THz-TDS Signal Analysis and Substance Identification

Weixin Xie\textsuperscript{1,2} Jing Li\textsuperscript{1,2} Jihong Pei\textsuperscript{2}

\textsuperscript{1} School of Electronic Engineering, Xidian University, Xi’an 710071, P. R. China; \textsuperscript{2} Intelligent Information Institute, Shenzhen University, Shenzhen, 518060, P. R. China

E-mail: jhpei@szu.edu.cn

Abstract. The terahertz time-domain spectroscopy (THz-TDS) imaging system can obtain high-dimensional signals of the substance fingerprint information. It is necessary to process properly to use some signal processing techniques especially for high dimensional signals. As a mathematical description language, geometric algebra (GA) provides not only a powerful algebraic framework for the multi-dimensional vector analysis and computing, but also a unified measurement and geometrical description for different geometric models. On the basis of the GA theories, a new signal analysis method of the THz-TDS is presented. Based on the characteristics of THz-TDS signals, signals are mapped into vectors in the high-dimensional real vector space. The vectors are represented with hyper-complex numbers. We can analyze the vectors using theories of GA. Based on the physical mechanism of the THz-TDS signal analysis, geometric distribution properties and algebraic relationships of THz-TDS signals are deduced. It is demonstrated that every complex refractive index of the sample relates to a unique 2-blade $B_2$, all vectors corresponding to the samples of the same substance are collinear and belong to the intrinsic 2-blade of the substance. In projective interpretation, the 2-blade $B_2$ represents a fixed line and all vectors related to the same substance are along that line. Accordingly, a novel substance identification method based on the relative THz-TDS is presented. In the method, two THz-TDS signals through the samples of the same substance but different thickness are measured. The intrinsic 2-blade $B_2$ of the substance is then determined by the outer product of these two corresponding vectors. Using the conformal split by the fixed bivector $B_2$, each vector corresponding to THz-TDS signals in the vector space $V$ can be linearly splitted into vectors in vector spaces $V_2$ and $V_n$. Since that 2-dimensional subspace $V_2$ is the support of $B_2$, the subspace is also a label for substances. So substances of samples can be identified on the magnitudes of projection vectors in that subspace. This method can contribute to the accurate classification and identification, and facilitate the feature extraction. Finally, experiments are presented and show that the substance identification method is feasible and effective.

1. Introduction
THz time-domain spectroscopy (THz-TDS) is one of the important detecting techniques in the THz research field, by which the time-dependent electric field of THz signals are measured [1-3]. Each THz-TDS signal records a temporal response of the THz reference pulse and can be represented by a pulse waveform, from which the THz spectrum including the amplitude and phase information is obtained by the Fourier transform [4]. Accordingly, the technique of terahertz spectroscopy is very attractive because many materials have an absorption band, which is called a fingerprint. THz-TDS image signal provides not only the spatial information but also the spectral information of the material.
at each pixel. One category of Terahertz spectroscopy studies is aiming to identify or differentiate substances on the basis of their terahertz time-domain spectroscopy (THz-TDS) signals and has been applied in the security checks, drug analysis, genetic engineering and chemistry and other fields.

Current processing and analysis of THz-TDS signals are generally using methods of the scalar data. Usually substances are identified using the amplitude waveform or the phase waveform in the frequency domain, focusing on the waveform properties, such as peaks, slopes and shifts [5,6]. Data processing techniques, such as neural network, wavelet, are mainly on the noise reduction of THz-TDS signals[7-9]. However, these scalar data processing methods have limitations. In practical applications, the analysis methods using waveform characteristics require high signal to noise ratio, and are not suitable to deal with signals of similar waveforms. Recently, the component spatial pattern analysis (CSPA) is proposed by Fukunaga, etc., [10,11]. Substances are identified using absorption spectra measured with a tunable THz wave source. Stability of the THz imaging system is required for high performance. And the substance identification is based only on amplitudes of THz waves regardless of their phase information.

One THz-TDS signal usually contains 1024 samples or more. Such high dimension would cause the data redundant and the high-dimensional disaster. It is necessary to process properly and effectively using some further signal processing techniques especially for high dimensional signals. As a mathematical description language, geometric algebra (GA) provides not only a powerful algebraic framework for the multi-dimensional vector analysis and computing, but also a unified measurement and geometrical description for different geometric models[12,13]. In recent years, geometric algebra has been successfully applied in the computer vision, color image processing, spectral image processing and high-dimensional data processing.

THz-TDS signals can be represented as hyper complex numbers [14-16]. Based on the physical mechanism of the THz-TDS signal analysis, geometric distribution properties and algebraic relationships of THz-TDS signals are analyzed under the framework of GA. From that, signals are demonstrated to belong to the intrinsic blade for each substance. Accordingly, a substance identification method via the conformal split is proposed. In this method, intrinsic blades for each substance are obtained using the normalized outer product of corresponding vectors to the substance with two different thicknesses. Substance of a sample is identified using the maximum magnitude of the corresponding vectors projected onto these blades. Finally, experiments are presented and show that the analysis in this paper is feasible and effective.

2. Physical Model of the THz-TDS Transmission System

The complex refractive index, \( n(\omega) = n(\omega) - j\kappa(\omega) \), comprises the optical constants: the index of refraction \( n(\omega) \) and the extinction coefficient \( \kappa(\omega) \). These constants of the sample can be calculated from the phase and amplitude of the transmitted THz waveforms.

A T-ray signal that passes through a parallel surfaced dielectric sample at the normal angle of incidence can be expressed as a function of the frequency, assuming no reflection, by \( \tilde{E}_{\text{sam}}(\omega) \). The signal measured with the same settings but without the sample is expressed by \( \tilde{E}_{\text{ref}}(\omega) \). Then the sample response normalized by the reference yields the complex transfer function of a material in the frequency domain [4]

\[
\tilde{h}(\omega) = \frac{\mathcal{E}_{\text{sam}}(\omega)}{\mathcal{E}_{\text{ref}}(\omega)} = \frac{4\tilde{n}(\omega)n_0}{\tilde{n}(\omega) + n_0} \exp\left\{-\kappa(\omega)\frac{\omega l}{c}\right\} \exp\left\{-j[n(\omega) - n_0]\frac{\omega l}{c}\right\}
\]

(1).

Taking the argument and logarithm of the transfer function gives, respectively,

\[
\angle \tilde{h}(\omega) = \angle \tilde{f}(\omega) - [n(\omega) - n_0] \frac{\omega l}{c}
\]

(2a)

\[
\ln |\tilde{h}(\omega)| = \ln |\tilde{f}(\omega)| - \kappa(\omega) \frac{\omega l}{c}
\]

(2b)
where \( n_0 \) is the refractive index of free air, \( l \) is the sample thickness and 
\[
\hat{f}(\omega) = 4\hat{n}(\omega)n_0 / [\hat{n}(\omega) + n_0]^{[4]} 
\] is a consequence of the Fresnel loss at the sample-air interfaces.

3. THz-TDS Signals Analysis and Substance Identification via the Conformal Split

Throughout this article, real vectors are denoted by boldface lowercase letters, real scalars are denoted by italic lowercase letters, complex vectors or scalars are denoted by the corresponding ones with a superscript \(^*\), while labels of substances are denoted as uppercase letters in superscripts.

Let any data vector \( \mathbf{x} = [x_i]_{i=0}^{m} \) be the discrete representation of a continuous signal \( x(\omega) \) sampled at \( m \) frequencies \( \{\omega_i, i = 1, 2, \ldots, m\} \). Then discrete representations of the complex refractive index \( \hat{n}(\omega) \) , the transfer function \( \hat{h}(\omega) \) and the Fresnel loss \( \hat{f}(\omega) \) are respectively \( \hat{n} = [\hat{n_i}]_{i=0}^{m} \), \( \hat{h} = [\hat{h_i}]_{i=0}^{m} \) and \( \hat{f} = [\hat{f_i}]_{i=0}^{m} \).

Note that, these discrete representations are complex vectors in an \( m \)-dimensional complex space. It is difficult to analysis THz-TDS signals or perform substance identification directly in such a high dimensional space. In this section, THz-TDS signals are first analyzed using GA. Two functions are defined mapping those complex vectors into real vectors. The linear relationship among those vectors is deduced on the basis of the physical model. From that, three theorems are proved and discussed. Following the fundamental decomposition theorem – the conformal split, the decomposition of vectors corresponding to THz-TDS signals are obtained. Consequently, a novel method of the substance identification based on THz-TDS signals is proposed using the conformal split.

3.1. Signal Analysis

To analysis THz-TDS signals using GA, two functions are defined mapping complex vectors into real vectors. Let \( C_m \) be an \( m \)-dimensional complex space, and \( V_n \) be a \( n \)-dimensional \((n=2m)\) vector space over real numbers with the orthonormal basis of \( n \) space hyper-imaginary units \( e_i, i = 1, 2, \ldots, n \), denoting the corresponding geometric algebra by \( G(V_n) \). For any \( \mathbf{c} = [\mathbf{c_i}]_{i=0}^{m} \in C_m \), functions are defined by

\[
g_1 : C_m \rightarrow V_n, \quad \hat{c} \mapsto \mathbf{c} = g_1(\hat{c}) = \sum_{i=0}^{m} \text{Re}(\hat{c}_i) e_i + \sum_{i=0}^{m} \text{Im}(\hat{c}_i) e_{m+i} \tag{3a}
\]

\[
g_2 : C_m \rightarrow V_n, \quad \hat{c} \mapsto \mathbf{c} = g_2(\hat{c}) = \sum_{i=0}^{m} \angle \hat{c}_i e_i + \sum_{i=0}^{m} \ln |\hat{c}_i| e_{m+i} \tag{3b}
\]

where \( \text{Re}(\hat{c}_i) \) and \( \text{Im}(\hat{c}_i) \) are real part and imaginary part of respectively.

Using the function (3b), the vector relates to the transfer function \( \hat{h} \) is

\[
\mathbf{h} = g_2(\hat{h}) = \sum_{i=0}^{m} \angle \hat{h}_i e_i + \sum_{i=0}^{m} \ln |\hat{h}_i| e_{m+i} \tag{4}
\]

Let \( \hat{\alpha}_i = -(n_i - n_0 + j \kappa_i) \omega l / c \) and \( \hat{\alpha} = [\hat{\alpha}_i]_{i=0}^{m} \), from the equation (2a) and (2b), we have

\[
\mathbf{h} = g_2(\hat{h}) = \sum_{i=0}^{m} \left[ \angle \hat{\alpha}_i - [n_i - n_0] \omega l / c \right] e_i + \sum_{i=0}^{m} \left[ \ln |\hat{\alpha}_i| - \kappa_i \omega l / c \right] e_{m+i}
\]

\[
= \sum_{i=0}^{m} \angle \hat{\alpha}_i e_i + \sum_{i=0}^{m} \ln |\hat{\alpha}_i| e_{m+i} + \sum_{i=0}^{m} \left[ -[n_i - n_0] \omega l / c \right] e_i + \sum_{i=0}^{m} \left[ -\kappa_i \omega l / c \right] e_{m+i}
\]

\[
= \sum_{i=0}^{m} \angle \hat{\alpha}_i e_i + \sum_{i=0}^{m} \ln |\hat{\alpha}_i| e_{m+i} + \sum_{i=0}^{m} \text{Re}(\hat{\alpha}_i) e_i + \sum_{i=0}^{m} \text{Im}(\hat{\alpha}_i) e_{m+i}
\]

that is \( g_2(\hat{h}) = g_2(\hat{f}) + l g_1(\hat{a}) \). Denoting \( \mathbf{a} = g_1(\hat{a}) \) and \( \mathbf{f} = g_2(\hat{f}) \), that equation can be simplified as:
The equation (5) reveals a linear relationship among real vectors correspond to the measurement of THz-TDS signals of the sample and the optical constants of the substance in the sample. The thickness of the sample is exactly the proportional scalar. To explicit deduce the following theorems, some notations and assumptions are clarified first.

For two substances, such as substance A and substance B, suppose $\hat{n}_A = \hat{n}_A^\lambda \hat{t}_A^\lambda \hat{\alpha}^\lambda$ and $\hat{n}_B = \hat{n}_B^\lambda \hat{t}_B^\lambda \hat{\alpha}^\lambda$ are their complex refractive indexes, $\hat{t}_A^\lambda$ and $\hat{t}_B^\lambda$ are their thickness values respectively, $\hat{\alpha}^\lambda$ and $\hat{\beta}^\lambda$ are the corresponding measurements obtained from the THz-TDS transmission system, we made up the following assumptions.

Assumption1. Substances could be identified on basis of their THz-TDS signals. That is the vector, $(\hat{\alpha})_l = g_2(\hat{h})$, which relates to the measurement from the THz-TDS system via the equation (4), corresponds to the unique $l$ and $\hat{n}$ in the T-ray range.

We do not consider other sources’ effecting on the measurements from the THz-TDS system in this paper.

Assumption2. It is assumed that there is no such substances A and B coexist that: their distinct complex refractive indexes, $\hat{n}_A^\lambda \neq \hat{n}_B^\lambda$, could satisfy any of the following relations for $\forall i = 1, 2, ..., m$ :

(a) $\exists \lambda \in R\ , \ \hat{n}_A^\lambda = \lambda \hat{n}_B^\lambda$;

(b) $\exists \lambda \in R\ , \ (\hat{n}_A^\lambda - n_o) = \lambda (\hat{n}_B^\lambda - n_o)$;

(c) $\exists \lambda \in R\ , \ \hat{n}_A^\lambda = 4\lambda \hat{n}_B^\lambda / \left[\hat{n}_B^\lambda + n_o\right]^2$;

(d) $\exists \lambda \in R\ , \ \frac{\hat{n}_A^\lambda}{\left(n_o + \hat{n}_A^\lambda\right)^2} = \lambda \frac{\hat{n}_B^\lambda}{\left(n_o + \hat{n}_B^\lambda\right)^2}$;

(e) $\hat{n}_A^\lambda \hat{n}_B^\lambda = n_o^2$;

That is they are neither linearly dependent nor complex conjugate of each other. Then for substances with distinct complex refractive indexes, their corresponding vectors $\hat{\alpha}$ and $\hat{\beta}$ are linearly independent of each other, that is

$\hat{n}_A^\lambda \neq \hat{n}_B^\lambda \ (\forall i = 1, 2, ..., m) \Rightarrow \hat{\alpha}^\lambda \wedge \hat{\beta}^\lambda \neq 0$.

Based on the assumptions and the equation (5), following properties of the vectors can be deduced.

Theorem1. Under the assumptions, every sequence of the complex refractive index, $\hat{n}$, relates to a unique $\hat{\alpha}$ and a unique $\hat{\beta}$ in $C_m$, consequently, every distinct $\hat{n}$ determines a unique vector $\hat{\alpha}$ and a unique vector $\hat{\beta}$ in $V_n$, and vice versa.

Proof For the substances A and B, which are with exactly the same complex refractive indexes

$\hat{n}_A^\lambda = \hat{n}_B^\lambda$,

i.e.

$\forall i, \hat{n}_i^\lambda = n_i^\lambda$ and $\kappa^\lambda = \kappa^\lambda$.

From the definition of $\hat{\alpha}_i = -(n_i - n_o + j \kappa^\lambda) \omega / c$, it is obviously $\forall i, \hat{\alpha}_i = \hat{\alpha}_i^\lambda$. That is

$\hat{n}_A^\lambda = \hat{n}_B^\lambda \iff \hat{\alpha}^\lambda = \hat{\alpha}^\lambda \iff \hat{\alpha}^\lambda = \hat{\alpha}^\lambda$.

$\forall i, \frac{\hat{n}_A^\lambda}{\left(n_o + \hat{n}_A^\lambda\right)} = \frac{\hat{n}_B^\lambda}{\left(n_o + \hat{n}_B^\lambda\right)} \iff \hat{n}_i^\lambda = \hat{n}_i^\lambda$ or $\hat{n}_A^\lambda \hat{n}_B^\lambda = n_o^2$. 


Based on our assumption 2(e), such substances A and B would not be considered coincidently that their complex refractive indexes could satisfy \( \hat{n}_1^A \hat{n}_1^B = n_0^2 \) \( \forall i = 1, 2, \ldots, m \).

\[ \therefore \quad \hat{n}^A = \hat{n}^B \iff \hat{f}^A = \hat{f}^B \iff f^A = f^B. \]

So, for distinct complex refractive indexes in the space \( C_m \), each determines a unique vector \( \alpha \) and a unique vector \( f \) in \( V_n \). Conversely, each distinct vector \( \alpha \) in \( V_n \) is corresponding to a unique \( \hat{n} \) in \( C_m \) and so is \( f \).

**Theorem 2.** Each complex refractive indexes \( \hat{n} \) relates to a unique 2-blade \( B_2 \) in \( V_n \), where \( B_2 = \frac{\alpha \wedge f}{|\alpha \wedge f|} \).

And it determines a unique 2-dimensional subspace \( V_2 \) in \( V_n \) which is the solution set of the equation \( x \wedge B_2 = 0 \).

**Proof** From our assumption 2, it can be deduced that

\[ \hat{n}^A = \hat{n}^B \iff \alpha^A \wedge a^B = 0, f^A \wedge f^B = 0. \]

And so, \( \hat{n}^A = \hat{n}^B \iff \exists \lambda_1, \lambda_2 \in R, \lambda_1 \lambda_2 \neq 0, \alpha^A = \lambda_1 a^B, \) and \( f^A = \lambda_2 f^B \)

\[ \therefore \quad \frac{\alpha^A \wedge f^A}{|\alpha^A \wedge f^A|} = \frac{\lambda_1 a^B \wedge \lambda_2 f^B}{|\lambda_1 a^B \wedge \lambda_2 f^B|} = \frac{\alpha^B \wedge f^B}{|\alpha^B \wedge f^B|}. \]

\[ \therefore \quad \hat{n}^A = \hat{n}^B \iff B_2^A = B_2^B. \]

**Theorem 3.** All vectors, \( h_i = f + l a \) \( \forall l \in R \) are collinear and belong to \( B_2 \).

**Proof** Using properties of the outer product, it is obviously true that

\[ h_i \wedge A_2 = (f + l a) \wedge a \wedge f = f \wedge a \wedge f + l a \wedge a \wedge f = 0 \] \( \forall l \in R \).

These theorems are important and usable in the substance identification. Theorem 2 demonstrates a unique 2-blade \( B_i \) relates to each substance. And theorem 3 reveals that vectors corresponding to THz-TDS signals from the same substances but with different thickness belong to the same blade. The geometric meaning of these theorems is interpreted in the figure 1. From the theorem 2, as the complex refractive indexes are different, the blades \( B_2^A \) and \( B_2^B \) are linearly independent \( B_2^A \wedge B_2^B \neq 0 \).

That in the geometric interpretation is they are not lying on the same plane. Since \( h_i^A \) and \( h_i^B \) are corresponding to the same substance A, they are collinear and on the same plane of \( B_2^A \), regardless of their different thickness values.
3.2. Substance Identification

As in GA, to every $r$-dimensional subspace $V_r$ in $V_n$ there corresponds an $r$-blade $B_r$ such that $V_r$ is the solution set of the equation $x \wedge B_r = 0$. $B_r$ is called the blade of $V_r$ while $V_r$ is called the support of $B_r$. [12]. The 2-blade $B_2$ defined in the theorem 2 determines a unique subspace $V_2$ for each substance, which is independent of thickness values of samples. So, the substance identification can be performed by mapping $n$-dimensional vectors $l_i$ into each 2-dimensional subspace $V_2$ (the support of $B_2$).

Accordingly, we proposed a substance identification method via the conformal split (CS).

In the study of D. Hestenes, geometric algebra can be factored multiplicatively according to the fundamental theorem: $G_n = G_2 \otimes G_{n-2}$, (The Kronecker product $\otimes$) [12]. A conformal split by a bivector determines a geometrically significant relation between the geometric algebras $G_n$ and $G_{n-2}$. Let $B_2$ be a fixed unit 2-blade in $G_{n-2}$ and let $x$ be a generic vector in $V_{n-2} = G_{n-2}$. A linear split of $V_n$ into vector spaces $V_2$ and $V_{n-2}$ is determined by the equation

$$x B_2 = x \cdot B_2 + x \wedge B_2 = x_0 + \rho X \quad (6)$$

where $V_2 = \{x_0 = x \cdot B_2\}$ and $V_{n-2} = \{\rho X = x \wedge B_2\}$ [13].

Given a fixed 2-blade $B_2$ of one substance, the vector $h_i$ in $V_n$ corresponding to THz-TDS signals from one sample, it can be linearly mapped into one 2-D subspace $V_2$ and the other subspace $V_{n-2}$ using the conformal split (6):

$$h_i B_2 = h_i \cdot B_2 + h_i \wedge B_2 = h_0 + \rho H \quad (7)$$

where $h_0 = h_i \cdot B_2$. Using the above equation, high dimensional vectors of THz-TDS signals is mapped into 2-D vectors.
In the process of mapping the vector $\mathbf{h}_i$ into the 2-D subspace intrinsic to one substance, it is necessary to obtain the unique 2-blade $\mathbf{B}_i$ of that substance first. In the condition the complex refractive index $\hat{n}$ of that substance is already known, the 2-blade $\mathbf{B}_i$ can be easily obtained through its definition in the theorem 2. In the general cases the complex refractive indexes of substances are unknown, the 2-blade $\mathbf{B}_i$ of each substance can be evaluated using experimental measurements, which is discussed in the following.

Intrinsic blades for “known” substances can be experimental computed. Given two THz-TDS signals measured respectively from samples of the same substance “A” but with thickness $l_1$ and $l_2$, they can be expressed by hypercomplex numbers, $\mathbf{h}_{1l}$ and $\mathbf{h}_{2l}$, in form of the equation (4). From the equation (5), it is deduced that

$$\mathbf{h}_{1l} \wedge \mathbf{h}_{2l} = (\mathbf{f} + l_1 \mathbf{a}) \wedge (\mathbf{f} + l_2 \mathbf{a})$$

$$= \mathbf{f} \wedge \mathbf{f} + l_1 \mathbf{f} \wedge \mathbf{a} + l_2 \mathbf{f} \wedge \mathbf{a} + l_1 \mathbf{a} \wedge l_2 \mathbf{a}$$

$$= (l_1 - l_2) \mathbf{a} \wedge \mathbf{f}$$

As in the definition of $\mathbf{B}_i$ in the theorem 2, $\mathbf{B}_i$ is a unit 2-blade, so

$$\mathbf{h}_{1l} \wedge \mathbf{h}_{2l} = \lambda \mathbf{B}_i, \quad (\lambda \in \mathbb{R})$$

Consequently the computation of $\mathbf{B}_i$ is as:

$$\mathbf{B}_i = \frac{\mathbf{h}_{1l} \wedge \mathbf{h}_{2l}}{|\mathbf{h}_{1l} \wedge \mathbf{h}_{2l}|}. \quad (8)$$

Blades of some substances to be identified can be obtained using the above equation. For the $i$th substance $A_i$, $i=1,2,\ldots$, it is measured first THz-TDS signals from the samples of the same substance but with two different thicknesses. With vectors corresponding to those signals, $\mathbf{h}_{1l}^{A_i}$ and $\mathbf{h}_{2l}^{A_i}$, the blade of that substance can be evaluated using the equation (8).

For the vector $\mathbf{h}_i^M$ corresponding to signal from one sample to be identified, it can be linearly decomposed into one 2-D subspace of and the other subspace of 2 dimension lower with respect to the blade $\mathbf{B}_i^{A_i}$ of each substance using the equation (7):

$$\mathbf{h}_i^M \wedge \mathbf{B}_i^{A_i} = \mathbf{h}_i^M \cdot \mathbf{B}_i^{A_i} + \mathbf{h}_i^M \wedge \mathbf{B}_i^{A_i} = \mathbf{h}_i^{M_0} + \rho \mathbf{H}_i^{M_0}.$$

Magnitudes of the projection of $\mathbf{h}_i^M$ onto the blades can be obtained, which is used as the criterion in the substance identification. The sample is then identified as the $k$-th substance based on the maximum magnitude:

$$k = \arg \max_i \left| d_i^M \right|. \quad (9)$$

where $d_i^M = \left| \mathbf{h}_i^{M_0} \right| = \left| \mathbf{h}_i^M \cdot \mathbf{B}_i^{A_i} \right|$.

Consequently, the THz-TDS signal of an unknown substance can be identified by comparing its mapping vectors in these blades intrinsic of “known” substances. The figure 2 illustrates the process of the substance identification (CS).
4. Results

4.1. Experiment I

Ten THz-TDS signals are obtained and also their related reference signals from five substance samples using our THz-TDS system in the transmission model. The substances are marked as CARD, OSA, PTFE, RB, and WB. Figure 2 (a) shows the amplitudes of the transfer functions for these signals. These waveforms are quite noisy and fluctuated. It is difficult to classify three substances, which are WB, RB and OSA, from each other as their waveforms are quite similar. Blades intrinsic to these five substances are computed as the outer product of the vector pairs corresponding to the signals from the same substance.

![Figure 2. Illustration of the substance identification (CS).](image)

Figure 3. THz-TDS signals

Fourteen raw THz-TDS signals (shown in the figure 3(b)) are also measured from these five distinct samples. Noisy signals simulated on these fourteen signals are identified using our method. Simulations are performed by adding Gaussian white noises to signals with zero mean and different variances (listed in the table 1). Each simulation is run 500 times and 500 noisy signals are obtained for each raw signal. The substances identification method is performed on these noisy signals. Table 1 shows the accurate rate of identification over 500 runs for each simulation. The average of the signal
The standard deviation (SNR) in each simulation are also calculated and listed out in the table 1. Simulation experiments show the high performance of the method on the identification of substances on the basis of their THz signals. This experiment demonstrates the feasibility of our method.

**Table 1. Accurate rate of identification**

| The standard deviation (σ) | SNR      | Accurate rate of identification (%) |
|----------------------------|----------|-------------------------------------|
| 0.001                      | 51.0786  | 100%                                |
| 0.01                       | 31.1131  | 100%                                |
| 0.05                       | 17.2672  | 98.10%                              |
| 0.1                        | 11.6075  | 86.50%                              |

4.2. Experiment II

A THz-TDS image was simulated. It was composed of 192*265 pixels and imaged at 32 frequencies (0.2THz~0.6THz). Experimental THz-TDS signals of substances ‘CARD’, ‘OSA’ and ‘PTFE’ were used. Each substance was simulated as being placed in a small block (a region of interest, ROI). Gaussian noises with zero mean and different variances (listed in the table 2) were added to the THz-TDS image. Such that, four simulated noisy THz-TDS images were obtained.

Substances identification on these THz-TDS images was performed using our method (CS). The visualized results are shown in the figure 4. For compare, substances identification was also performed using the component spatial pattern analysis (CSPA) [10,11]. Accurate rates (listed out in the table 2) of both method in each experiment are calculated. In the noisy conditions σ =0.5 and 0.1, the accurate rates of our method (CS) are higher than those of the CSPA method. That shows the advantage of our method (CS).

**Figure 4.** results of the substance identification (CS)

**Table 2. Accurate rate of identification**

| The standard deviation (σ) | SNR      | Accurate rate of identification (%) |
|----------------------------|----------|-------------------------------------|
| 0.001                      | 50.6070  | 100%                                |
| 0.01                       | 30.6097  | 100%                                |
| 0.05                       | 16.7254  | 98.83%                              |
| 0.1                        | 10.9694  | 90.78%                              |

5. Conclusion
In conclusion, a novel analysis of the THz-TDS signals using GA is presented. THz-TDS signals, described in the high dimensional vector space, are mapped into the hyper-complex numbers in GA. Geometric properties of vectors corresponding to signals from the same substance but different thickness are deduced using the language of GA. From that, signals are demonstrated to belong to the intrinsic blade for each substance.

A substance identification method via the conformal split is proposed. In this method, intrinsic blades for each substance are obtained using the normalized outer product of vectors corresponding to the THz-TDS signals from the same substance but two different thicknesses. And for each sample to be identified, the corresponding vector is projected onto each blade. The maximum magnitude of the projected vector is used to identify the sample's substance.

Experiments demonstrate the feasibility and high accuracy of our method on the applications of identify substances on the basis of their THz transmission spectra.

Acknowledgment
This study is supported by the National Natural Science Funds (No. 61071206, P.R.China).

References
[1] D. M. Mittleman, S. Hunsche, etc., T-ray tomography [J]. Optics letters, 1997, Vol.22(12):904-906
[2] W. X. Xie, J. H. Pei, “Review of Terahertz signal processing and analysis”, ACTA Electronica sinica, 35(10), 1973-1978, (2007)
[3] Pei Ji-hong, Hu Yong, Xie Wei-xin, PCA-based visualization of terahertz time-domain spectroscopy image[A]. ,MIPPR2007: Multispectral Image Processing[C]. Wuhan, China, Proc. of SPIE, 2007 Vol. 6787, 67871M-1: 67871M-7Nov 15-17
[4] W. Withawat, B.M. Fischer, D. Abbott, “Material thickness optimization for transmission-mode terahertz time-domain spectroscopy”, Optics express, 16(10), 7382-7396, (2008)
[5] H. B. Liu, X. C. Zhang, Dehydration kinetics of D-glucose monohydrate studied using THz time-domain spectroscopy [J]. Chemical Physics Letters, 2006, Vol.429:229–233
[6] Pupeza, R. Wilk, M. Koch, Highly accurate optical material parameter determination with THz time-domain spectroscopy [J]. Optics express,2007, Vol.15(7):4335-4350
[7] J.H. Pei*, P.L. Ye and W.X. Xie, Optimal wavelet analysis for THz-TDS pulse signals[A]. Photonics and Optoelectronics Meetings, POEM 2008[C]. Wuhan, China
[8] Sillas Hadjiloucas, Roberto K. H. Galvão, John W. Bowen, Analysis of spectroscopic measurements of leaf water content at terahertz frequencies using linear transforms [J]. Opt. Soc. Am. A, Vol. 19(12), 2002: 2495-2509
[9] Roberto K. H. Galvão, Sillas Hadjiloucas and John W. Bowen, Clarimar J. Coelho, Optimal discrimination and classification of THz spectra in the wavelet domain[J]. OPTICS EXPRESS, OSA, 2003, Vol.11(12):1462-1472
[10] K. Fukunaga, Y. Ogawa, etc, Application of terahertz spectroscopy for character recognition in a medieval manuscript[J]., IEICE Electronics Express, 2008, Vol.5(7):223-228
[11] K. Kawase, Y. Ogawa, Y. Watanabe, Non-destructive terahertz imaging of illicit drugs using spectral fingerprints, OPTICS EXPRESS, OSA, 2003, Vol.11(20):2549-2554
[12] D. Hestenes, Universal geometric algebra, [J], A quarterly J. of Pure and Applied Mathematics, 1988, Vol. 62(3-4)
[13] D. Hestenes, The design of linear algebra and geometry, [J], Acta Applicandae Mathematicae, 1991, Vol. 23, pp. 65-93
[14] W. Xie, J. Li, J.H. Pei, A Clifford Algebra Analysis Method of THz-TDS Images [A]. 9th International Conference on Signal Processing (ICSP 2008 Proceedings) [C]. Beijing, China, IEEE Press, 2008.128-131
[15] W. Xie, J. Li, J.H. Pei, An Analysis of THz-TDS Signals using Geometric Algebra [A]. Photonics and Optoelectronics Meetings, POEM 2008, Wuhan, China
[16] W. Xie, J.H. Pei, J. Li, A Dimensionality Reduction Method for THz-TDS Signals via the Recursive Projective Splits based on PCA, Terahertz Science and Technology, the On-line International Journal, JTST, Vol.2(2), 2009