Concentration analysis of breast tissue phantoms with terahertz spectroscopy

BAO C. Q. TRUONG, ANTHONY J. FITZGERALD, SHUTING FAN, AND VINCENT P. WALLACE*

Terahertz and Optical Technology Group, Department of Physics, The University of Western Australia, Perth, 6009, Australia
*vincent.wallace@uwa.edu.au

Abstract: Terahertz imaging has been previously shown to be capable of distinguishing normal breast tissue from its cancerous form, indicating its applicability to breast conserving surgery. The heterogeneous composition of breast tissue is among the main challenges to progressing this potential research towards a practical application. In this paper, two concentration analysis methods are proposed for analyzing phantoms mimicking breast tissue. The dielectric properties and the double Debye parameters were used to determine the phantom composition. The first method is wholly based on the conventional effective medium theory while the second one combines this theoretical model with empirical polynomial models. Through assessing the accuracy of these methods, their potential for application to quantifying breast tissue pathology was confirmed.

OCIS codes: (110.6795) Terahertz imaging; (170.3880) Medical and biological imaging; (170.6935) Tissue characterization.

References and links
1. L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, “Global cancer statistics, 2012,” CA Cancer J. Clin. 65(2), 87–108 (2015).
2. R. G. Pleijhuis, M. Graafland, J. de Vries, J. Bart, J. S. de Jong, and G. M. van Dam, “Obtaining Adequate Surgical Margins in Breast-Conserving Therapy for Patients with Early-Stage Breast Cancer: Current Modalities and Future Directions,” Ann. Surg. Oncol. 16(10), 2717–2730 (2009).
3. J. R. Harris, M. E. Lippman, M. Morrow, and C. K. Osborne, Diseases of the Breast, 4 ed. (Lippincott Williams & Wilkins, 2010).
4. P. C. Ashworth, P. O’Kelly, A. D. Purushotham, S. E. Pinder, M. Kontos, M. Pepper, and V. P. Wallace, “An intra-operative THz probe for use during the surgical removal of breast tumors,” in 2008 33rd International Conference on Infrared, Millimeter, and Terahertz Waves, 2008), 1–3.
5. A. Portieri, M. Grootendorst, and T. Fitzgerald, “Intra-operative terahertz probe for detection of breast cancer,” in 2015 8th UK, Europe, China Millimeter Waves and THz Technology Workshop (UCMWT), 2015), 1–2.
6. C. Yu, S. Fan, Y. Sun, and E. Pickwell-Macpherson, “The potential of terahertz imaging for cancer diagnosis: A review of investigations to date,” Quant. Imaging Med. Surg. 2(1), 33–45 (2012).
7. T. C. Bowman, M. El-Shenawee, and L. K. Campbell, “Terahertz Imaging of Excised Breast Tumor Tissue on Paraffin Sections,” IEEE Trans. Antenn. Propag. 63(5), 2088–2097 (2015).
8. A. J. Fitzgerald, V. P. Wallace, M. Jimenez-Linan, L. Bobrow, R. J. Pye, A. D. Purushotham, and D. D. Arnone, “Terahertz Pulsed Imaging of Human Breast Tumors,” Radiology 239(2), 533–540 (2006).
9. A. J. Fitzgerald, S. Pinder, A. D. Purushotham, P. O’Kelly, P. C. Ashworth, and V. P. Wallace, “Classification of terahertz-pulsed imaging data from excised breast tissue,” J. Biomed. Opt. 17(1), 016005 (2012).
10. P. C. Ashworth, E. Pickwell-MacPherson, E. Provenzano, S. E. Pinder, A. D. Purushotham, M. Pepper, and V. P. Wallace, “Terahertz pulsed spectroscopy of freshly excised human breast cancer,” Opt. Express 17(15), 12444–12454 (2009).
11. M. R. Grootendorst, A. J. Fitzgerald, S. G. Brouwer de Koning, A. Sanatoalla, A. Portieri, M. Van Hemelrijck, M. R. Young, J. Owen, M. Cariati, M. Pepper, V. P. Wallace, S. E. Pinder, and A. Purushotham, “Use of a handheld terahertz pulsed imaging device to differentiate benign and malignant breast tissue,” Biomed. Opt. Express 8(6), 2932–2945 (2017).
12. C. B. Reid, E. Pickwell-MacPherson, J. G. Lauffer, A. P. Gibson, J. C. Hegden, and V. P. Wallace, “Accuracy and resolution of THz reflection spectroscopy for medical imaging,” Phys. Med. Biol. 55(16), 4825–4838 (2010).
13. T. Bowman, A. Walter, and M. El-Shenawee, “Margin assessment of three-dimensional breast cancer phantoms using terahertz imaging,” in SPIE BiOS, (SPIE, 2016), 5.
1. Introduction

Breast cancer is the most commonly diagnosed cancer as well as the deadliest form of cancer among females, which is accountable for approximately 25% of all cancer cases and 15% of all cancer deaths [1]. The standard first line treatment for early breast cancer is breast conserving surgery (BCS); 70% of breast cancer patients undergo this type of surgery. BCS involves removal of the breast tumor with the twin goal of achieving a tumor-free margin of 2 mm and conserving the maximum amount of healthy tissue [2]. For about 15%-20% of all cases, there is a failure to remove the cancer with an adequate margins; this leads to local recurrence [3, 4]. As a result, secondary operations are required to ensure complete removal of all the diseased tissue; this has various negative consequences for the patients regarding morbidity, cosmetic results, infection, survival rates, and additional overall cost of the treatment. Therefore, accurately identifying the tumorous areas in breast tissue during BCS is crucial for complete resection and minimizing the number of secondary surgeries. Histopathological examination to assess the surgical margin is current clinical practice; it is time-consuming and, hence, not capable of real-time monitoring. There is a clear need for intraoperative screening techniques to aid surgeons in conducting BCS successfully.

Terahertz (THz) imaging is among the recent potential intraoperative techniques for BCS [4, 5]. Lying between the millimeter waves and the infrared (IR) spectral region, THz radiation is non-ionizing, which makes it ideal for biomedical applications, while still being able to achieve good resolution for imaging. The high sensitivity of THz radiation to polar substances, especially water, also plays a crucial role in its application to medical imaging as biological tissue has high water content [6]. Breast tissue samples from lumpectomies were imaged by terahertz systems, showing contrast between healthy and cancerous tissues [7, 8]. Fitzgerald et al. found good correlation in the shape and area of breast tumor between THz images and histology [9].

Previous studies of breast cancer using THz imaging have found that the inhomogeneity of breast tissue poses a challenge for accurate detection of the tumor [7, 9]. The spectroscopic investigation of freshly excised cancer breast tissue by Ashworth et al. indicates the contrast seen in breast tissue THz images is the result of differences in refractive index and absorption coefficient between sample groups with only one predominant component (e.g. adipose tissue, fibrous tissue or tumor) [10]. However, it is the lack of the specific understanding of how tissue composition with more than one component can affect the differentiation between healthy and cancerous tissue. The presence of adipose and fiber as background tissues in tumor samples potentially causes misclassification of the tumor regions [11]. Diagnosis of the tissue pathologies can be of great help to improve cancer classification with more informative inputs. Therefore, it is essential to seek an answer for the possibility of identifying tissue composition through the THz imaging data. A preliminary study of tissue composition has been conducted with water and lipid emulsions [12]. The authors employed three different methods – linear spectral decomposition, spectrally averaged dielectric coefficient, and Debye relaxation coefficient methods, to specify the sample composition. The linear spectral decomposition method gave the best results overall and is suitable for three component materials as might be expected in breast tissue. However a limitation of this method is that it requires prior knowledge of the absorption coefficients of each pure component in the mixture. The spectrally averaged dielectric coefficient method gave the worst resolution of the three methods, and was limited in only being applied to two component tissue models. The Debye relaxation coefficient gave similar results to the linear spectral decomposition method but was not suitable for breast tissue composition because it could not operate with
more than two components as found in breast tissue. Other limitations of this method are that it relies on the transmission data to fit linear regression models but applies them for reflection measurements, which are inherently noisier and may differ slightly due to systematic variations between transmission and reflection spectroscopy setups. It also relies on a measurement of the change in Debye relaxations to determine changes in the concentrations of all components. However, the outcomes of that study are sufficient for motivating further investigation of the potential of THz spectroscopy regarding the analysis of tissue composition and pathology.

More recent work has advanced the use of tissue model phantoms that can be used to represent the optical properties of breast tissue in the THz frequency range. Phantoms reported in [13, 14] were designed to resemble the THz properties of fresh infiltrating ductal carcinoma, fibroglandular tissue, and fatty tissue of the breast as published in the literature [10]. These phantoms were made to be highly tunable to breast dielectric properties in solid form and were composed of water, oil, surfactant TX151 and agar in varying proportions. The advantages of these phantoms is that they can be used to investigate breast cancer margin assessment and allow the exploration of carbon-based contrast agents to improve cancer detection [15].

This study aims to advance existing findings of breast cancer detection using THz imaging in reflection through investigating the possibility of specifying composition of phantoms mimicking the three component nature of breast tissue. The two-component and three-component breast phantoms were created with controllable compositions to be capable of representing the broad interpatient variation of breast tissue and therefore the wide range of dielectric properties observed when imaging breast with THz [11]. This range occurs due to the inhomogeneous nature of this tissue type and the variations of composition with age and breast density [16]. The three major components we used for these phantoms, when combined, were intended to give similar dielectric properties to the combined main components of breast tissue which are, on average water, lipid and collagen. Together these make up over 90% of the breast material. Water comprises an estimated 35% of breast tissue [16], while lipid, which contributes to the dominant proportion of adipose tissue (and breast is on average 45% adipose) makes up an estimated 30% of the breast [17]. The third major component of breast, collagen is the primary component of fibroglandular stroma cells. Fibroglandular tissue density varies widely with age but on average is around one third of the breast giving approximately 30% of breast tissue in the form of collagen [18]. The advantage of using three primary components for a phantom was that it was possible, using a single solid phantom type, to model the range of variations in breast tissue composition expected to be seen with age and density.

The main contribution of this paper is to propose the application of two analytical models based on effective medium theory (EMT). These methods are proposed to solve the problems of the previous models, overcoming the need for prior knowledge of dielectric properties and accounting for the medium dynamics in modelling. These two approaches facilitate analysis of composition concentration in three-component phantoms given the prior information obtained from the two-component phantoms. The first method combines the EMT models with optimization techniques to approximate dielectric properties of pure components from the two-component phantoms and then applies them to analysis of the three-component phantoms. The second one employs the stepwise regression method to develop empirical functions describing the relationship of a double Debye parameter with each component and use them in conjunction with the EMT for estimation of the concentrations in the three-component phantoms.
2. Experiments and methodology

2.1 Sample preparation

The main materials used for making breast tissue phantoms in this study include lipids/oil, gelatin and deionized water. These were selected due to their similarity to breast tissue components such as fat and proteins. These materials together allow for the creation of phantoms that replicate the real breast tissue response with controllable concentrations of individual components. There are two types of the phantoms investigated in this study, made of either two components or three components. The two-component phantoms are made from emulsions of water and lipid or water and gelatin while the three-component ones primarily consist of oil, gelatin, and deionized water. Oil represents the fat content of the tissue and gelatin, a main protein of connective tissue in mammals, plays the role of proteins in the tissue.

Similar to a previous study [12], a commercial water and lipid emulsion – Intralipid 20% (Sigma Aldrich, UK) is used to avoid difficulties in mixing oil and water effectively. The intralipid contains 20% of soybean oil stabilized with phospholipid. The concentration of the original intralipid was reduced by the dilutions of this emulsion with deionized water and increased by letting the water evaporate inside a fume hood. This produced a series of water and lipid emulsions with the volume percentage of lipid varying from 4% to 67%.

Gelatin powder was dissolved in deionized water to create a series of gelatin mixtures; weight percentages of these mixtures ranged from 5% to 30% with 5% increments are converted into volume percentages using the average mass density of proteins – 1.4 g/cm³ [19].

The three-component phantoms were created from mixing water, gelatin and olive oil with varying proportions. Containing all three main components makes the phantoms resemble a wide range of compositions observed for breast tissue. A small amount of p-toluic acid, n-propanol and formaldehyde were used to reduce the aggregation of gelatin in water solution and improving the cross-linking of the gelatin molecules. The main steps to produce the three-component phantoms include: mix p-toluic acid and n-propanol with the deionized water; add gelatin to the water solution and stir with a homogenizer; add pre-heated oil to the gelatin mixture and stir vigorously in the homogenizer to achieve a uniform emulsion; finally add - with stirring, formaldehyde solution to the emulsion. The final emulsion was poured into a petri dish, cooled down to room temperature, and then left to set in the next 24 hours. More details of the procedure are also presented in the previous studies [20, 21]. Table 1 provides specific concentrations of each main component in the three-component breast phantoms.

| Sample | %Vol. Water | %Vol. Gelatin | %Vol. Oil |
|--------|-------------|---------------|-----------|
| WGL1   | 81.9        | 13.2          | 0.0       |
| WGL2   | 74.7        | 11.8          | 9.3       |
| WGL3   | 64.5        | 10.3          | 21.3      |
| WGL4   | 56.4        | 9.4           | 30.7      |
| WGL5   | 46.4        | 7.4           | 43.1      |
| WGL6   | 36.4        | 5.9           | 54.8      |

2.2 Spectroscopy system

All terahertz time-domain spectroscopy measurements were conducted in reflection mode using a TeraPulse4000 system (TeraView Ltd, Cambridge, UK). The system produces THz pulses with a spectral range from 0.06 to 4 THz; more details of the system can be found in [22]. The broadband THz pulses are focused on the top of a quartz window where the sample is placed (see Fig. 1). A reference measurement is also taken without any sample on the quartz. Reflected signals are acquired from either a quartz-sample (in the sample...
measurement) or quartz-air (in the reference measurement). Both reference and sample pulses are required for the calculation of optical properties and complex permittivity of the samples described in Section 2.3. The temperature of the laboratory was kept constant at 24°C throughout the experiments.

2.3 Data acquisition

Each measured sample/reference waveform - \(E_{\text{sample}}(t)\) and \(E_{\text{air}}(t)\), includes two reflected pulses (see Fig. 1). \(E_{\text{ref}}(t)\) is a reflection from the air and quartz interface when the terahertz radiation arrives the bottom surface of the quartz window. The method to extract the dielectric constants in the reflection geometry proposed by [23] was used in this study after validating that it provided accurate values for water at a range of temperatures with known optical properties. This method is capable of tackling difficulties in arbitrary incidence angle and polarization of the electric field. An incident angle of 30° and a polarization angle of 90° were applied due to the similarity between our reflection spectroscopy system and that described in [23]. Testing the method with these values gave accurate comparison of the known properties of water. The use of the reference signal \(E_{\text{ref}}(t)\) facilitates the calculation of a frequency-dependent calibration factor as explained in [23] that compensates for any difference in propagation axis and coupling to the detector antenna that exists between the reference beam (from the bottom interface of air and quartz) and the sample beam (from the top interface of quartz and sample). The deconvolution of the sample reflection with the reference reflection, while windowing out the air-quartz reflection from the bottom interface, allows removal of noise due to systematic and environmental factors. An equation based on the calibration factor and the ratio of the sample reflection spectrum to the reference reflection spectrum and the Fresnel’s equations is employed for the extraction of refractive index and absorption coefficient [23]. Due to the complexity of this equation, this study applied MATLAB nonlinear optimization solver – \texttt{fmincon}, using sequential quadratic programming (SQP) to solve the equation and extract the optical properties [24].

2.4 Parameter extraction

The double Debye model (DD) has been used effectively to model the dielectric properties of a wide variety of biological tissue in the terahertz regime [25]. The high content of water in human tissue is the primary motivation for the application of two Debye relaxation processes. The first Debye process is associated with breaking the existing tetrahedral structures of the water molecules under the excitation of an incident THz radiation. Then, new tetrahedral sites
are formed in the second relaxation due to the reorientation of the previously freed molecules [26, 27]. The double Debye function takes the following form:

\[
\hat{\varepsilon}(\omega) = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_\infty}{1 + j\omega\tau_1} + \frac{\varepsilon_z - \varepsilon_\infty}{1 + j\omega\tau_2},
\]

where \(\hat{\varepsilon}(\omega)\) represents the frequency-dependent complex permittivity of the sample, \(\varepsilon_s\) is the static dielectric constant at the low-frequency limit, \(\varepsilon_z\) is the intermediate dielectric constant associated with the transition between the two Debye relaxation processes and \(\varepsilon_\infty\) is the high-frequency limit of the complex permittivity. \(\varepsilon_s - \varepsilon_\infty\) and \(\varepsilon_z - \varepsilon_\infty\) indicate the amplitude dispersion of permittivity with the relaxation time of \(\tau_1\) and \(\tau_2\) respectively [28].

The frequency-dependent extracted refractive index \(n(\omega)\) and absorption coefficients \(\alpha(\omega)\) of the samples are used to calculate the complex permittivities of the samples through the relation \(\hat{\varepsilon}(\omega) = \left(\left(n(\omega) - 1\right) - \frac{c}{2\omega}\alpha(\omega)\right)^2 = \varepsilon' - j\varepsilon''\), where \(\omega\) is the angular frequency and \(c\) is the speed of light in vacuum. After extracting the optical properties, these were converted to dielectric values and the nonlinear optimization algorithm (SQP) employed to fit the DD model to the complex permittivity spectrum of each phantom sample over a frequency range up to 2 THz. All five parameters of the DD model were extracted through the fitting algorithm for the tissue phantom samples except for the water sample that has \(\varepsilon_s\) pre-calculated using its previously-published relation with measurement temperature [28].

Since the applied analysis method in this study is based on the effective medium theory, the set of only permittivity-related parameters consisting of \(\varepsilon', \varepsilon'', \varepsilon_s, \varepsilon_z, \varepsilon_\infty\) is investigated. \(\varepsilon', \varepsilon''\) are the average real and imaginary values of the sample complex permittivity over the applied frequency range. In this paper, the analysis parameters - \(\varepsilon', \varepsilon'', \varepsilon_s, \varepsilon_z, \varepsilon_\infty\), are represented by \(\varepsilon \subset X = (\varepsilon', \varepsilon'', \varepsilon_s, \varepsilon_z, \varepsilon_\infty)\), which is regularly used in the next section.

### 2.5 Analysis methodology

#### 2.5.1 Correlation analysis

Rank correlation analysis using the Spearman’s coefficient was employed to analyze the relationship between \((\varepsilon', \varepsilon'', \varepsilon_s, \varepsilon_z, \varepsilon_\infty)\) and the sample concentrations [29]. Since this relationship can be either linear or nonlinear, the Spearman method, which measures the degree of monotonic association between two variables, is preferable to the popular Pearson’s coefficient that examines the linearly related variables [30]. The parameters with the highest correlation coefficients is taken to the further analysis described in the following sections.

#### 2.5.2 Effective medium theory

The dielectric properties of heterogenous mixtures can be handled effectively using the effective medium theory (EMT) [31]. The Clausius-Mossotti relation based on the Lorentz field provides the basis for the first traditional EMT model – the Maxwell-Garnet (MG) function, that considers particles spherical and independent from each other [32]. Then, this MG model was advanced by Polder and van Santeen (PvS) with consideration of a range of different ellipsoidal shapes of the particles while Bruggeman (BM) considers a symmetrical interaction between composites within the mixtures [33, 34]. An application of the EMT to the terahertz field is the proposed quasi-static EMT model that takes account of distributions and shapes of additive particles in a host matrix as well as the arbitrary concentrations of the
composites [35]. The BM model was also applied by Bennett et al to calculating the dielectric constants of a composite skin model with its pre-defined hydration profile in their study of terahertz reflectometry for the skin [36]. The general BM equation for any $K$ number of components with considering the shape of particles is of the form [31]

$$\sum_{k=1}^{K} \eta_k \left( \frac{\varepsilon_k - \varepsilon_r}{\varepsilon_k + (P_k - 1) \varepsilon_r} \right) = 0 \quad (2)$$

where $\eta_k$, $\varepsilon_k$, and $P_k$ respectively represent the volume fraction, the $\varepsilon$ value, and the depolarizing factor of the $k^{th}$ component, and $\varepsilon_r$ is the $\varepsilon$ value of the mixture. The depolarizing factor is defined using the ratios between semi-axes of the general ellipsoid of the component particles [37]. In the popular spherical case, which is also considered in this paper, $P_k \to 1/3$ substituted into Eq. (2) yields the widely known BM model

$$\sum_{k=1}^{K} \eta_k \left( \frac{\varepsilon_k - \varepsilon_r}{\varepsilon_k + 2 \varepsilon_r} \right) = 0, \quad \text{with} \sum_{k=1}^{K} \eta_k = 1. \quad (3)$$

Application of the BM model requires prior information of the dielectric properties for each pure component in the mixture. These values of water and oil can be directly determined through experiments and calculation procedure described in Sections 2.3 and 2.4. Considering the biological aspect, proteins often take a hydrated form that enables more engagement with metabolic functions. Analysis of hydrated proteins, such as gelatin in liquid water, provides more accurate understanding of biological protein dynamics than that of dried powder proteins [38, 39]. Therefore, the following optimization based on Eq. (3) was used

$$\min \sum_{i=1}^{N} \eta_i \left( \frac{\varepsilon_i - \varepsilon_{r,w}}{\varepsilon_i + 2 \varepsilon_{r,w}} \right) + \eta_i \left( \frac{\varepsilon_i - \varepsilon_{r,\ell}}{\varepsilon_i + 2 \varepsilon_{r,\ell}} \right) \mid \varepsilon_i \geq 0. \quad (4)$$

Here $\varepsilon_i$ is the known permittivity value of the first pure component in an arbitrary two-component mixture with its permittivity $\varepsilon_r$, $N$ is the total number of the two-component samples. Instead of conducting spectroscopic measurements for gelatin powder, the dielectric properties of “pure” gelatin were approximated using Eq. (4) and water-gelatin samples.

Application of the BM model to two-component phantoms: Given the pre-defined values of the pure components in the two-component samples, the concentration values of the two components can be obtained by solving a system of two equations from Eq. (3) in the two variables ($\eta_1, \eta_2$).

Application of the BM model to three-component phantoms: In this case, Eq. (3) becomes an underdetermined system of two equations in three variables ($\eta_1, \eta_2, \eta_3$). Since this equation system has many solutions which are inconsistent, an optimization approach can be applied to tackle this problem. Particularly, instead of using only one parameter from $X_\varepsilon$, two or more can be used simultaneously to improve the stability of the solution, as described below,

$$\min \sum_{i=1}^{L} \sum_{j=1}^{M} \eta_k \left( \frac{\varepsilon_{l,k} - \varepsilon_{r,j}}{\varepsilon_{l,k} + 2 \varepsilon_{r,j}} \right) \mid \eta_k \geq 0, \sum_{k=1}^{K} \eta_k = 1, \quad (5)$$

With $X_\varepsilon \subset X_\ell$, $L$ - the total number of parameters used in $X_\varepsilon$, and $K$ is the number of mixture components ($K \leq 3$ in this study). The SQP solver mentioned earlier was employed for solving both Eq. (4) and Eq. (5) [24]. In fact, Eq. (5) can be effectively applied for the
two-component phantoms as well. However, this does not mean that Eq. (5) can perform better than Eq. (3) in this case.

2.5.3 Empirical EMT-based method

The empirical EMT-based method (E-EMT) is based on the assumption that the relationship between the investigated parameters and the component concentrations could take polynomial forms in which the concentration is a function of each parameter \( \varepsilon \), such as,

\[
C_L(\varepsilon) = \sum_{n=0}^{N_L} a_n \varepsilon^n
\]

(6)

\[
C_G(\varepsilon) = \sum_{n=0}^{N_G} b_n \varepsilon^n
\]

(7)

where \( C_L \) and \( C_G \) are the concentration of the lipid and gelatin respectively. \( N_L \) and \( N_G \) are the largest degrees of the polynomial functions while \( a_n \) and \( b_n \) are their coefficients of the \( n \)-degree exponents. A stepwise regression procedure for model training was performed on the data of intralipid and water-gelatin emulsions to determine the constant coefficients \( N_L \), \( N_G \), \( a_n \), and \( b_n \) in Eq. (6) [40]. This form of regression facilitates a repeated procedure of adding or removing terms from a generalized linear regression model based on statistical improvement in fitting the model. The \texttt{stepwiselm} function from the statistics and machine learning toolbox of MATLAB was employed to perform the stepwise regression method [41].

With the obtained coefficients from the training, function Eq. (6) is used to calculate the lipid concentration in three-component phantoms. Although Eq. (6) was developed based on the two-component emulsions of lipid and water, similar relationship between \( \varepsilon \) and lipid concentration throughout all phantom samples, which is be shown later in the section of results, supports the applicability of this function to the three-component emulsions. Contrary to the case of lipid concentration, \( \varepsilon \) the gelatin concentration seems to respond differently due to changes in gelatin concentration in different types of phantom samples. The reason is that gelatin molecules can attract a large number of water molecules in aqueous environment and form a hydrated gelatin complex, which has different dielectric response from both pure gelatin and bulk water [42]. This ability of to trap water molecules depends on the concentration of the gelatin in the solution [43]. Therefore, to identify the gelatin concentration with Eq. (7) in the three-component phantom, the response of water-gelatin must be separated from the lipid content. Considering the three-component samples as only having two components - lipid and a complex of water and gelatin, the BM model Eq. (3) can be effectively applied for calculating an intermediate value \( \varepsilon_{wg} \) that represents the \( \varepsilon \) value of the water-gelatin complex alone in each three-component phantom. Given the pre-defined lipid concentration \( \eta_L \) of a three-component phantom by Eq. (5), \( \varepsilon_{wg} \) can be obtained through

\[
\varepsilon_{wg} = \varepsilon_{wgf} \left( \frac{1 - \eta_L + A}{1 - \eta_L - 2A} \right)
\]

(8)

where \( \varepsilon_{wgf} \) is the \( \varepsilon \) value of the three-component phantom, \( A = \eta_L \left( \frac{\varepsilon_L - \varepsilon_{wg}}{\varepsilon_L - 2\varepsilon_{wg}} \right) \), and \( \varepsilon_L \) is the \( \varepsilon \) of the lipid. Then, substituting the value of \( \varepsilon_{wg} \) into Eq. (6) allows calculation of the gelatin concentration in the three-component phantom. The concentration of water is obtained by subtracting the achieved concentrations of oil, gelatin, and others.
3. Results and discussion

![Fig. 2. The measured real and imaginary part of complex permittivities of water and olive oil compared to existing literature values.](image)

The measured frequency-dependent complex permittivities for oil and water are shown in Fig. 2 compared to literature values for water from [26] and oil from [21]. The dielectric properties for all breast tissue phantoms were calculated from experiments and are shown in Fig. 3. To demonstrate the ability of the three component phantoms to model the range of THz properties from the possible variety of breast tissue compositions, the values from [44] were converted to dielectric values and are compared to the values for the ternary mixtures in Fig. 3c. It can be seen that the ternary phantoms represent the range of clinically relevant THz properties for breast tissue margin assessment.

The real and imaginary permittivities of oil and water used in the BM model (Fig. 2) were averaged over many measurements. The decrease in water concentration leads to a drop in both the real and imaginary parts of the phantom dielectric properties. The frequency range from 0.4 to 2 THz seems to have the most reliable spectra in which the low and high-frequency noise is minimal. In fact, our investigation of various spectral ranges, especially with variation in the low-frequency limit, using the methods proposed in this paper also demonstrates the optimality of this frequency range. Thus, the values of $\varepsilon''$, $\varepsilon'''$, $\varepsilon'', \varepsilon$, $\varepsilon_\infty$, $\varepsilon_\infty''$ were extracted from all the phantom samples over the range of 0.4 to 2 THz.

| Parameter | Water Concentration | Lipid Concentration | Gelatin Concentration |
|-----------|---------------------|---------------------|-----------------------|
| $\varepsilon$ | $0.79 \,(p<0.001)$ | $-0.52 \,(p<0.001)$ | $-0.23 \,(p = 0.27)$ |
| $\varepsilon_2$ | $0.88 \,(p<0.001)$ | $-0.97 \,(p<0.001)$ | $0.35 \,(p = 0.09)$ |
| $\varepsilon'''$ | $0.70 \,(p<0.001)$ | $-0.63 \,(p<0.001)$ | $0.02 \,(p = 0.91)$ |
| $\varepsilon''$ | $0.81 \,(p<0.001)$ | $-0.93 \,(p<0.001)$ | $0.41 \,(p = 0.04)$ |
| $\varepsilon''''$ | $0.98 \,(p<0.001)$ | $-0.88 \,(p<0.001)$ | $0.04 \,(p = 0.83)$ |

Table 2 shows the Spearman correlation coefficients of all five analysis parameters with the component concentration in all phantom samples. The concentration of water and lipid is highly correlated with these parameters while their correlation coefficients with gelatin are low and not statistically significant. This can be explained by the small proportions of gelatin content as compared to those of lipid and water. $\varepsilon_2$, $\varepsilon'''$, and $\varepsilon''''$ have the highest correlation with the concentrations of all the phantom components. The high correlation of $\varepsilon_2$ with the lipid concentration in intralipid emulsions was also reported previously by [12].
Fig. 3. The complex dielectric properties permittivities of (a) intralipid samples with increasing oil concentration, (b) gelatin-water emulsion samples with increasing gelatin concentration, (c) three-component phantoms with increasing solute (gelatin/oil) concentration compared to the dielectric properties reported in the literature for breast tissue. The notation w:g:o represents the volume percentage of water, gelatin, and oil, respectively.

3.1 EMT

Even though it is theoretically possible to apply all the five analysis parameters mentioned earlier to this method, $\varepsilon'$ and $\varepsilon''$ prove to be the most suitable for the concentration analysis using the EMT approach for all three types of the breast phantoms.

Intralipid emulsions: By solving Eq. (4) with the data of these two-component phantoms, the obtained values of $\varepsilon''$ and $\varepsilon'''$ for pure lipid are approximately 2.02 and 0.02. These values are quite close to those measured from oil as can be seen in Fig. 2. This figure also presents the averaged complex permittivity of water, which were calculated from the experimental data with distilled water.
Fig. 4. Plots show the results of the concentration analysis with EMT method of (a) intralipid emulsions, (b) water-gelatin emulsions, (c) water-gelatin-oil emulsions.

With the pre-defined values of the pure components, their concentrations in the intralipid emulsions were able to be estimated by Eq. (3). Figure 4(a) illustrates the good agreement between the measured and estimated concentrations of lipid in the intralipid emulsions.

**Water-gelatin emulsions:** The $\varepsilon'$, $\varepsilon''$ of pure gelatin, which were also estimated through applying Eq. (4) to the gelatin-based two-component mixture, has the values of 4.23 and 1.48 respectively. These values agree well with the complex permittivity of hydrated gelatin reported in [21]. The estimated concentration percentages of gelatin using Eq. (3) are plotted against their measured values in Fig. 4(b), showing a relatively good fit.

**Water-gelatin-oil emulsions:** With the extracted complex permittivities of water and oil from the experimental data, Eq. (5) was applied to estimate the component concentrations in the three-component phantom samples. The estimated values of these volume concentrations are compared with their actual values in Fig. 4(c). The estimation of lipid concentration effectively simulates the real data, especially in the samples having the volume concentration of lipid less than 40%. At the increasingly higher lipid concentration, the fitted values begin to diverge from the measured ones. The estimation of water concentration values shows the opposite trend with the samples having water concentration of less than 70% diverging. The
low accuracy in estimating the water and lipid concentration seems to affect the estimation of gelatin concentration as its content is accountable for a much smaller proportion compared with the others in the three-component phantoms.

![Graph showing concentration analysis results](image)

**Fig. 5.** Plots show the results of the concentration analysis with E-EMT method of (a) intralipid emulsions, (b) water-gelatin emulsions, (c) water-gelatin-oil emulsions.

### 3.2 E-EMT

**Intralipid emulsions:** Based on the results of Spearman correlation analysis in Table 2, $\varepsilon_2$, $\varepsilon^\prime$, $\varepsilon^\prime\prime$ were chosen for implementation of the E-EMT method. We found that this method is specifically effective for application to $\varepsilon_2$. Thus, the analysis outcomes of applying this DD parameter are used for validation the effectiveness of the E-EMT method and, as a result, presented in this paper. $\varepsilon_2$ values of intralipid samples and pure olive oil were used as input data for the E-EMT to define the relationship between lipid concentration and this parameter.
Table 3. The accuracies of the two proposed analysis methods corresponding to each type of the breast tissue phantoms. Each result is presented by the mean difference (the estimation error) in percentage (%) between the estimated and true concentration values with its standard deviation.

| METHOD | PHANTOM          | LIPID   | GELATIN | WATER   |
|--------|------------------|---------|---------|---------|
| EMT    | Intralipid       | 2.3(±1.8)| -       | 2.3(±1.6)|
|        | Water-gelatin    | -       | 2.0(±1.3)| 2.0(±1.3)|
|        | Water-gelatin-oil| 2.9(±3.8)| 6.0(±3.7)| 6.1(±4.2)|
| E-EMT  | Intralipid       | 2.7(±3.2)| -       | 2.7(±3.2)|
|        | Water-gelatin    | -       | 1.4(±1.1)| 1.4(±1.1)|
|        | Water-gelatin-oil| 1.8(±2.2)| 1.5(±1.2)| 2.1(±2.3)|

This function takes the form $C_{\text{lipid}}(\varepsilon_z) = a_0 + a_1\varepsilon_z$ with $a_0 = 172.49$ and $a_1 = 38.03$.

Figure 5(a) depicts the good fit between the true lipid concentrations and their estimations by the linear function.

Water-gelatin emulsions: Similarly, based on the data from the water-gelatin emulsions, the relationship between gelatin concentration and $\varepsilon_z$ is described by $C_{\text{gelatin}}(\varepsilon_z) = b_0 + b_1\varepsilon_z + b_2\varepsilon_z^2 + b_3\varepsilon_z^3 + b_4\varepsilon_z^4$, where $b_0 = 8.5390 \times 10^7$, $b_1 = -7.5386 \times 10^7$, $b_2 = 2.4966 \times 10^7$, $b_3 = -36708 \times 10^6$, $b_4 = 2.0246 \times 10^4$. The low correlation coefficient of this DD parameter with gelatin concentration shown in Table 2 explains the necessity to use such a high-degree polynomial to fit the data. As can be seen in Fig. 5(b), the gelatin concentration in the water-gelatin emulsions is not perfectly simulated by the polynomial model. Although a polynomial of higher degree can improve the fitting, the function could overfit the data. In fact, the stepwise regression algorithm has removed the higher-degree terms while training $C_{\text{gelatin}}(\varepsilon_z)$.

Water-gelatin-oil emulsions: According to the procedure of the E-EMT method described earlier, the trained model $C_{\text{lipid}}(\varepsilon_z)$ was applied to calculate the lipid concentrations in the three-component phantoms. Then, the function in Eq. (8) was employed to identify the $\varepsilon_{z,\text{wg}}$ of the water-gelatin complex with respect to each phantom. Finally, the trained $C_{\text{gelatin}}(\varepsilon_z)$ allows the calculation of gelatin concentration for each component in the three-component emulsions. All estimated values are plotted against the measured in Fig. 5(c), showing an excellent agreement.

3.3 Methodology comparison

The accuracy of the proposed methods was assessed through the estimation error which is defined by the mean difference between the estimated and true concentration values. The corresponding standard deviation of the error was also calculated to assess the robustness of the analysis methods. Table 3 shows the detailed accuracy values of the concentration analysis for each group of the examined breast tissue phantoms.

The E-EMT method generally outperforms the EMT one, except for the case of the intralipid emulsion. The measured data of pure olive oil was used in conjunction with the intralipid emulsion data for training $C_{\text{lipid}}(\varepsilon_z)$ to make this model more generalized for different kinds of lipid. This improves the calculation accuracy of lipid concentration with the E-EMT for the three-component phantoms while at the same time adding more error to the analysis of intralipid emulsions those only contain soybean oil. The high-degree polynomial...
of \(C_{\text{gelatin}}(\varepsilon_2)\) in the E-EMT method could help achieve the better accuracy for calculating the gelatin concentration in both water-gelatin and three-component emulsions. This empirical relationship between \(\varepsilon_2\) and gelatin concentration was carefully developed based on our experimental data, but its physical association has not been understood. Regarding the three-component emulsions, the E-EMT proves to be a more suitable approach with the considerably lower errors in estimating each discrete concentration than the EMT. Notably, the estimation of both gelatin and water concentration was significantly improved with the estimation error reducing from 6.0(±3.7) and 6.1(±4.2) to 1.5(±2.2) and 2.1(±2.3) respectively, whereas the improvement in estimating the oil concentration was achieved with a lesser degree (from 2.9(±3.8) to 1.8(±2.2)). With the achieved accuracy of the E-EMT method, a difference of approximately 5% in volume concentration of all the three components is detectable given that the error of the weight measurement is negligible.

Our aim is to transfer these models to apply them to THz imaging of fresh breast tissue. The sensitivity seen is encouraging for detecting the changes in water and fibrous concentration that might be expected in cancerous breast tissue based on the published values of THz optical properties that indicate changes of order of 10% [10]. We believe as a first approximation the results obtained indicate that the analysis methods may give useful results for determining the three main components of breast tissue, however the models may require some refinements to account for the other components in breast tissue and the degree of variation of those.

Since the commercial intralipid emulsions have been among the most popular tissue-mimicking phantoms for optical experiments, developing empirical model of lipid concentration in tissue phantom based on these lipid-water mixtures help reduce the uncertainty of the modelling. Testing the model with phantoms built from a different type of lipid, particularly olive oil of ternary phantoms versus soybean oil in this study, allows generalization of the proposed modelling technique despite that this increases the error of our analysis approach. Also, the generality of the lipid concentration model in Eq. (6) could be improved further through involving more types of lipid in training.

In theory, the models could potentially handle more than three components. However, the main challenge of the two models is the hypothesis on the use of the effective medium theory for the phantoms and the limitation of the Bruggeman model. The E-EMT tackles the limitation of BM model when dealing with complex mixtures by empirically determining the lipid concentration first and then applying the BM for only two components (water and gelatin). So, if additional components change the behaviour of lipid in the phantoms, which results in the failure of pre-calculating the lipid concentration, the E-EMT applicability would be compromised. The E-EMT might be also applicable for other dielectric models provided that those models can fit the complex permittivity of breast tissue or its experimental phantoms. In fact, our previous study demonstrated that the DD model could not effectively fit the dielectric properties of breast tissue with high content of fat and fiber in a broader frequency ranges from 0.1 THz to 2 THz and proposed application of a non-Debye model [45]. However, in this study, the DD proves to be sufficient for fitting the dielectric properties of the phantom samples in the examined range from 0.4 THz to 2 THz, which, as mentioned earlier, is the most reliable for analysis. Further study on the selection of applicable dielectric models and frequency range for fitting will be considered in our future work.

4. Conclusions

The EMT method with the proposed optimization approach has been shown to be useful for determining the composition profile of two-component breast tissue phantoms including intralipid and water-gelatin emulsions. However, this method proves to be not suitable for the three-component tissue phantoms as it fails to estimate the component concentrations of the phantoms those have low water content (e.g. less than 70% according to our current study).
The E-EMT method with the application of $\varepsilon_2$ from the DD model offers a more accurate approach for concentration analysis of the three-component phantoms as compared to the EMT. The estimated concentration values for all three components in these samples fit the true values effectively, especially those of the gelatin content which could not be accurately achieved by the EMT due to the complicity of the gelatin/water network formed in aqueous environment. The outcomes suggest that detection of a concentration change of less than 5% in the phantom composition is achievable. Nonetheless, the E-EMT method is limited due to its empirical nature which is heavily dependent on the designed phantom samples and system errors. As an initial study introducing such an analysis method, for simplification purposes, we do not take these limitations into account and only focus on the overall performance but will investigate their impact in further study that assesses the accuracy and resolution in details. Both of the proposed analysis methods show that the terahertz reflection spectroscopy is capable of determining composition of the breast tissue phantoms which consist of up to three main components. This supports the possibility of applying THz imaging to analyzing and specifying the breast tissue pathologies. Technological advances in THz systems together with improvements in the analysis methodologies will make further progress towards the application of THz imaging to BCS in the future.

**Funding**

Cancer Council Western Australia, Youngberg Women’s Cancer Research Fellowship (VPW); Australian Government through the Australian Research Council (Discovery Project, DP150100635); National Health and Medical Research Council (Development Grant APP1074894); UWA Research Collaboration Award funded by the New Century Endowment Fund.

**Acknowledgments**

The authors would like to thank Prof Adam Gibson, Medical Physics & Biomedical Engineering, University College London for useful discussions on developing the tissue phantoms used in this work.

**Disclosures**

The authors declare that there are no conflicts of interest related to this article.