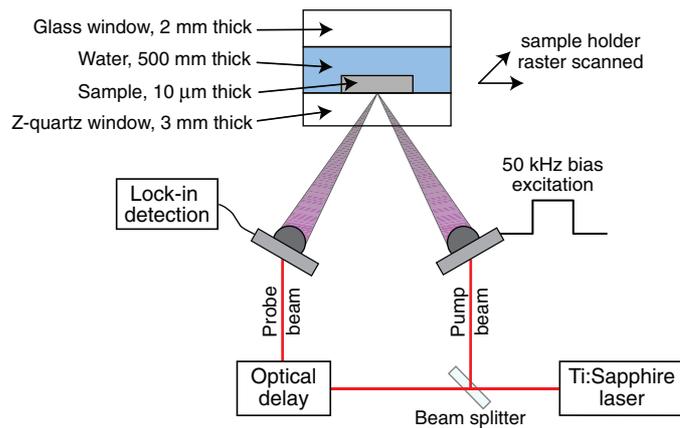


Figure 1. Schematic of the THz imaging system and the sample configuration used for the measurements of thin tissue slices.

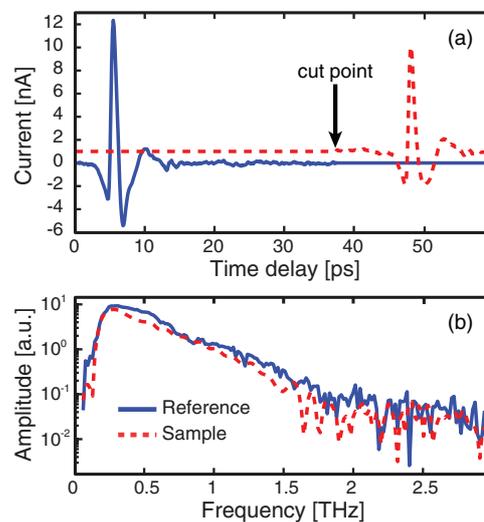


Figure 2. (a) Time-domain spectrum. The reference and sample spectra are extracted from the full spectrum by setting all values to zero after and before the cut point, respectively. The sample spectrum is shifted up for clarity. (b) Amplitude of the Fourier components obtained after FFT of the spectra above.

any system response and extract the sample response. To reduce atmospheric absorption, the THz beam path is purged with nitrogen gas to keep the relative humidity below 10%. Measurements are performed at ambient temperature. The THz beam is focused on the quartz–sample interface at an incidence angle of 20°.

Since the signals reflected from such thin samples are very weak, image filtering is applied to increase the image signal-to-noise ratio (SNR). Efficient filtering requires the pixel size to be much smaller than the expected resolution. As the beam spot size is about 500 μm at 1 THz, a typical pixel size is set to 50 μm. The system bandwidth is approximately 0.2–3 THz, yet atmospheric absorption induces image noise and prevents the use of data above 1.5 THz in practice. A time-domain spectrum is shown in figure 2(a) and the amplitude of the Fourier

components is shown in figure 2(b). In addition to the THz data, a visible picture registered to the THz-imaged area is acquired.

2.2. Samples

Anonymous paraffin-embedded tissue blocks are obtained from the archives of the Human Pathology Department of Tokyo Medical and Dental University. This work was approved by the local ethics committee. 10 μm thick tissue slices are prepared on quartz windows. The samples can be measured directly, processed back into a water solution or after specific staining (hematoxylin, silver, etc). All THz images presented hereafter are obtained by using unstained samples measured in water. Spacers and a top window (see figure 1) are used to prevent denaturation of the sample during data acquisition.

In this study, the results obtained on two different tumor samples are presented. The first sample is a part of an excised lung cancer. Pathological inspection classified the malignancy as poorly differentiated squamous cell carcinoma with necrotic parts in the center of the sample. The second sample is a part of an invasive ductal carcinoma of the pancreas. Clusters of cancer cells visible in the globally healthy areas indicate that the tumor is ill-defined. Cancer spread extends to the lymph node.

2.3. Digital slide data

Tissue slices stained with hematoxylin-eosin are prepared for microscopic inspection. Digital slide data are acquired on a MIRAX system (Zeiss). The visible image acquired simultaneously with the THz data allows registering the THz images to the digital slide data. Contrasted areas of the THz images can then be compared with high magnification microscope pictures, to visualize each corresponding cellular arrangement.

3. Analysis

3.1. Spectrum analysis

For each pixel, the reference and sample spectra are extracted from the full time-domain spectrum as shown in figure 2. Both are then Fourier transformed, and all the following analysis is carried out in the frequency domain. The complex refractive index at each pixel is obtained using a method similar to the one described by Jepsen *et al* (2007). The calculation considers the influence of the incidence angle and the polarization of the THz beam. The error induced by the beam shift between the reference and sample data is corrected using a calibration factor. In addition to Jepsen's model, a term taking into account multiple reflections within the sample is included. The successive steps are illustrated in figure 3.

The calibration factor obtained by measuring a sample of known refractive index (water) is applied on the experimental ratio between the sample and the reference electric fields at each frequency. The calibration corrects for the influence of the beam shift between the reference and sample beams as well as for possible errors in the reference window thickness or refractive index values.

The theoretical relationship between the reference and sample Fourier components of the electric field at each frequency comes from the following formula:

$$E_{\text{reference}} = E_{\text{emitted}} * (r_S^{\text{AQ}} \cos^2 \theta_0 + r_P^{\text{AQ}} \sin^2 \theta_0), \quad (1)$$

$$E_{\text{sample}} = E_{\text{emitted}} * (r_S \cos^2 \theta_0 + r_P \sin^2 \theta_0), \quad (2)$$

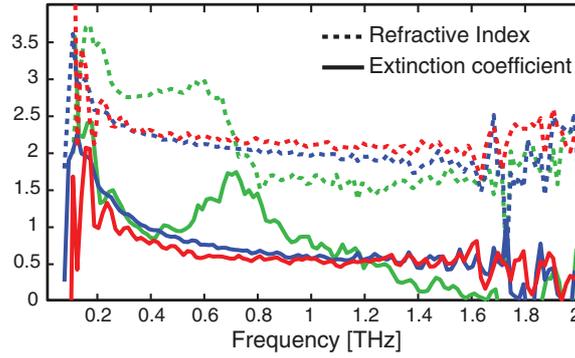


Figure 3. Complex refractive index calculation steps. The dashed lines show the real part of the complex refractive index, the solid lines show the extinction coefficient. The greens lines are obtained by neglecting the polarization dependence, the beam shift and multiple reflections within the sample. The blue lines are obtained after calibration, taking into account the effect of the beam shift. The red lines show the result of the comprehensive calculation, including polarization and etalon effects.

where r is the reflection coefficient, θ_0 is the incidence angle and E_{emitted} is the THz electric field emitted at a given frequency. The subscripts S and P refer to the polarization states, and the superscript AQ refers to the air–quartz interface. The reflection coefficient used in (2) can be written as

$$r_{S/P} = t_{S/P}^{\text{AQ}} t_{S/P}^{\text{QA}} * P * \text{FP}_{S/P}, \quad (3)$$

where t is the transmission coefficient, P is a function of the complex refractive index and thickness of the reference window taking into account propagation. FP is a function of the complex refractive index and the thickness of the sample, and refractive indices of the reference window and water. It represents the contribution of multiple reflections within the sample. Terms independent of the complex refractive index of the sample can then be grouped as follows:

$$\text{ReferenceTh} = r_S^{\text{AQ}} \cos^2 \theta_0 + r_P^{\text{AQ}} \sin^2 \theta_0, \quad (4)$$

$$\text{SampleThS} = t_S^{\text{AQ}} t_S^{\text{QA}} * P * \cos^2 \theta_0, \quad (5)$$

$$\text{SampleThP} = t_P^{\text{AQ}} t_P^{\text{QA}} * P * \sin^2 \theta_0. \quad (6)$$

Therefore, the theoretical ratio between the amplitude of the reference and sample electric fields at a given frequency can be written as

$$\frac{E_{\text{ref}}}{E_{\text{sam}}} = \frac{\text{ReferenceTh}}{\text{SampleThS} * \text{FP}_S + \text{SampleThP} * \text{FP}_P}. \quad (7)$$

The complex refractive index of the sample is finally extracted through minimization of the difference between the theoretical and corrected experimental ratios.

3.2. Image rendering

To improve the image SNR, circular average and median filters are applied at each frequency over 3×3 neighboring pixels. The filtered three-dimensional frequency-dependent complex refractive index dataset is then used for segmentation. Clustering and PCA algorithms as well

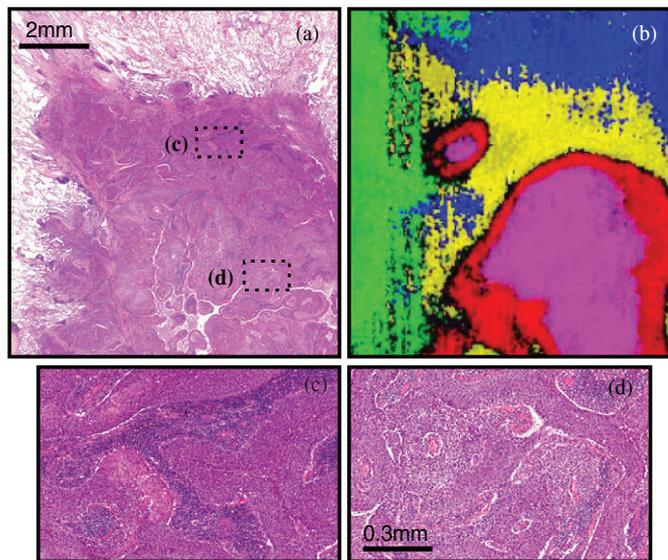


Figure 4. (a) Visible image of the hematoxylin-eosin-stained lung sample obtained from digital slide data. (b) Corresponding segmented THz image. (c), (d) High magnification close-ups corresponding to the areas marked on (a).

as validation tools used are included in the Fuzzy Clustering and Data Analysis Toolbox for MATLAB (MathWorks), made available by Balasko *et al* (2005).

All images shown hereafter are obtained by applying the fuzzy C-means clustering algorithm over the 0.8–1.2 THz refractive index (real part) sub-datasets. The optimal number of clusters was first estimated using Xie and Beni's index and further refined to produce meaningful images. Pixel color is determined by the membership degrees in each cluster. Pixels that cannot be classified in any of the clusters are displayed in black.

Further analysis of the data was performed using the PCA ability to visualize clusters on two-dimensional and possibly three-dimensional score plots. Separation of the clusters along different axes representing principal components suggests that the contrast originated from independent physical parameters. Although these parameters are not identified, water is a probable candidate since it has long been recognized as a potential contrast agent in the THz regime (Woodward *et al* 2003). Other candidates may include cells' sizes and arrangement, and their protein or DNA contents.

4. Results

4.1. Squamous cell carcinoma of the left lower lung

The tumor in the lung sample is well defined even though peripheral invasion can be found. Thorough investigation reveals two tissue subtypes within the tumor and two distinct types of healthy tissues as well as the existence of a necrotic area close to the center of the tumor mass.

Figure 4(a) shows the visible image obtained from digital slide data and corresponding to the THz image. Figure 4(b) is a segmented cluster image produced using the refractive index data. In agreement with the pathologic observations, two distinct healthy areas are identified (blue and green). The top blue area reveals the atelectatic lung parenchyma due to the cancer

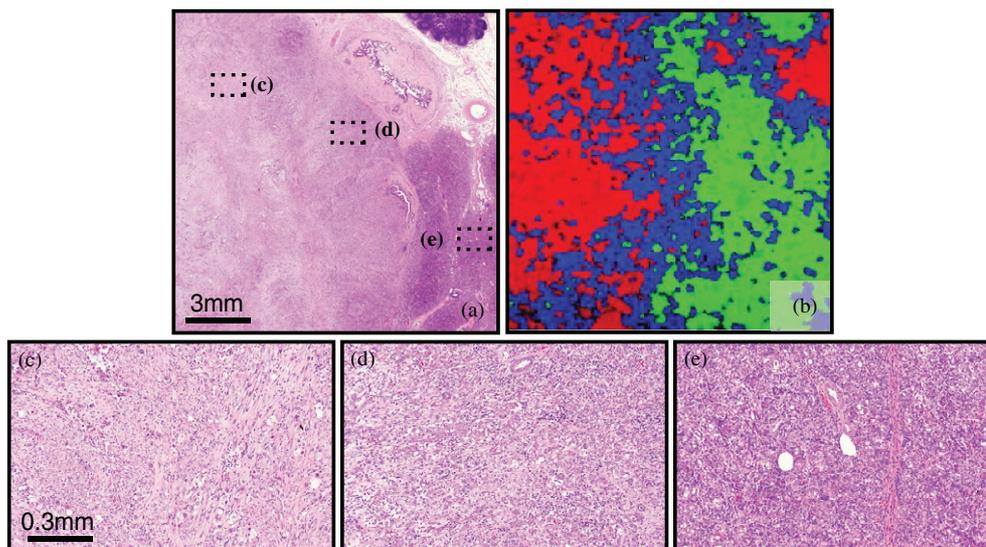


Figure 5. (a) Visible image of the stained pancreas sample obtained from digital slide data. (b) Corresponding segmented THz image. (c)–(e) High magnification close-ups corresponding to the areas marked on (a).

spread. Besides differentiation of tumors from healthy tissues, the THz image also reveals that the tumor is segmented in three areas (yellow, red and magenta). Figures 4(c) and (d) are high magnification close-up representatives of the cellular structure corresponding to the yellow and magenta areas, respectively. The contrast difference can be explained by the presence of massive lymphocytic infiltration in the yellow areas and the presence of tumor cell necrosis in the magenta areas. Red areas are thought to be intermediate and visible distinction is difficult.

Our measurements show that a low magnification visible image does not allow for a clear distinction between the various areas within the tumor. However, the inspection of high magnification images confirms that the segmentation obtained on the THz image corresponds to differences of cellular type and arrangement at the microscopic level. In particular, the concentration of lymphocytes is much higher in the magenta and red areas than in the yellow area.

4.2. Invasive ductal carcinoma of the pancreas

The tumor in the pancreas sample is so invasive and ill-defined at the invasion front that the borders cannot be visualized macroscopically. Inspection of high magnification images is required to distinguish cancer cells intermingled with healthy tissue components.

Figure 5(a) shows the visible image obtained from digital slide data. The corresponding segmented cluster image in figure 5(b) suggests a reasonable cancer border, with healthy areas appearing in green. A degree of invasiveness can also be visualized, although the low resolution prevents from detecting smaller cancerous cells' clusters. Figure 5(e) is a high magnification image representative of the cellular structure of the healthy areas. The THz image indicates that the tumor is composed of two tissue subtypes. Indeed, the cellular structure visible in figure 5(c) and corresponding to the red cluster seems to be associated with the amount of fibrotic tissue.

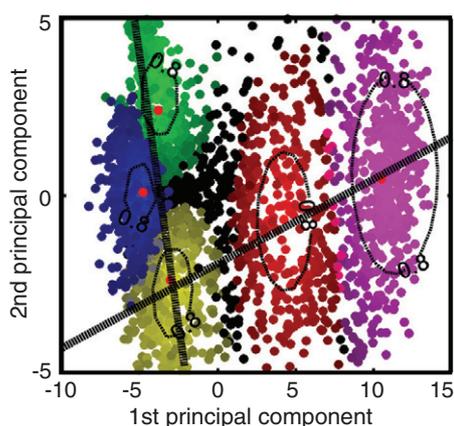


Figure 6. Principal component analysis projection along the first and second components applied to the data of figure 4(b). The ellipses show the core of each cluster, defined by the pixels with membership degree over 0.8. Blue and green dots correspond to healthy tissues while yellow, red and magenta dots correspond to tumor tissues.

4.3. Toward understanding the origin of contrast: lung cancer

Broader application of THz spectroscopic imaging to cancer diagnosis is currently hindered by the lack of understanding of the physical mechanisms behind the origin of the contrast on THz images. More information could be obtained when physically significant parameters were extracted, for example using dielectric spectroscopy methods such as fitting to a Debye model to evaluate the water content.

PCA identifies independent linear combination of measurement parameters (principal components), thus eliminating redundant information. The principal components are ordered in such a way that the variance of the data is maximal for the first component and decreasing for successive components. Therefore, most of the variance can usually be explained using only the first two or three principal components, reflecting the most meaningful physical parameters and thus simplifying data visualization. Different principal components represent independent physical parameters, so that PCA may point out different contrast mechanisms.

The score plot of figure 6 was obtained by applying PCA to the data corresponding to the segmented image of figure 5(b). The quasi-normal direction of the two axes suggests that the contrast between healthy and cancer tissues has a different origin than the contrast between subtypes of tumor tissues.

5. Conclusion

We demonstrated the possibility of obtaining medically relevant segmented images of standard tissue slices using THz spectroscopic imaging and clustering algorithms. By registration of the THz image to digital slide data, we were able to navigate through THz data and link the contrasts to cellular-level differences. We identified cancer tissues even in the ambiguous case of the pancreas tumor and classified subtypes of tumor tissues. Our results open the way to large-scale studies to better understand the contrast mechanisms and build up a database associating THz spectra with specific cellular arrangements. We believe that this could be a first step toward THz imaging aid for cancer diagnosis. THz images could macroscopically

point out areas on which pathologists should focus their attention and provide a first diagnosis as to the nature and complexity of the observed cancer.

References

- Balasko B, Abonyi J and Feil B 2005 Fuzzy clustering and data analysis toolbox Department of Process Engineering, University of Veszprem, Hungary <http://www.fmt.vein.hu/softcomp/fclusttoolbox/>
- Berry E, Handley J W, Fitzgerald A J, Merchant W J, Boyle R D, Zinov'ev N N, Miles R E, Chamberlain J M and Smith M A 2004 Multispectral classification techniques for terahertz pulsed imaging: an example in histopathology *Med. Eng. Phys.* **26** 423–30
- Chan W L, Deibel J and Mittleman D M 2007 Imaging with terahertz radiation *Rep. Prog. Phys.* **70** 1325–79
- Ferguson B and Zhang X C 2002 Materials for terahertz science and technology *Nat. Mater.* **1** 26–33
- Fitzgerald A J, Wallace V P, Jimenez-Linan M, Bobrow L, Pye R J, Purushotham A D and Arnone D D 2006 Terahertz pulsed imaging of human breast tumors *Radiology* **239** 533–40
- Jepsen P U, Moller U and Merbold H 2007 Investigation of aqueous alcohol and sugar solutions with reflection terahertz time-domain spectroscopy *Opt. Express* **15** 14717
- Kawase K 2004 Terahertz imaging for drug detection and large-scale integrated circuit inspection *Opt. Photonics News* **15** 34–9
- Molnar B, Berczi L, Diczhazy C, Tagsherer A, Varga S V, Szende B and Tulassay Z 2003 Digital slide and virtual microscopy based routine and telepathology evaluation of routine gastrointestinal biopsy specimens *J. Clin. Pathol.* **56** 433–8
- Nakajima S, Hoshina H, Yamashita M, Otani C and Miyoshi N 2007 Terahertz imaging diagnostics of cancer tissues with a chemometrics technique *Appl. Phys. Lett.* **90** 041102
- Pickwell E and Wallace V P 2006 Biomedical applications of terahertz technology *J. Phys. D: Appl. Phys.* **39** R301
- Woodward R M, Wallace V P, Pye R J, Cole B E, Arnone D, Linfield E H and Pepper M 2003 Terahertz pulse imaging of *ex vivo* basal cell carcinoma *J. Invest. Dermatol.* **120** 72–8