

## APPLICATION OF AUTO REGRESSIVE MODELS OF WAVELET SUB-BANDS FOR CLASSIFYING TERAHERTZ PULSE MEASUREMENTS

XIAOXIA YIN,\* BRIAN W.-H. NG<sup>†</sup> and DEREK ABBOTT<sup>‡</sup>

*Center for Biomedical Engineering and  
School of Electrical and Electronic Engineering  
The University of Adelaide, SA 5005, Australia*

*\*xyin@eleceng.adelaide.edu.au*

*†bwng@eleceng.adelaide.edu.au*

*‡dabbott@eleceng.adelaide.edu.au*

BRADLEY FERGUSON

*Tenix Systems Pty Ltd.  
Mawson Lakes, SA 5095, Australia*

*School of Electrical & Electronic Engineering  
The University of Adelaide, Adelaide, SA 5005, Australia  
brad.ferguson@tenix.com*

SILLAS HADJILOUCAS

*Cybernetics, School of Systems Engineering  
The University of Reading  
Whiteknights, Reading, Berkshire, RG 6AY, UK  
s.hadjiloucas@reading.ac.uk*

Received 8 February 2007

Revised 12 June 2007

This paper presents an approach for automatic classification of pulsed Terahertz (THz), or T-ray, signals highlighting their potential in biomedical, pharmaceutical and security applications. T-ray classification systems supply a wealth of information about test samples and make possible the discrimination of heterogeneous layers within an object. In this paper, a novel technique involving the use of Auto Regressive (AR) and Auto Regressive Moving Average (ARMA) models on the wavelet transforms of measured T-ray pulse data is presented. Two example applications are examined — the classification of normal human bone (NHB) osteoblasts against human osteosarcoma (HOS) cells and the identification of six different powder samples. A variety of model types and orders are used to generate descriptive features for subsequent classification. Wavelet-based de-noising with soft threshold shrinkage is applied to the measured T-ray signals prior to modeling. For classification, a simple Mahalanobis distance classifier is used. After feature extraction, classification accuracy for cancerous and normal cell types is 93%, whereas for powders, it is 98%.

*Keywords:* Terahertz; Auto Regressive Moving Average (ARMA); Yule-Walker Algorithm; Prony's Method; Soft Threshold Wavelet Shrinkage De-noising; Discrete Wavelet Transform (DWT); Wavelet Packet Transform (WPT); Mahalanobis Distance Classifier.

## 1. Introduction

The terahertz (THz) part of the electromagnetic spectrum lying between the microwaves and infrared (100 GHz–10 THz) is of significant importance to the biological sciences because complementary information to traditional spectroscopic measurements on low-frequency bond vibrations, hydrogen bond stretches and torsions in liquids and gases may be obtained. The vibrational spectral characteristics of bio-molecules which lie in this range (wavenumbers between 3.33–333  $\text{cm}^{-1}$ ) make T-rays a promising sensing modality for clinical diagnosis. Since THz photons (or T-rays) have significantly lower energies (e.g. only 0.04 meV at 100 GHz) than other radiation used for medical diagnosis such as X-rays, it has been considered by many as non-invasive. Although nonlinear interactions between biological tissue and coherent THz radiation have been predicted by Fröhlich<sup>1</sup> and experimentally verified by the careful work of Grundler and the analysis by Kaiser<sup>2</sup> in the 1990s, the widely held view at the moment is that any measurement technique that operates at THz frequencies should be considered as non-invasive. Such a conclusion is based on assumptions that in the linear absorption processes involved when THz pulses interact with biological tissue, the Gibbs free energy conveyed in the THz light beam is insufficient to drive chemical reactions. For example, the molar energy at a frequency  $f$  of 100 GHz would be given from  $E = Nhf$  where  $N = 6.023 \times 10^{23} \text{ mol}^{-1}$  (Avogadro's number) and  $h = 6.626 \times 10^{-34} \text{ Js}$  (Planck's constant), the calculated value of only  $E = 0.04 \text{ kJ mol}^{-1}$  is so low (approximately 100 times lower than the amount of molar energy required for ATP hydrolysis) that for most practical purposes, we may assume that the interference with biochemical processes would be minimal.

Other advantages of performing imaging based on the optical properties of biological tissue with THz radiation are the better penetration length than infrared light, and lower scattering. Organ differentiation on the basis of tissue water content using microwave transmission or reflection measurements is impractical because the diffraction limited minimum spot size for a free-space beam is too large to avoid beam spillover around most tissues and organs. From a technological point of view, THz imaging needs to compete with positron emission tomography (PET) imaging, which has picomolar sensitivity but poor spatial resolution and magnetic resonance imaging (MRI) that offers millimolar sensitivity with high spatial resolution. Indeed, a diffraction limited imaging system operating at 1 THz would have a spatial resolution of 300  $\mu\text{m}$  which should be considered sufficient for many biomedical applications. Since the human body is 70% composed of water, a large part of the energy in the excitation pulse is attenuated, rendering advanced signal processing techniques for de-noising and feature extraction, an integral part of the work.

Recent advances in T-ray sources and detectors have made it possible to image opaque objects, and discriminate tumor cells from normal tissue.<sup>3</sup> While much effort has been devoted to improving the signal to noise ratio and repeatability of measurements as well as reliability in the function of the spectrometers, the

further processing of THz transients has only recently received some attention in the literature.<sup>4,5</sup> T-ray classification relies on observing changes in pulse amplitude, phase as well as dispersion characteristics of the tissue under study. The non-stationary nature of time-domain pulses obtained in T-ray spectrometry justifies their decomposition in the wavelet domain as it can provide better de-noising. Furthermore, compared to Fourier-based techniques, a wavelet decomposition of the experimental signal can provide better time-frequency localization characteristics<sup>6</sup> facilitating subsequent classification tasks.

This paper introduces a novel parametric modeling procedure to the wavelet-decomposed T-ray transient signals that improves feature extraction and classification. The success of the proposed algorithm in classifying human bone osteoblasts (HBO) against human osteosarcoma cells (HOS) as well as at differentiating between six types of powder samples is discussed. The choice of biological samples is made on the basis that there have been suggestions that THz transient spectrometry can be used for the early detection of cancerous tissue (the contrast mechanism being the enhanced blood circulation in the cancerous tissue). The reported work complements previous work in classifying basal cell carcinomas (a form of skin cancer), which has been conducted by researchers at the University of Cambridge and TeraView Limited.<sup>7</sup> It needs to be emphasized that the rationale for picking bone cancer cells is that techniques for culturing bone cells on a Petri dish are well-established and readily accessible. Our positive results with T-ray detection of cancerous bone cells will motivate future research to explore other classes of cancer cells. The motivation for using THz pulse transients for extracting information on densities, thicknesses and number of absorber molecules per unit volume in different powder samples stems from the fact that substance detection is an increasingly important area in the pharmaceutical (drug polymorphs and isomorphs<sup>8-10</sup>) as well as security industries (e.g. fingerprinting of explosives and illicit drug detection<sup>11,12</sup>). Our goal is to demonstrate a feature extraction methodology that is non-specific to the data sets in hand. This is of significant importance to the THz community as data driven classifiers prohibit proper inter-comparison between results obtained in different labs and therefore preclude the development of standards, guidelines and specifications that could be adopted by the biomedical, pharmaceutical as well as security sectors, which are envisaged to become emerging markets for THz-transient spectrometers.

In this sense, the requirement that our proposed algorithm should perform well in two very different classification tasks represents a departure from previous THz works presented in the literature. Section 2 briefly described the principle of terahertz imaging and the related hardware used to generate the experimental data sets whereas Sec. 3 provides the methodology used to perform the classification tasks. Section 4 presents the results from a Mahalanobis distance classifier and discusses issues related to the proposed feature extraction method. Section 5 highlights the important aspects of the work and provides directions for future research.

## 2. THz Spectroscopy and Imaging

The time-resolved THz spectrometer used in the reported studies utilizes a short coherence length infrared source (centered at around 800 nm) to generate a sub-100 femtosecond duration pulse train with repetition frequency of around 80 MHz. Each infrared pulse, is split into separate pump and probe beams. The pump beam is used to excite an optical rectification crystal, which acts as a T-ray emitter, and the T-rays produced (duration around 200 fs) are collimated and focused onto a sample by a pair of parabolic mirrors. The T-rays emerging from the sample are re-collimated by another pair of mirrors, before being combined with the probe beam in a T-ray detector crystal. As a result, the sample modified T-ray and the probe beams propagate through the THz detector crystal co-linearly. The pump beam, which is also transmitted through a chopper, travels through an optical delay stage that is modulated accordingly so that the pump and probe beams arrive at the detector in a time-coincident manner. By moving the delay line through the zero path difference of the two arms of the interferometer, the cross-correlation of the optical and THz signal is obtained. The electro-optic detector crystal produces an output which is proportional to the birefringence observed from the interaction of the THz pulse with the time-coincident infrared pulse replica within the crystal. This output is proportional to the T-ray response of the sample and this signal is measured with the use of an optical photodetection scheme. A lock-in amplifier is also used to demodulate the signal, this avoids  $1/f$  (flicker) noise problems, which are present in this detector-limited measurement scheme. Terahertz pulsed imaging (TPI) is achieved by performing a 2D raster scan after translating the sample in both the  $x$  and  $y$  directions while keeping it at the focal plane of the parabolic mirrors. A typical setup<sup>13</sup> is shown in Fig. 1.

A further advantage in using THz pulses instead of infrared as a measurement modality is the fact that the wavelengths are longer, and noise due to the motion of the translation stage is sufficiently small compared to the wavelength permitting the extraction of phase information. A limitation of current measurement techniques is the low power per spectral bin and consequently, the small signal to noise ratio in the measured complex insertion loss of the sample. It has previously been shown that a combination of wavelet transform techniques and statistical models can mitigate the effects of noise and extract effective features (frequency-dependent dispersion, attenuation and phase delay properties of the sample) for classification.<sup>14</sup> Normal human bone (NHB) osteoblast cells were obtained from patients and cultured from small pieces of trabecular bone for four to six weeks to obtain a confluent culture. Human osteosarcoma (HOS) cells were cultured from an immortalized cell line. The confluent culture was obtained within one week. A 25 ml polystyrene flask with flat bottom was used to culture the cells under a 5% carbon dioxide environment and a temperature of 37°C. In order to perform T-ray imaging, the rectangular flasks were tipped and placed in the T-ray ( $x-y$ ) translation stage. A T-ray image was obtained

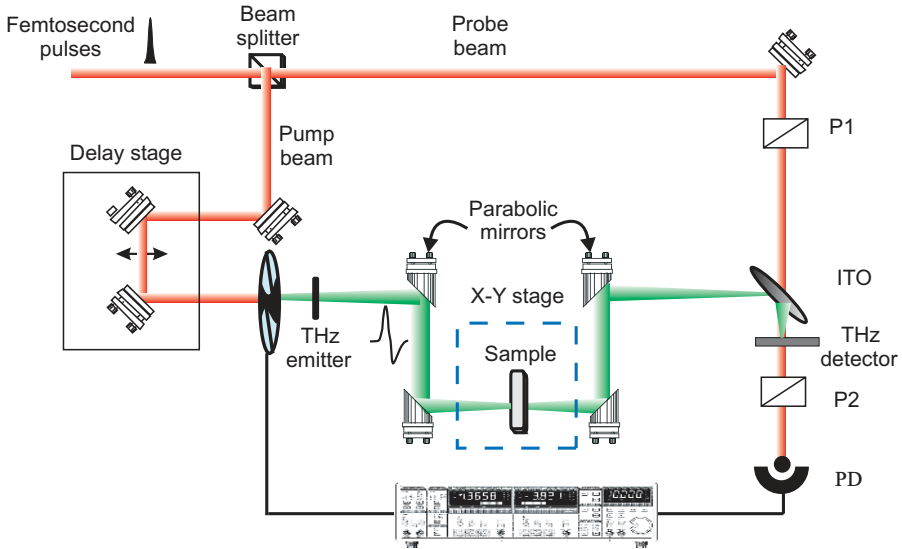


Fig. 1. Illustration of a femtosecond laser-based T-ray functional imaging system based on a pump-probe configuration. PD = infrared photodiode; P1, P2 = polarizers; ITO = indium tin oxide coated beam splitter.

at ten different positions to provide spectroscopic data, with a distance interval of  $50\ \mu\text{m}$ . The above procedure was performed for each of the three flasks — normal cells, cancerous cells and the container with media solution — and iterated a further five times until 50 pixels of T-ray responses were obtained for all three flasks.<sup>15</sup> This was regarded as a sufficient amount of data for a verification of classification effectiveness for different cells. For the powder sample classification experiment, six different powdered substances are used: sand, talcum, salt, powdered sugar, wheat flour, and baking soda with thicknesses of 2, 3 and 4 mm. The variations in thickness enabled investigation in thickness-independent classification. A seventh set of measurements of an empty sample holder was used as a reference to provide a background spectrum. The sample holder consisted of two teflon blocks separated by a translation-stage based control mechanism. This guaranteed a consistent powder density. A similar T-ray pulse imaging system is used to record the THz transients. The teflon sample holder was mounted on an  $x$ - $y$  translation stage. Since teflon is dispersionless and has a very low absorption coefficient at THz frequencies, there is minimal distortion of the T-ray pulse as it propagates through the holder. A 2D T-ray image of the sample was obtained via a raster scan; such an image allowed the effects of different scattering paths and minor variations in powder thickness and density to be observed. The data in this paper is obtained from a truncated raster scan with measurements taken from 50 co-linear locations, or pixels (with a spacing of  $100\ \mu\text{m}$ ). The integration time of 30 minutes per sample corresponds to an integration time of 36 seconds per pixel.

### 3. Wavelet De-noising, Parametization and Classification

#### 3.1. Wavelet de-noising

Fourier transformation of the TPI data described in the previous section provides the frequency dependent characteristics of a target sample. The sampled T-ray transients are a function of discrete time, to which the Discrete Wavelet Transform (DWT) is applied. To realize the DWT, simple digital filter banks are utilized in a recursive structure to calculate wavelet transform coefficients of T-ray signals. Formally, the representation of signals with their wavelet transform coefficients is known as a multi-resolution analysis (MRA). The theory underlying MRA allows a systematical method for constructing (bi-)orthogonal wavelets<sup>16</sup> and leads to the fast wavelet transform (FWT, also known as Mallat's algorithm).<sup>17,18</sup> In principle, wavelet-based techniques are very well suited to studies of non-stationary time-domain data sets, highlighting the variability of features at different time-frequency scales.<sup>6,19</sup> The main concern about the current T-ray TPI measurements is its corruption by different types of noise which limit the practical usefulness of this mode of imaging. One of the main problems is the distortion of the T-ray pulse as it propagates through the optical system as well as the sample. As a result, the received T-ray signals are strongly dependent on the acquisition conditions, and similar samples may produce variations in the measured signals in different portions of the image.<sup>14</sup> In order to keep image acquisition within realistic timescales, a short integration time per pixel is adopted as common practice, this makes the choice of the THz-transient de-noising process critical for the correct interpretation of TPI data. The use of perfect reconstruction quadrature mirror filter banks has been extensively discussed by Vaidyanathan<sup>20</sup> for the purpose of de-noising and generating bases of compact support. The works of Vetterli and Kovacevic<sup>21</sup> as well as that of Strang and Nguyen<sup>22</sup> further complemented the above, elaborating more on sub-band coding, Sherlock and Monro<sup>23</sup> discussed how to apply FIR filters of arbitrary length to describe the space of orthonormal wavelets, further parameterizing the wavelet coefficients at each decomposed level, and Tuqun and Vaidyanathan<sup>24</sup> proposed a state-space approach to the design of globally optimal FIR energy compaction filters. Since, in our work, there is no requirement for adopting an algorithm with a perfect reconstruction property, as our ultimate goal is feature extraction and classification, our constraints are more relaxed compared to those used in filtering or signal compression applications. Divine and Godtliebse<sup>25</sup> suggested that for feature exploration purposes, it is possible to assume stationarity over some time interval and smooth the wavelet spectrum along the time axis using an auto regressive (AR) model. Recently, Paiva and Galvão<sup>26</sup> also discussed a wavelet-packet decomposition tree algorithm that establishes frequency bands where sub-band models are created. Both approaches propose the modeling of the approximation and detail wavelet coefficients in order to further extract statistically significant features and a similar approach is adopted in our work. A typical de-noising procedure consists of decomposing the original signal using

the discrete wavelet packet transform (DWPT) or the discrete wavelet transform (DWT),<sup>19,27-29</sup> thresholding the detail coefficients, and reconstructing the signal by applying the appropriate inverse transform (IDWT or IDWPT, respectively). In our work, we adopted (i) the bior6.8 (DWPT) and (ii) the db20 (DWT) wavelet families for de-noising after comparing the classification results obtained using the following: db1, db8, db20, sym1, sym2, sym4, sym8, sym12, coif2, coif5, bior1.1, bior2.8 and bior6.8, and adopting a three-level decomposition. These wavelets are chosen as representatives for general classes of orthonormal, non-orthonormal and bi-orthogonal wavelets, respectively. For the de-noising of femtosecond THz transients, a three-level decomposition is usually sufficient<sup>3</sup> and unnecessary computational load associated with more decomposition levels can be avoided. The Stein's Unbiased Estimate of Risk (SURE) and the "heuristic" SURE methods<sup>6</sup> are used separately to estimate the soft threshold parameter ( $\lambda^S$ ) for the cancer and powder classification experiments, respectively.

**3.2. Formulation of Wavelet transform and heuristic SURE thresholding**

In our work, each time domain signature corresponding to data from a single pixel was represented by a data vector  $\mathbf{x}$  of length  $J$ , where the  $n$ th element of  $\mathbf{x}$ , denoted by  $x_n$ , represents the measured signal at the  $n$ th sampling instant. The filter bank transform can be regarded as a change in variables from  $\mathcal{R}^J$  to  $\mathcal{R}^J$  performed according to the following operation

$$w_m = \sum_{n=0}^{J-1} x_n v_m(n), \quad m = 0, 1, \dots, J - 1 \tag{3.1}$$

where  $w_m$  is a transformed variable and  $v_m(n) \in \mathcal{R}$  is a transform weight. It proves convenient to write the transform in matrix form as:

$$\mathbf{w}_{1 \times J} = \mathbf{x}_{1 \times J} \mathbf{V}_{J \times J} \tag{3.2}$$

where  $\mathbf{x} = [x_i : i = 0, 1, \dots, J - 1]$  is the row vector of original variables,  $\mathbf{w}$  is the row vector of new (transformed) variables and  $\mathbf{V}$  is the matrix of weights. Choosing  $\mathbf{V}$  to be unitary (that is,  $\mathbf{V}^T \mathbf{V} = \mathbf{I}$ ), the transform is said to be orthogonal and it therefore consists of a simple rotation in the coordinate axes (with the new axes directions determined by the columns of  $\mathbf{V}$ ). As described in the wavelet literature, the discrete wavelet transform can be calculated in a fast manner by using a finite-impulse-response (FIR) filter bank structure. In this filter bank, the low-pass filtering result undergoes successive filtering iterations with the number of iterations  $k$  chosen by the analyst. The final result of the decomposition of data vector  $\mathbf{x}$  is a vector resulting from the concatenation of row vectors  $\mathbf{c}_k$  and  $\mathbf{d}_k$  (termed respectively approximation and detail coefficients at the  $k$ th scale level) in

the following manner:

$$\mathbf{w} = [\mathbf{c}_k | \mathbf{d}_k | \mathbf{d}_{k-1} | \cdots | \mathbf{d}_1] \tag{3.3}$$

with coefficients in larger scales (e.g.  $\mathbf{d}_k, \mathbf{d}_{k-1}, \mathbf{d}_{k-1}, \dots$ ) associated with broad features in the data vector, and coefficients in smaller scales (e.g.  $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3, \dots$ ) associated with narrower features such as sharp peaks. Let  $h[0], h[1], \dots, h[2N - 1]$  and  $g[0], g[1], \dots, g[2N - 1]$  be the impulse responses of the low- and high-pass filters  $H$  and  $G$ , respectively. Assuming that filtering is carried out by circular convolution, the procedure for generating the approximation coefficients from the data vector  $\mathbf{x}$  consists of flipping the filtering sequence and moving it alongside the data vector. For each position of the filtering sequence with respect to the data vector, the scalar product of the two is calculated. Dyadic down-sampling is then performed to generate coefficients  $c[i]$ . The detail coefficients  $d[i]$  are obtained in a similar manner by using the high-pass filtering sequence. Filtering in the wavelet domain consists of changing some of the above elements of  $\mathbf{w}$  by applying soft thresholding so that a new vector  $\mathbf{w}_f$  is produced and then applying the inverse transform. A soft threshold operation with threshold  $\lambda^s$  is employed:

$$\mathbf{w}_f(\mathbf{x}, \lambda^s) = \text{sgn}(\mathbf{x}) \max(0, |\mathbf{x}| - \lambda^s). \tag{3.4}$$

A soft threshold has better continuity<sup>6,30</sup> than hard thresholding, while at the same time, provides shrinkage of the wavelet coefficients dominated by noise. The nonlinear shrinking of coefficients in the wavelet domain is a non-parametric method. We adopt the heuristic Stein’s Unbiased Risk Estimate (SURE)<sup>30</sup> algorithm, which is estimated adaptively at each wavelet decomposition level  $k$ . This operates on detail coefficients  $\mathbf{d}_k$ . Each detail coefficient  $\mathbf{d}_k$  is composed of elements of the signal  $\mathbf{d}_f$  as well as superimposed noise  $\eta_k$ , so that  $\mathbf{d}_k = \mathbf{d}_{f_k} + \eta_k$ .

The Sure approach assumes a vector  $\mu_k = (\mu_k[i] : i = 1, \dots, n)$  of length  $n = 2^{-k}J$ , where  $d_k[i] \sim \eta(\mu_k[i], 1)$  is multivariate normal observations with that mean vector. The soft threshold estimator  $\hat{\mu}$  of  $\mu$  is given from  $\hat{\mu} = \hat{\mu}(\mathbf{d}_k) = \mathbf{d}_k + \mathbf{g}(\mathbf{d}_k)$  with  $\mathbf{g}_k = (g_k[i])_{i=1}^n$ , a function from  $R^n$  into  $R^n$ . Assuming that  $\mathbf{g}_k(\mathbf{d}_k)$  is weakly differentiable, Stein’s method calculates the energy loss  $\|\hat{\mu} - \mu\|$  in an unbiased fashion using the following

$$E \|\hat{\mu}_k(\mathbf{d}_k) - \mu_k\|^2 = n + E\{ \|\mathbf{g}_k(\mathbf{d}_k)\|^2 + 2\nabla \cdot \mathbf{g}_k(\mathbf{d}_k) \} \tag{3.5}$$

where  $\nabla \cdot \mathbf{g}_k \equiv \sum_i \frac{\partial}{\partial d_k[i]} g_k[i]$   $E \|\hat{\mu}_k^{(\lambda_k^s)}(\mathbf{d}_k) - \mu_k\|^2$ .

By applying Stein’s results<sup>6,31</sup> at the  $k$ th decomposition level, we have

$$\text{SURE}_k(\lambda_k^s; \mathbf{d}_k) = n - 2 \cdot \#\{i : |d_k[i]| < \lambda_k^s\} + \sum_{i=1}^n \min(|d_k[i]|, \lambda_k^s)^2, \tag{3.6}$$

where  $\hat{\mu}_k$  denotes the soft threshold estimator  $\hat{\mu}_k[i]^{(\lambda_k^s)}(\mathbf{d}_k) = \eta_{\lambda_k^s}^s(d_k[i])$ . Thus the threshold  $\lambda^s$  that minimizes  $\text{SURE}_k(\lambda_k^s; \mathbf{d}_k)$  is obtained.<sup>30</sup> Here,  $\#$  means the number of the coefficients that are smaller than the threshold of noisy signals.



The de-noised T-ray signal in the wavelet domain can be re-written as

$$\mathbf{w}_f = [\mathbf{c}_{f_k} | \mathbf{d}_{f_k} | \mathbf{d}_{f_{k-1}} | \dots | \mathbf{d}_{f_1}]. \tag{3.7}$$

### 3.3. AR and ARMA parametrization of wavelet coefficients

#### 3.3.1. AR model parameter estimation

In the subsequent wavelet decomposition, we represent the new approximation and detail coefficients with  $\bar{\mathbf{c}}$  and  $\bar{\mathbf{d}}$ , respectively. At every level  $k$  of this new decomposition, we perform firstly autoregressive (AR) and then moving average (MA) modeling of the coefficients. We use the forward linear prediction<sup>32,33</sup> at the  $k$ th wavelet transform step given by the linear difference equation in the time domain:

$$\hat{\tilde{c}}_k^{AR}[n] = -a_k[1]\bar{c}_k[n-1] - a_k[2]\bar{c}_k[n-2] - \dots - a_k[P]\bar{c}_k[n-P] + w_k[n] \tag{3.8}$$

where  $\hat{\tilde{c}}_k^{AR}[n]$  represents the current prediction of wavelet approximation coefficients through AR modeling, and  $P$  represents the prediction order of the model. The modeling residual  $w_k[n]$  is assumed to be a white Gaussian process. The least-squares form of the Yule-Walker equation is used to estimate the corresponding AR model parameters:<sup>32,34</sup>

$$\mathbf{R}_{\bar{\mathbf{c}}_k} \mathbf{a}_k = [S_k \quad \mathbf{0}]^T \tag{3.9}$$

where,  $\mathbf{R}_{\bar{\mathbf{c}}_k}$  is the covariance matrix of the data,  $\mathbf{a}_k = [a_k[0], \dots, a_k[P]]^T$  is the vector of the AR model coefficients and  $S_k$  is the minimum sum of squared errors. The data vector  $\bar{\mathbf{c}}_k$  of the autocorrelation method is determined by the selection of the end points as  $n_I = 0$ ,  $n_F = N_s + P - 1$ , and  $N_s$  is selected to be the length of the detailed wavelet coefficient at the  $k$ th wavelet transform level, with  $N_s = n = 2^{-k}J$ .

The prediction error variance (PEV), which is used for validation purposes of AR model, is calculated as

$$\sigma_{\epsilon^{k,P}}^2 = \frac{1}{n_F - n_I + 1} S_k. \tag{3.10}$$

It should be noted that the prediction error variance is the theoretic criterion for any AR/ARMA model order selection task. It measures the precision of a model's predictions. The measurement errors are reduced by the model fitting process if the PEV is less than 1, while any errors in the data measurements are multiplied if the PEV is greater than 1. It is desirable for PEV to be close to zero, in which case, the predictive power of the model will be more accurate. This is a crucial first step to estimate the model parameters.<sup>35</sup>

#### 3.3.2. ARMA model parameter estimation

An ARMA model is a combination of AR and moving average (MA) models. The MA model at the  $k$ th wavelet decomposition level, based on approximation

coefficients  $\bar{c}_k$  is developed from the following difference equation:

$$\hat{c}_k^{MA}[n] = \bar{c}_k[0]w_k[n] + b_k[1]w_k[n - 1] + \dots + b_k[Q]w_k[n - Q]. \tag{3.11}$$

The MA prediction output  $\hat{c}_k^{MA}[n]$  is seen as a sum of weighted average of the past  $Q_k$  input samples of a white noise process  $w_k(n)$ . The full ARMA model can be obtained by combining both AR and MA models, as below:

$$\sum_{i=0}^{P_k} a_k[i]\bar{c}_k[n - i] = \sum_{j=0}^{Q_k} b_k[j]w_k[n - j] \tag{3.12}$$

where,  $a_k[0]$  is normalized to 1.

The MA parameters are estimated by Durbin’s method, which is an improvement of the basic Prony method that assumes that  $\hat{c}_k[n] = \bar{c}_k[n]$  for  $n = 0, 1, \dots, N_s - 1$ , setting  $N_s = P_k + Q_k + 1$  with  $P_k = Q_k$ . The idea behind Durbin’s method is to turn an MA modeling problem into a set of two normalized and significantly over-parameterized AR modeling problems denoted by A and B and to consequently solve them (i.e. adopting the condition  $N_s > P_k + Q_k + 1$ ). The purpose of this over-parameterization process is to facilitate the solution. In our work, the order of AR model ( $P_k$ ) is selected to be five times the MA model order  $Q_k$  ( $P_k = 5Q_k$ ) as this has been found to provide a good fit.<sup>32</sup> Then the Yule-Walker matrix equation of the ARMA model is partitioned as:

$$[\mathbf{R}_{\bar{c}_{k_B}} \quad \mathbf{R}_{\bar{c}_{k_A}}]^T \mathbf{a}_k = [\mathbf{b}_k \quad \mathbf{e}_{k_A}]^T \tag{3.13}$$

where  $\mathbf{b}_k = [b_k[0] \dots b_k[Q]]$ . Here  $\mathbf{e}_{k_A}$  is the prediction error that is defined as  $\mathbf{e}_{k_A} = \bar{c}_k[n] - \hat{c}_k[n]$ , with  $n = 0, 1, \dots, N_s - 1$ . The related AR coefficient vector  $\mathbf{a}_k$  is determined from minimizing  $S_{k_A} = |\mathbf{e}_{k_A}|^2$  and observing that  $[\mathbf{R}_{\bar{c}_{k_A}}] \mathbf{a}_k = \mathbf{e}_{k_A}$  is equivalent to the Yule-Walker equations for the AR model in Eq. (3.9). The difference between the two formulations lies in the calculation of MA parameters  $\mathbf{b}_k$ . Instead of forcing  $\mathbf{b}_k$  to match the left side of the equation  $[\mathbf{R}_{\bar{c}_{k_B}}] \mathbf{a}_k = [\mathbf{b}_k]$  in Eq. (3.13), Durbin’s method considers the existence of an error term of this equation. A higher order AR model is used as a substitute for an MA model. It is worth noting that (i) the coefficient vector  $\mathbf{a}_k$  is obtained from the solution of the Yule-Walker equation normalized by the gain — first item of AR coefficients, before substituting  $S_k$  with unity in Eq. (3.9) and (ii) we use the coefficient vector  $\mathbf{a}_k$ , derived from setting  $P_k = 5Q_k$ , to construct the data matrix that provides the new coefficients of AR modeling with the MA model order  $Q_k$ . Durbin’s method guarantees the stability of the ARMA model eliminating windowing effects.

### 3.4. Feature extraction

The objective of feature extraction is to isolate the relevant features mentioned earlier from the T-ray signals to improve classification performance. Feature matrix calculations, assuming AR, MA and ARMA models of different order, are performed.

The averages of the modeling coefficients are computed over the three decomposition levels of the wavelet transform employed on each dataset. The model coefficient averages are then joined to produce feature vectors with a dimension equal to the number of subbands in the wavelet decomposition. The feature vectors obtained from two different AR orders, and MA orders are combined, respectively, to form the final AR and MA feature matrices. The ARMA feature matrix is obtained by combining two different orders of AR and MA vectors together. The extracted AR and MA feature vectors are calculated at each decomposition level  $k$  from

$$\langle \mathbf{a}_k^{P^*} \rangle = \frac{1}{P^*} \sum_{i^*=1}^{P^*} a_k[i^*], \quad \text{where } 2 \leq P^* \leq 7 \quad (3.14)$$

$$\langle \mathbf{a}_k^P \rangle = \frac{1}{P} \sum_{i=1}^P a_k[i], \quad \text{where } 3 \leq P \leq 8 \quad (3.15)$$

$$\langle \mathbf{b}_k^{Q^*} \rangle = \frac{1}{Q^*} \sum_{j^*=1}^{Q^*} b_k[j^*], \quad \text{where } 2 \leq Q^* \leq 7 \quad (3.16)$$

$$\langle \mathbf{b}_k^Q \rangle = \frac{1}{Q} \sum_{j=1}^Q b_k[j], \quad \text{where } 3 \leq Q \leq 8 \quad (3.17)$$

where  $\langle \mathbf{a}_k^P \rangle$ ,  $\langle \mathbf{a}_k^{P^*} \rangle$ , and  $\langle \mathbf{b}_k^Q \rangle$ ,  $\langle \mathbf{b}_k^{Q^*} \rangle$  denote the averaged value (DC value) of AR and MA model coefficients at different model orders. The new feature matrix using the AR modeling procedure is given from

$$\mathbf{DC}_k^{AR} = [\langle \mathbf{a}_k^{P^*} \rangle | \langle \mathbf{a}_k^P \rangle], \quad \text{where } P^* \leq P. \quad (3.18)$$

A similar procedure is adopted for the MA modeled datasets

$$\mathbf{DC}_k^{MA} = [\langle \mathbf{b}_k^{Q^*} \rangle | \langle \mathbf{b}_k^Q \rangle], \quad \text{where } Q^* \leq Q. \quad (3.19)$$

Finally, we obtain new feature matrices combining the coefficients from both the AR and MA modeling procedures

$$\mathbf{DC}_k^{ARMA} = [\langle \mathbf{a}_k^{P^*} \rangle | \langle \mathbf{b}_k^Q \rangle]. \quad (3.20)$$

This combination aims at improving the generalization ability of the classification process.

The complete procedure for calculating  $\mathbf{DC}_k^{ARMA}$  is depicted in Fig. 2. The implementation used in this paper is motivated by a desire to obtain fixed length feature vectors to facilitate comparison. Since various AR/ARMA model orders are used, this naturally leads to a varying number of model coefficients, and hence affect the dimensionality of the extracted feature vectors. Averaging the AR/ARMA coefficients — transforms the varying dimension of AR/ARMA model coefficients to a fixed number of dimensions. Another parameter that can affect the feature matrix length is the number of decomposition levels in wavelet transforms. Varying  $k$ , leads to a different number of sub-bands. Figure 3 illustrates the preprocessing steps as well as the adopted feature extraction algorithm.

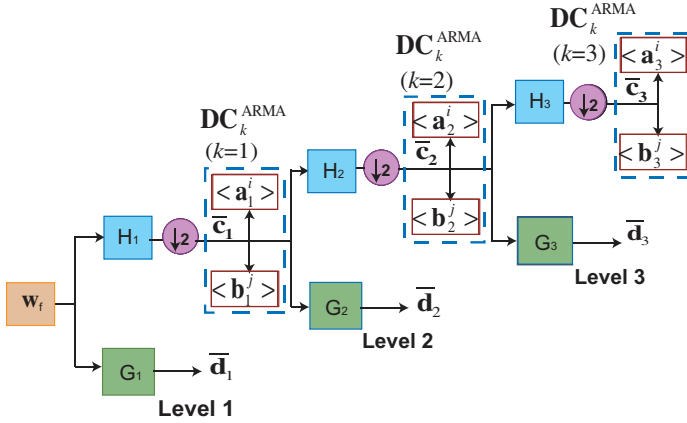


Fig. 2. Illustration of the  $DC^{ARMA}$  feature matrix calculated over three levels of wavelet decomposition, where  $H$  and  $G$  denote the low- and high-pass filters, respectively,  $w_f$  is the de-noised T-ray measurement. The arrow indicates the diadic downsampling process. Similar illustration related to  $DC^{AR}$  and  $DC^{MA}$  feature matrix are assumed.

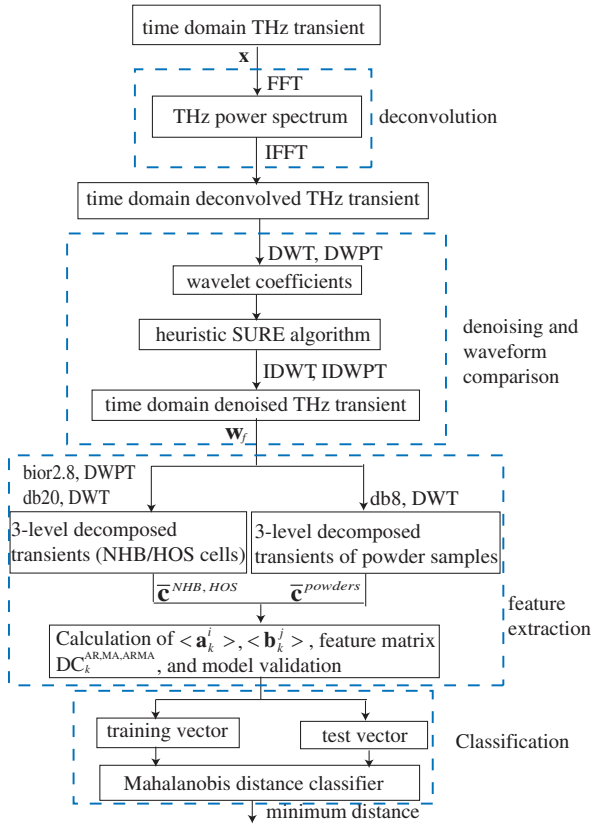


Fig. 3. Illustration of the current classification algorithm using AR/ARMA modeling and wavelet decomposition.

### 3.5. Classification

The classification is realized using a Mahalanobis distance classifier.<sup>36</sup> The Mahalanobis distance is defined as the distance from the mean value of the pointed class to a given point, where the given class is normalized by the different training vectors along the different directions.<sup>36</sup> For a given class,  $m$ , the distance from a feature matrix  $\mathbf{DC}_k^l$  to the class mean  $\alpha_m$ , is defined as

$$\rho_m(X) = \sqrt{(\mathbf{DC}_k^l - \alpha_m)^T \mathbf{C}^{-1} (\mathbf{DC}_k^l - \alpha_m)} \quad (3.21)$$

where  $\mathbf{C}$  is covariance matrix of the feature vectors,  $\mathbf{DC}_k^l$  with  $l = 1, 2, 3$  representing the averaged coefficients matrix related to AR ( $l = 1$ ), MA ( $l = 2$ ), and ARMA ( $l = 3$ ) modeling of wavelet approximation coefficients at three decomposition levels, that is,  $\mathbf{DC}_k^1$  being  $\mathbf{DC}_k^{AR}$ ,  $\mathbf{DC}_k^2$  being  $\mathbf{DC}_k^{MA}$ ,  $\mathbf{DC}_k^3$  being  $\mathbf{DC}_k^{ARMA}$ . In practice, the covariance matrix is estimated from the training vectors. During classification, the minimum Mahalanobis distance from feature vector  $\mathbf{DC}_k^l$  to each class center  $\alpha_m$  is used to assign the appropriate class label.

## 4. Results and Discussion

### 4.1. Wavelet preprocessing

Figure 4 illustrates the performance of wavelet de-noising, after adopting the WP SURE de-noising procedure for both normal (NHB) and cancerous cells (HOS) after performing de-convolution to eliminate the features due to the container. The background reference signal is used for the de-convolution processing, which isolates the TPI sample response from that of the container. The de-convolution procedure is performed by dividing the respective Fourier transforms of the sample

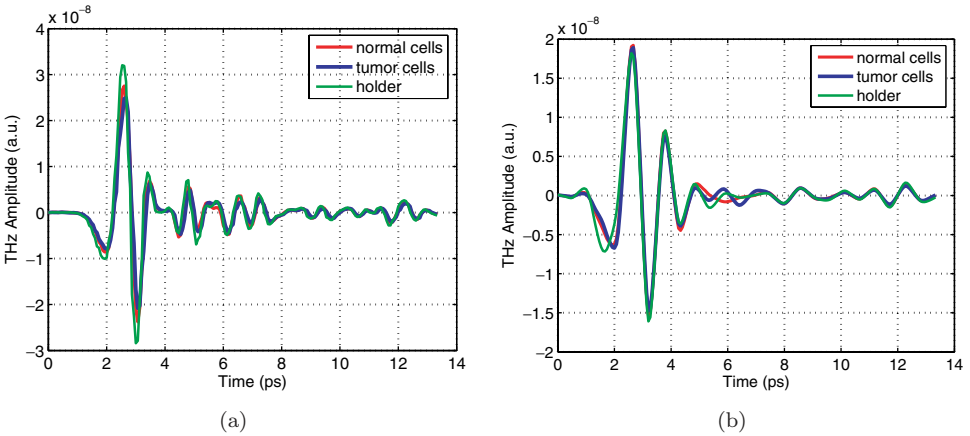


Fig. 4. Comparisons of the time domain signals related to a single pixel of T-ray CT images for a normal cell (NHB) and a cancerous cell (HOS) together with the empty container (holder) (a) before and (b) after employing WP SURE de-noising of bior6.8 wavelet.

signature with that of the background time domain signatures that corresponds to signal pixel.

The comparisons of the signals in the time domain are performed for one pixel of normal and cancerous cell responses together with their containers before and after employing WP SURE de-noising. The effectiveness of wavelet package SURE soft threshold shrinkage de-noising is demonstrated in the three visually separable T-ray pulsed responses of Fig. 4b, in comparison with the original measured T-ray signals of Fig. 4a.

**4.2. Classification of NHB and HOS cells**

For the classification of cancer cells, the averages of two different orders of AR coefficients are extracted for use as features. Wavelet transform depths of 2 and 3 are compared, and the biorthogonal spline wavelet of order 6.8 is used to compute the discrete wavelet packet transform. A Mahalanobis distance classifier is trained using a portion of the T-ray responses corresponding to each of the three classes — cancer cells, normal cells and the reference signals — then the remainder of the pixel responses are classified. The resultant classification accuracy is used to measure performance. In addition, error prediction covariance at different orders of AR modeling and the different levels of wavelet transform are calculated and employed as another metric for comparison.

Table 1 shows squared error variances of seven orders of AR model corresponding to the approximation coefficients generated using a bior6.8 wavelet family with three levels of wavelet decomposition on an arbitrarily chosen T-ray response. The columns of this table correspond to the AR model order, while the rows correspond to the levels of DWPT prior to calculating the AR models. Data are scaled by a factor of  $10^{-16}$ , demonstrating very successful modeling. AR models are practically error-free as a result of the limitations of our computation hardware. Similar observations can be made for values in Tables 3 and 4. It can be seen that the value of the error variance decreases with increasing AR order. Increasing the DWPT depth tends to increase the value of the error variance. The absolute error variance spans over two orders of magnitude, demonstrating that the choice of number of decomposition levels and the AR model order adopted can have a significant effect on the validation process of the model.

Table 1. Squared error variances of AR models fitted to DWPT sub-bands using cell samples (results scaled by a factor of  $10^{-16}$ ).

DWPT depth	AR model order						
	2	3	4	5	6	7	8
1	0.01155	0.00767	0.00759	0.00722	0.00426	0.00318	0.00316
2	0.1596	0.1403	0.0763	0.0685	0.0512	0.0480	0.0435
3	0.2202	0.2032	0.2013	0.2011	0.2010	0.2010	0.2010

Table 2. Percentage classification accuracy of T-ray pulses traveling through cell samples. The range of AR model orders considered is 2 to 8; three levels of DWPT with a bior6.8 wavelet are used throughout. Above the diagonal of this table are results from Method (A); below the diagonal are results from Method (B). The values in bold show the best and worst performances.

AR order	AR order						
	2	3	4	5	6	7	8
2		88.9	91.7	90.3	91.7	87.5	87.5
3	<b>90.6</b>		90.3	<b>93.1</b>	86.1	87.5	88.9
4	87.7	<b>90.6</b>		83.3	88.9	84.7	87.5
5	87.0	<b>90.6</b>	84.1		86.1	80.6	87.5
6	87.7	<b>90.6</b>	87.7	84.8		86.1	84.7
7	87.7	86.2	86.2	83.3	82.6		<b>79.2</b>
8	87.0	85.5	84.8	84.1	86.2	<b>71.7</b>	

Table 2 shows the classification accuracy results obtained from two classification methods, which are labeled Method (A) and Method (B) for notational simplicity. In Method (A), the classifier is trained using half of the pixel responses (chosen in an arbitrary manner) for the three data classes (normal, cancerous cells and reference TPI responses), and tested by the remaining half of the pixel responses. In Method (B), a leave-one-out (LOO) error estimator is used. The leave-one-out estimator uses the Mahalanobis distance classifier which is trained using a randomly selected set of  $N - 1$  responses, out of a total of  $N$  responses for the three classes and tested by the remaining response. The procedure is repeated  $N$  times to obtain the average classification accuracy. In the experiments, all pairwise combinations of AR orders are considered, which implies a total of  $\binom{7}{6} = 42$  combinations. From Table 2, it can be seen that the classification accuracies obtained were in the 79.2% to 93.1% range for Method (A) and 71.2% to 90.6% range for Method (B). Considering the results in Tables 1 and 2 together, it can be observed that there is a trade off between the classification accuracy and model complexity. Combinations of lower-order AR models generally outperformed combinations with high-order models. The best classification accuracy is achieved by performing a combination of the features generated from AR models of orders 3 and 5. The classification results after adopting the LOO method reaches 90.85%.

### 4.3. Classification of powder samples

In the powder classification experiment, the responses are obtained from samples of different thicknesses: 2, 3 and 4 mm. All the data from the six-classes of powders plus the background reference signal at thicknesses of 2 and 4 mm are used to train the classifier, and all the data of powder samples at a thickness of 3 mm are used to test the classifier. After wavelet de-noising using the SURE procedure, a db8 discrete wavelet transform (DWT) is applied to the de-noised powder data. Subsequently, ARMA modeling is performed to extract the relative feature matrix. The performance of Durbin’s algorithm for ARMA modeling is characterized by

Table 3. Squared error variances of ARMA fitted to DWT sub-bands on T-ray signals of powder samples (scaled by a factor of  $10^{-17}$ ).

DWPT depth	ARMA model order						
	2	3	4	5	6	7	8
1	0.2405	0.1397	0.1337	0.1147	0.0948	0.0935	0.0797
2	4.458	2.854	1.280	0.976	0.674	0.673	0.660
3	32.09	32.06	26.97	26.93	24.35	24.06	23.06

Table 4. Squared error variances of ARMA obtained from three different methods: correlation, Prony’s and Durbin’s, as applied on powder samples.

Method	Correlation	Prony’s	Durbin’s
Minimum value	$0.3399 \times 10^{-19}$	$0.3581 \times 10^{-4}$	$0.797 \times 10^{-17}$
Maximum value	$0.1153 \times 10^{-16}$	0.0115	$0.3209 \times 10^{-15}$

the squared error variances. These results are presented in Table 3. The correlation method allows the lowest error variance among the methods investigated, shown in Table 4. The error variance of Durbin’s method is much lower than the basic Prony’s method, though a little higher than that of the correlation method. However, the low variance achieved by the correlation method comes at a cost of a larger model order (AR only).

The ARMA model produces features that separate the classes quite effectively. The classification accuracy in Table 5 improves by at least 3% over either AR or MA models alone, as illustrated in Tables 6 and 7, respectively. Compared to the correlation and normal Prony’s methods used to fit an AR or ARMA model, that produced a maximum classification accuracy of 96% and 95%, respectively, Durbin’s algorithm improves the maximum classification accuracy by 2%.

Figure 5 shows the scatter plots of the feature vectors for 700 random samples at thicknesses of 2 and 4 mm. In order to better understand the learning vectors and show a better grouped dataset for each powder sample, we choose the second- and fifth-order ARMA model at three levels of wavelet decomposition to generate the scatter plots. The seven classes (comprising the six powder samples and a reference)

Table 5. Percentage classification accuracy of several powder samples, obtained using an AR modeling matrix.

AR order	AR order						
	3	4	5	6	7	8	
2	88.8	75.9	96.1	92.7	95.0	92.7	
3		85.4	89.1	88.8	81.8	96.9	
4			88.2	96.6	85.2	96.9	
5				81.2	75.6	86.0	
6					75.9	91.3	
7						75.63	



Table 6. Percentage classification accuracy of several powder samples, obtained using an MA modeling matrix.

MA order	MA order					
	3	4	5	6	7	8
2	84.9	84.9	81.5	84.9	85.2	84.0
3		85.7	78.7	72.0	77.0	79.8
4			73.4	79.8	80.1	81.0
5				72.0	81.0	79.3
6					70.6	77.9
7						73.1

Table 7. Percentage classification accuracy of several powder samples, obtained using an ARMA modeling matrix.

AR order	MA order					
	3	4	5	6	7	8
2	95.8	98.0	96.1	95.5	96.6	96.9
3		85.7	88.0	87.4	85.2	83.8
4			92.6	92.4	92.7	89.9
5				81.2	80.1	82.1
6					77.0	84.3
7						78.2

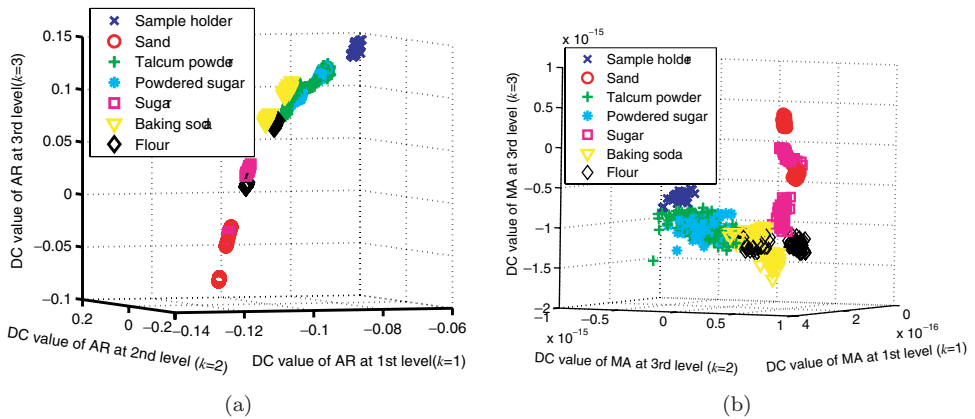


Fig. 5. Scatter plots showing the learning vectors for discriminating powders with the thickness of 2 and 4 mm, corresponding to (a) second order AR model and (b) fifth order of MA model, over three steps of wavelet transform.

in Fig. 5a, which uses AR model coefficients, are grouped together, although the class populations show some degree of overlap. The scatter plot in Fig. 5b, which uses MA model coefficients, is not clearer than the case with the AR model coefficients. However, it is found that the combination of the two model coefficients to generate an ARMA feature matrix at the two specified orders actually led to good classification performance, with classification accuracy of 96%.

The ARMA classification experiments are performed on a computer equipped with a 2.4 GHz Pentium 4 CPU. The average time spent classifying the six-classes powdered samples and their reference is 51.2 seconds for 42 runs, or 1.2 seconds per run. It is therefore feasible to use this ARMA modeling algorithm to perform real-time classification of pulses.

## 5. Conclusion and Future Work

The use of an orthogonal transform such as the wavelet transform is well justified for decomposing the time-domain signals obtained using a THz-transient spectrometer. De-noising in the wavelet domain is isometric to de-noising in the time domain so the adopted methodology is appropriate to applications where legislation of the proposed algorithm is necessary. The adoption of SureShrink thresholding is also appropriate as it is smoothness adaptive (the level-dependent thresholds are arrived at after regarding the different resolution levels of the transform as independent multivariate normal estimation problems). The further modeling of the wavelet coefficients to produce new feature vectors has been shown to produce new feature vectors rich in information content irrespective to the morphology of the samples. The use of Durbin's algorithm in estimating the ARMA coefficients at each decomposition level yielded small error variances ensuring that features present in the wavelet domain are not lost with this modeling process. The classification performance obtained for both the cellular samples as well as the powder samples using the Mahalanobis distance classifier is encouraging and provides further motivation to use more elaborate data-driven schemes in future work such as multi-layer perceptrons or support vector machine classifiers. It is worth noting, however, that classification work using an SVM classifier scheme without adopting a wavelet decomposition step, but using four-dimensional feature set based on amplitude and phase data has had a 89% success rate, at correctly classifying cancer cells.<sup>37</sup>

The adoption of a THz imaging system in a histopathology lab for the identification of cancerous cells, in conjunction with our classification methodology which is capable of differentiating between small differences in the transmission properties of the samples under study, is expected to provide significant advantages as current techniques are laborious, expensive and the results obtained are subjective to human interpretation. A further advantage of the adopted methodology is that it can be implemented in real-time so that measurements obtained with Terahertz-pulse imaging systems by practitioners from other disciplines can be displayed directly as feature maps in a manner that is transparent to the user. The work described will lead to the further proliferation of THz transient spectrometers in the biomedical, security and pharmaceutical technology sectors. The ultimate aim of our work is to achieve heterogenous layer classification in a three-dimensional space based on T-ray tomographic imaging. Work is ongoing to transfer the classification

methodology developed in this work to results from *in vivo* and *ex vivo* cancer cells that must be discriminated *in situ* through a number of heterogeneous layers. More recent work from our lab<sup>38</sup> in 3D tomographic classification of homogeneous layers has demonstrated that such goals can be achieved.

## Acknowledgments

Xiaoxia Yin acknowledges financial support from the Roger Pysden Memorial Fellowship for 2006, Australia Business, ABL State Chamber.

## References

1. Fröhlich H, The biological effects of microwaves and related questions, *Adv Electron Electron Phys* **53**:85–152, 1980.
2. Grundler W, Kaiser F, Experimental evidence for coherent excitations correlated with cell growth nanobiology, *Nanobiology* **1**:163–176, 1992.
3. Hadjiloucas S, Galvão RKH, Becerra VM, Bowen JW, Martini R, Brucherseifer M, Pellemans HPM, Haring Bolivar P, Kurz H, Chamberlain JM, Comparison of state space and ARX models of a waveguide's THz transient response after optimal wavelet filtering, *IEEE Trans Microw Theory Tech MTT* **52**(10):2409–2419, 2004.
4. Löffler T, Siebert K, Czasch S, Bauer HG, Roskos T, Visualization and classification in biomedical terahertz pulsed imaging, *Phys Med Biol* **47**:3847–3852, 2002.
5. Yin XX, Ng BW-H, Ferguson B, Mickan SP, Abbott D, One dimensional wavelet transforms and their application to T-ray pulsed signal identification, in *Proceedings of SPIE Photonics: Design, Technology and Packaging*, Vol. 6038, Brisbane, Australia, pp. 499–509, 2006.
6. Donoho DL, De-noising by soft thresholding, *IEEE Trans Inf Theor* **41**(3):613–627, 1995.
7. Woodward RM, Cole B, Wallace VP, Arnone DD, Pye R, Linfield, EH, Pepper M, Terahertz pulse imaging in reflection geometry of human skin cancer and skin tissue, *J Invest Dermatol* **120**:3853–3863, 2002.
8. Strachan CJ, Taday PF, Newnham DA, Gordon KC, Zeitler JA, Pepper M, Rades T, Using terahertz pulsed spectroscopy to quantify pharmaceutical polymorphism and crystallinity, *J Pharm Sci* **94**:837–846, 2005.
9. Zeitler JA, Taday PF, Newnham DA, Pepper M, Gordon KC, Rades T, Terahertz pulsed spectroscopy and imaging in the pharmaceutical setting, *J Pharm Pharmacol* **59**(2):209–223, 2007.
10. Watanabe Y, Kawase K, Ikari T, Ito H, Ishikawa Y, Minamide H, Spatial pattern separation of chemicals and frequency-independent components by terahertz spectroscopic imaging, *Appl Optics* **42**(28):5744–5748, 2003.
11. Federici JF, Schulkin B, Huang F, Gary D, Barat R, Oliveira F, Zimdars D, THz imaging and sensing for security applications-explosives, weapons and drugs, *Semicond Sci Technol* **20**:S266–S280, 2005.
12. Kawase K, Ogawa Y, Minamide H, Ito H, Terahertz parametric sources and imaging applications, *Semicond Sci Technol* **20**:S258–S265, 2005.
13. Ferguson B, Wang S, Zhong H, Abbott D, Zhang X-C, Powder detection with T-ray imaging, in *Proceedings of SPIE Terahertz for Military and Security Applications*, Vol. 5070, pp. 7–16, 2003.

14. Ferguson B, Abbott D, De-noising techniques for terahertz responses of biological samples, *Microelectron J* **32**(12):943–953, 2001.
15. Ferguson B, Liu H, Hay S, Findlay D, Zhang X-C, Abbott D, *In vitro* osteosarcoma biosensing using THz time domain spectroscopy, in Neev J, Reed M (eds.), *Proceedings of SPIE BioMEMS and Nanotechnology*, Vol. 5275, Bellingham, Australia, pp. 304–316, 2004.
16. Daubechies I, Orthonormal bases of compactly supported wavelets, *Comm Pure Appl Math* **41**(7):909–996, 1988.
17. Qian SE, *Time-Frequency and Wavelet Transforms*, 1<sup>st</sup> ed, Prentice Hall, Inc., New Jersey, USA, 2002.
18. Mallat SG, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans Pattern Anal Mach Intell* **14**(7):674–693, 1989.
19. Mallat SG, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 1999.
20. Vaidyanathan PP, *Multirate Systems and Filter Banks*, Prentice Hall PTR, New Jersey, USA, 1993.
21. Vetterli M, Kovacevic J, *Wavelets and Subband Coding*, Prentice-Hall PTR, New Jersey, 1995.
22. Strang G, Nguyen T, *Wavelets and Filter Banks*, 1<sup>st</sup> ed, Wellesley-Cambridge Press, Wellesley, USA, 1996.
23. Sherlock BG, Monro DM, On the space of orthonormal wavelets, *IEEE Trans Signal Process* **46**(6):1716–1720, 1998.
24. Tuqun J, Vaidyanathan PP, A state-space approach to the design of globally optimal FIR energy compaction filters, *IEEE Trans Signal Process* **48**(10):2822–2838, 2000.
25. Divine DV, Godtlielsen F, Bayesian modeling and significant features exploration in wavelet power spectra, *Nonlin Process Geophys* **14**:79–88, 2007.
26. Paiva HM, Galvão RKH, Wavelet-packet identification of dynamic systems in frequency subbands, *Signal Process Arch* **86**(8):2001–2008, 2006.
27. Daubechies I, *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992.
28. Jensen A, la Cour-Harbo A, *Ripples in Mathematics: The Discrete Wavelet Transform*, Springer-Verlag, Berlin, 2001.
29. Percival DB, Walden AT, *Wavelet Methods for Time Series Analysis*, Cambridge University Press, Cambridge, England, 2000.
30. Rangarajan R, Venkataramanan R, Shah S, Imaging denoising using wavelet (wavelets and time frequency), <http://bbs.matwav.com/index.jsp> (Accessed on August 19, 2005).
31. Johnstone MI, Donoho LD, Adapting to unknown smoothness via wavelet shrinkage, *J Stat Assoc* **90**(432):1220–1224, 1995.
32. Therrien CW, Oppenheim AV, *Discrete Random Signals and Statistical Signal Processing*, Prentice Hall, New Jersey, USA, 1992.
33. Proakis JG, Manolakis DG, *Digital Signal Processing: Principles, Algorithms, and Applications*, Prentice-Hall, Inc., NJ, USA, 1996.
34. Jain S, Deshpande G, Parametric modeling of brain signals, *IEEE Biotechnol Bioinf* **12**(15):85–91, 2004.
35. Liang G, Wilkes DM, Cadzow JA, ARMA model order estimation based on the eigenvalues of the covariance matrix, *IEEE Trans Signal Process* **41**(10):3003–3009, 1993.
36. Schürmann J, *Pattern Classification: A Unified View of Statistical and Neural Approaches*, John Wiley and Sons, Inc., New York, 1996.

37. Withayachumnankul B, Ferguson B, Rainsford T, Findlay D, Mickan SP, Abbott D, T-ray relevant frequencies for osteosarcoma classification, in *Process of SPIE Photonics: Design, Technology, and Packaging II*, Adelaide, Australia, 2005.
38. Yin X-X, Ng BW-H, Ferguson B, Mickan SP, Abbott D, 2-D wavelet segmentation in 3-D T-ray tomography, *IEEE Sensors J* **7**(3):342–343, 2007.