

# Wavelet-Based Dimensionality Reduction for Hyperspectral THz Imaging

Henrike Stephani <sup>1\*</sup>, Michael Herrmann <sup>2</sup>, Frank Bauer <sup>3</sup>, and Bettina Heise <sup>3</sup>

<sup>1</sup>Fraunhofer ITWM and Technical University, Kaiserslautern, Germany

Fraunhofer Platz 1, 67663 Kaiserslautern, Germany

<sup>2</sup>Fraunhofer IPM, Kaiserslautern, Germany

<sup>3</sup>Johannes Kepler University, Linz, Austria

\*1 Email: [henrike.stephani@itwm.fraunhofer.de](mailto:henrike.stephani@itwm.fraunhofer.de)

(Received 15 July 2010; accepted 20 September 2010)

**Abstract:** With terahertz time-domain spectroscopy, hyperspectral images can be acquired where each pixel contains a full spectrum of the range of several terahertz (THz). An enormous amount of data is generated. Therefore, advanced methods for automated data analysis and image processing are required. We present a wavelet-based approach for channel reduction and feature selection for a subsequent clustering leading to an image segmentation. The main focus of our method is set on the appropriate dimensionality reduction adapted to the THz spectral characteristics of the samples under investigation. A feature reduction to less than 5% is achieved, thereby enabling a channel-wise image processing on the reduced data set. Furthermore, unsupervised classification is chosen for an automatized segmentation including all channel information represented in the wavelet domain. Relevant characteristics of the THz spectra are preserved by our feature selection, in particular the distribution of the peak position and peak depth. The proposed method for channel reduction is verified by extensive simulations at first. Finally, it is demonstrated on various real-world measurements of chemical compounds. The improved performance of the analysis on the reduced feature set could be shown in comparison with the evaluation on the full data set.

**Keywords:** THz-TDS Imaging, Hyperspectral Image Processing, Feature Selection

**doi:** [10.11906/TST.117-129.2010.09.12](https://doi.org/10.11906/TST.117-129.2010.09.12)

## 1. Introduction

In terahertz time-domain spectroscopy (THz-TDS) femto-second laser pulses are used to probe material properties in the terahertz (THz) range. A spectrum of broad bandwidth – about 100 GHz up to 3 THz or more – is acquired. Within this frequency band, chemical compounds have characteristic absorption lines. THz-TDS is particularly useful for non-destructive testing and security control [1]. Using measurements of broad bandwidth not only for one-dimensional data analysis but rather for imaging has become increasingly popular in the last years. Regarding multiple channels for imaging – so-called hyperspectral imaging – has been a topic in ultrasonic and infrared imaging already [2] and is especially important in THz imaging as the number of channels is even higher.

Every pixel of such an image represents a spectrum consisting of hundreds of values. Due to this high dimensionality, image processing and interpretation can be difficult. In addition, some of this information is redundant and noisy. Therefore, we propose to apply a wavelet-based feature selection. Wavelets are particularly suitable for THz spectra because of their shape being similar to the shape of the characteristic peaks and the efficiency with which they can be calculated. By this method we can reduce the number of features from over 500 down to under 20.

To further process hyperspectral imaging data, researchers of that area, such as [3,4,5] use classification. Typical feature selection algorithms in classification are aimed at optimizing

the classification result. They are usually based on entropy values of a specific dataset [6,7]. In contrast to that, the suggested wavelet-based feature selection method is applicable on arbitrary THz spectra of chemicals. This allows incremental data acquisition.

To prove the validity of these methods we will present an extensive simulation scheme of spectra that have the same characteristic peaks as well as the same noise as THz spectra. One thereby can compare different feature sets and theoretically prove the adequacy of the method. We then apply the proposed feature selection on real-world measurements of 5 different chemical compounds. For automatic data organization in this paper two different hierarchical clustering methods are used.

To illustrate the necessity of feature selection we further employ a typical 2D image processing filter – anisotropic diffusion – on the wavelet feature channels of the real-world measurements. Such channel-wise image processing is not feasible on the full number of channels. It nevertheless is recommendable because it leads to a noise reduction based on the neighborhood information of each pixel and not only on the spectral information.

## 2. Methods

Classification is often applied for the interpretation of hyperspectral images. In this section the general functionality of the classification utilized here – hierarchical unsupervised classification – will be explained, as will be the classical approaches on feature selection. Thereby the proposed feature selection method is embedded into the methods currently in use

### 2.1 Unsupervised classification

To be able to automatically group new measurements without the necessity of labeling data beforehand, we use unsupervised classification also known as clustering where no training data is needed. Instead, these methods are based on distances between the given data and classify it by maximizing the interclass distance and minimizing the intraclass distance. There is a broad variety of clustering algorithms. We will utilize an agglomerative hierarchical clustering. We engage two different distance measures and cluster updating functions. Hierarchical clustering in general has the advantage of needing no input parameters and being very flexible in the possibilities of data analysis. Because the focus of this paper lies on the validity of the feature selection proposed, we refer the reader to [7] or [8] for further information about clustering.

### 2.2 Feature Selection

Working with extremely high dimensional data, as given in case of a spectral data set, may result in high computational costs. Furthermore we have to face the so-called “curse of dimensionality”, as the number of samples in relation to the number of features is very low. Therefore, researchers perform dimensionality reduction or feature selection where either a subset or a transform of the original features is chosen for the clustering process.

Although this topic is not new, typical feature selection algorithms in use are based on optimizing the entropy of either the features themselves or the clustering outcome [9,10]. It is not possible to compare incrementally acquired data sets. Hence, we propose a method that

reduces the dimensionality independently of the respective measurement.

### 2.3 Wavelet-Based Feature Selection

The wavelet transformed has been applied in THz applications in different ways, e.g. for noise reduction in the signal domain [11] or for image processing [12]. In contrast to that we propose to apply it on the frequency domain spectra of each pixel for dimensionality reduction. Using wavelet coefficients for this purpose has been done before in other time series applications in combination with clustering [13,14]. In these cases, the most discriminating coefficients were used to represent the whole data set. A drawback of this method is that it is still based on a closed data set. To avoid this we propose to use coefficients of a fixed level as a transform of the original spectra.

The wavelet transform of a time series consists in the loss-free decomposition of a signal. This decomposition depends on time as well as on frequency. The discrete wavelet transform of a signal can be calculated very efficiently due to the typical dyadic downsampling scheme used for that purpose. In this scheme the whole spectrum is iteratively filtered by the actual wavelet functions and a scaling function [15,16]. Therefore, coarse features already appear in low scales.

The features we are looking for, are the peaks of the spectra that characterize different chemical compounds. In THz radiation of solids these peaks are about 100 GHz broad. With such broad characteristics, using only a small percentage of the original channels should suffice to preserve the relevant information.

## 3. Simulation

The information content of THz spectra is mainly encoded in the position and depth of the peaks. This information has to be preserved or even enhanced by a feature selection. Hence, we validate our proposed method with respect to the peak detection that can be achieved on a certain feature set. For this validation we simulate spectra that have characteristics of the same nature as the typical peaks. The simulation does not aim at the prediction of the spectra of specific compounds but rather at simulating a large number of THz spectra with a known ground truth. We will start by simulating the basic shape of the spectra, continue with noise considerations and then simulate the characteristic peaks.

### 3.1 Basic Shape

THz-TDS is based on pulsed THz systems where time-domain signals (pulses) are recorded. The pulse is proportional to the second derivative of a transient of electrical polarization [17]. To obtain the basic shape of a measurement, we therefore simulate the polarization transient. We thus gain the reference and the sample pulse and use both for calculating the transmittance spectrum. Fig. 1 shows example reference and sample basic shape. The figure illustrates how signal and spectrum can be calculated from that. The power and bandwidth of the spectrum are determined by the slope of the transient.

### 3.2 Noise

The next step is to include typical noise that appears in THz-TDS spectra in the simulation.

In this paper we only consider the signal noise and do not account for the influence of humidity and temperature because it strongly depends on the specific measurement situation.

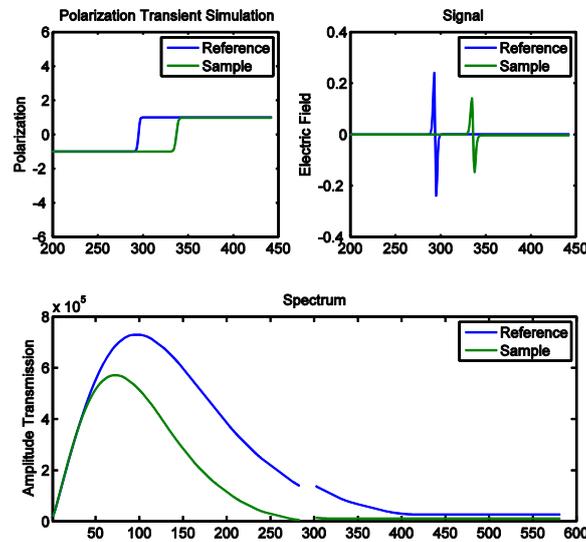


Fig.1 Upper left shows the polarization transients, upper right their second derivatives that are equivalent to the signals. The spectra at the bottom plot are the Fourier transforms of the signals.

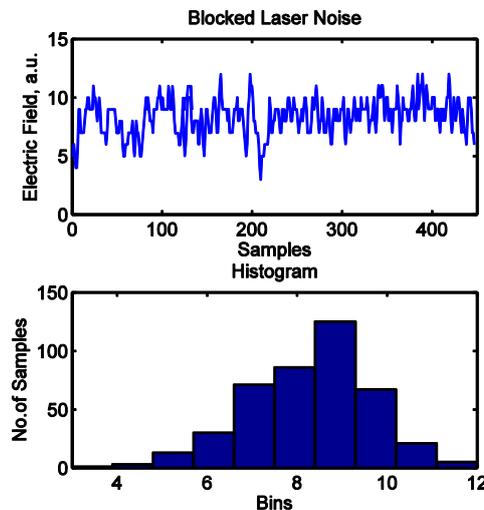


Fig.2 Upper plot shows the distribution of the blocked laser signal, the lower plot shows the histogram of the values.

[18] showed that in THz-TDS the variance of the transmission modulus  $\rho(\omega)$  can be written as  $\sigma_{\rho}^2(\omega) = A(\omega)\rho^2(\omega) + B(\omega)\rho(\omega) + C(\omega)$ , where  $A(\omega), B(\omega)$ , and  $C(\omega)$  are coefficients depending on emitter, detector and shot noise respectively.

The laser is a strong noise source in THz measurements and enters the THz setup through the antennas on the emitter and detector side. On the emitter side it is the dominant noise source during a THz pulse ([18], there called “emitter noise”), however, we do not explicitly treat it in our study because it is present only during a very short period of time. With

sufficiently long waveforms as measured for spectroscopy, detector noise is increasingly important. It is also partly generated from laser noise but has other components particularly from electronic noise in the antenna and preamplifier. The relative intensity of these noise sources depends on the laser model, the antenna type, the current-to-voltage resistor in the first amplifier, and the modulator (“chopper”) frequency.

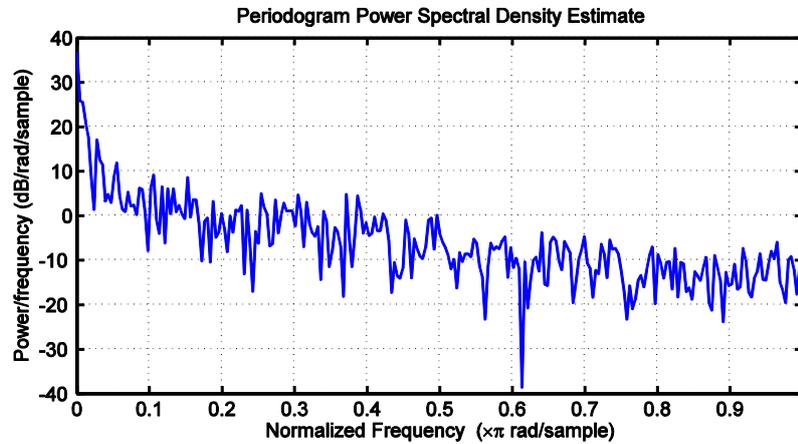


Fig.3 Power spectral density of the blocked laser beam noise is shown. Its  $1/f$  character can be seen.

To determine the characteristics of this noise we have to take measurements with a blocked laser beam. As this is usually not taken separately, we consider the signal before the laser pulse of several reference measurements, i.e. measurements without a sample between emitter and detector. Fig. 2 shows the noise signal and its histogram. The shape suggests that the noise is normally distributed, with its mean and standard deviation as parameters. This assumption was verified with an  $\chi^2$ -Test.

In addition to the distribution we analyze the power spectral density. In Fig. 3 it can be seen that the density is not constant over the spectrum but declining with increasing frequency. This noise form is called  $1/f$ -noise.

This suggests that the laser may be the dominant noise source and therefore, the noise that shall be used in our simulation is  $1/f$  noise with a Gaussian distribution.

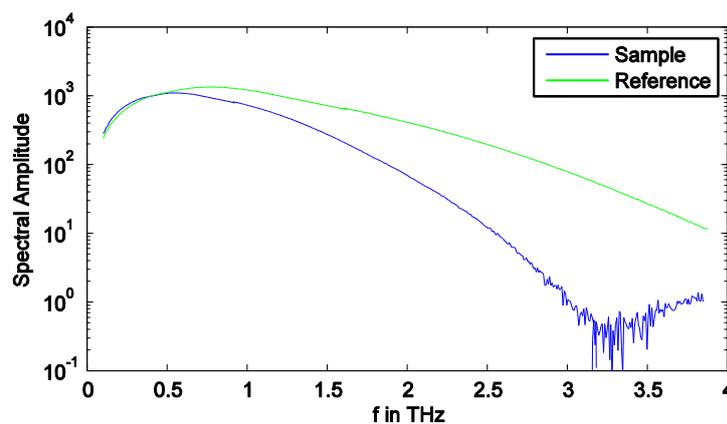


Fig.4 Logarithmic transmission amplitude of reference and sample's basic shape with the  $1/f$  noise added.

We now add the noise to the simulated signals. The dynamic range of THz measurements is determined by the reference measurement normalized with the noise floor and typically moves in a range of  $10^3$  [19]. We simulate the noise with respect to that and the result can be seen in Fig. 4.

### 3.3 Peaks

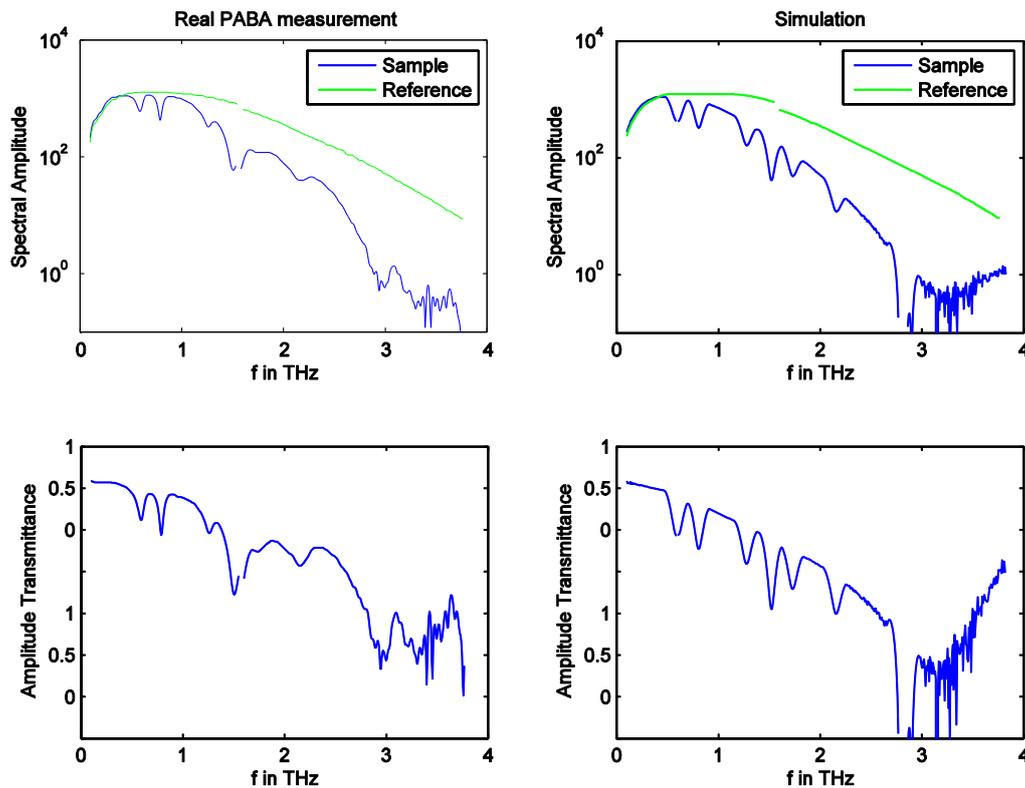


Fig.5 Left hand side shows simulated spectra, Right hand side measured ones. On the top level reference and sample spectra can be seen while on the bottom level the transmittance is plotted.

The goal of any feature selection applied to THz spectra is to preserve the differentiating information. That information mainly consists in the position and depth of the peaks. Hence, we will validate our proposed method with respect to that.

To simulate the peaks, their typical shape has to be described first. THz measurements of solids show peaks of a width of around  $100\text{GHz}$ . The depth varies greatly from an almost complete absorption up to only slight shifts from the background. The position of the peaks covers the whole bandwidth of the spectrum. In this paper we shall consider a bandwidth of  $3\text{THz}$  and a dynamic range of  $10^3$ .

In Fig. 5 on the left hand side a PABA spectrum and its reference spectrum can be seen, on the right hand side a simulated spectrum with the same number and position of peaks is shown. Below that, the respective transmittance spectra are plotted. This should serve as an illustration of the similarity of the simulated spectra with measured ones. The peak simulation is done using splines.

### 4. Evaluation of the Method

We now apply the wavelet decomposition proposed for dimensionality reduction. To illustrate why doing so is sensible, a plot of two different spectra of chemical compounds is shown in Fig. 6 on the top. The middle and the bottom graphs show the wavelet coefficients of level 4 and 5 (i.e. 32 and 16 coefficients respectively). Both are calculated with a Daubechies wavelet with a support of 20 points. Although the peaks, i.e. the characteristics of

the original spectrum, do not appear at the same position of the reduced feature sets, nevertheless in both sets of wavelet coefficients the two different compounds can be easily distinguished from each other. Looking at the 16 coefficients feature set in Fig. 6 makes visual differentiation even easier than using one of the other two feature sets.

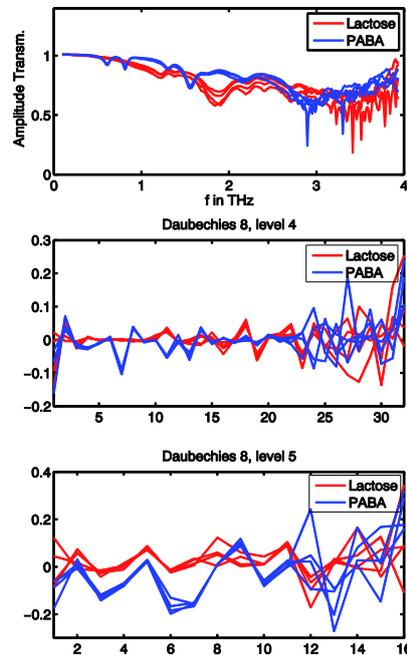


Fig.6 Top shows Simulation of spectra with four peaks, all with a width of about 100GHz. Middle and bottom show wavelet coefficients of fourth (32) and fifth (16) sub-sampling level.

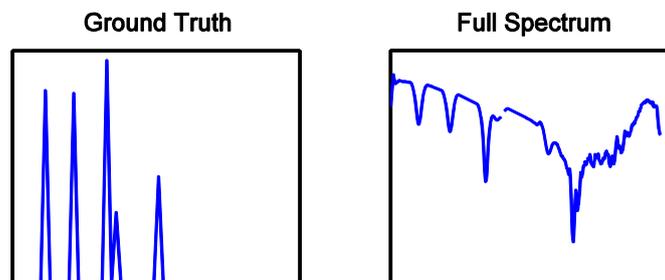


Fig.7 Schematic view of the different “spectra” that are being compared here.

We now perform a systematic simulation of these spectra. As we create the spectra ourselves we know the ground truth of the two important features, namely the position and depth of each peak. Assuming the overall absorption of the spectrum to be constant, the ideal feature selection for a THz spectrum would consist in exactly these two features. Hence, one has to compare every proposed feature selection with this. In Fig. 7 we can see a schematic view of how the ground truth is used here. On the left hand side the sparse vector containing only 5 non zero values – indicating the respective peak positions and depth – is shown while the right hand side shows one filtered spectrum that could belong to these values.

We validate the feature selection now by comparing the different feature sets with the ground truth: The distance matrix of the ideal features is represented by  $D$ . We compare it to the distances between the full spectra represented by  $D_{sp}$  and the distances between the 16-

and 32-dimensional wavelet representation  $D_{wa16}$  and  $D_{wa32}$ .

For the correlation analysis Pearson's correlation coefficient [20] is used by calculating between each two columns  $D_{sp}(:, j)$  and  $D(:, j)$  :

$$C_{D_j D_{sp_j}} = \frac{\sum_{i \in I} (D_{(i,j)} - \overline{D_{(:,j)}})(D_{sp_{(i,j)}} - \overline{D_{sp_{(:,j)}}})}{(n-1)(s_{D_{(:,j)}} s_{D_{sp_{(:,j)}}})} \quad (1)$$

where  $s_{D_{(:,j)}}$  and  $s_{D_{sp_{(:,j)}}}$  are the standard deviations of the  $j$ th column of the matrix  $D$  and  $D_{sp}$  respectively. Furthermore  $I=\{1, \dots, n\}$ , where  $n$  is the number of channels. The accuracy measure is then defined as the mean correlation over all columns  $\overline{C_{D_j D_{sp_j}}}$ . For the wavelet coefficients,  $\overline{C_{D_j D_{wa_j}}}$  is calculated analogously.

For a good visualization of the result we simulate spectra with respect to the two different characteristics: firstly we simulate spectra with a different number of peaks and secondly with different depth of the peaks. In both cases we systematically simulate 20 different sets of features and each set of features appears in 5 simulated spectra. Thus every simulation consists of 100 simulated spectra and is then repeated 10 times. The results shown here are based on a total number of 1000 simulations each.

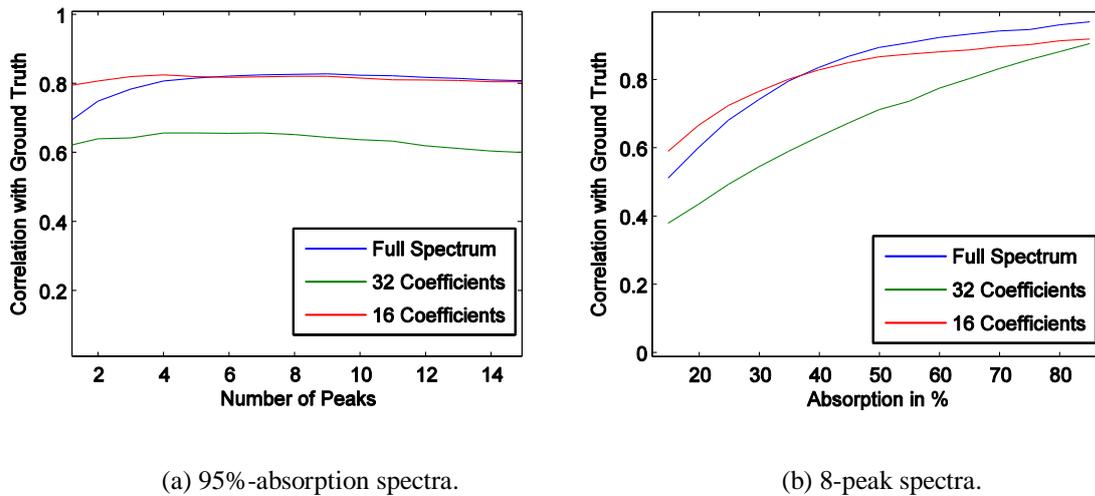
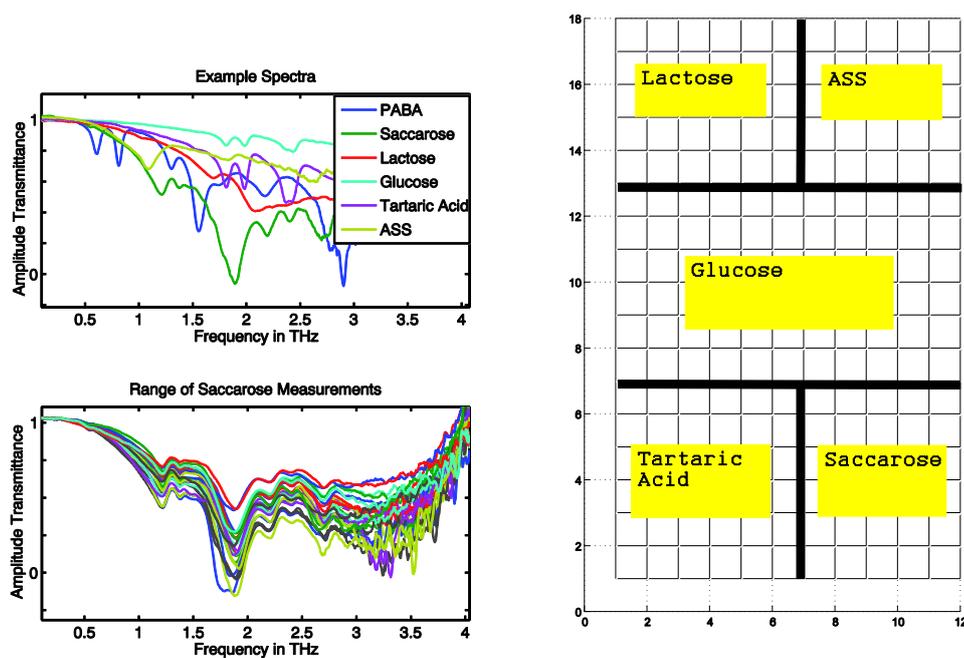


Fig. 8 Correlation analysis of Spectra with different features.

Fig. 8 shows the correlation analysis of the simulation of spectra with different peak numbers. One can note that with less than 6 peaks the correlation declines notably. This can be explained by the fact that the more peaks are within a spectrum the less the idealized feature selection differs from the actual feature selection because of the vector becoming less sparse. Generally it can be seen, that the feature selection based on 16 wavelet coefficients is as highly or more highly correlated with the ideal one as the full spectrum is. Interestingly using 32 coefficients does not bring a further improvement to the result but presents consistently worse results. This might seem counter-intuitive at first. But as illustrated in Fig. 6, using 16 coefficients, while suppressing some characteristics on the one hand, leads to a more distinct separation of the spectra on the other hand. This leads to a higher correlation with the ground truth. The latter only consists of the distances between the peak positions and depths and hence rewards feature sets that express a good separation rather than fine features.

In a next step we vary the depth of the peaks. In the above simulation the peaks were equally simulated to be of high absorption namely 95% of the actual value. We now use spectra with 8 peaks each and shift the depth of these peaks from 95% down to 15% absorption. Fig. 8 shows the result of this simulation. The lower the absorption, i.e. the more similar the peaks are to the actual spectrum, the less correlated the full spectrum and the wavelet coefficients are to the ground truth. One can note though, that for the lower peaks – meaning from 35% on – the reduced 16 feature vector detects the peaks better than the full spectrum does.



(a) Spectra of 5 different chemical compounds.

(b) Scheme for sample arrangement.

Fig.9 Data set used for application.

The results shown in Fig. 8 (a) and 8 (b) are based on a Daubechies10 wavelet feature selection. The same procedure was systematically repeated with different wavelets with different support size. Daubechies10 wavelets showed the highest correlation and are therefore used for the further application to real-world data.

## 5. Application to Real-World Data

The correlation analysis of the simulated spectra has proven the dimensionality reduction with 16 Daubechies 10 wavelet coefficients to be valid for THz spectra. That means that the important information namely the peak position and depth is preserved by this selection. We will now firstly practically undermine these theoretical findings by applying this feature selection to a real-world data set and compare clustering results of the different feature sets. Secondly we will show the advantage of having a thus reduced set of channels by demonstrating its potential in image processing.

The data used for this purpose consists in an image put together from six  $6 \times 6$ -pixel measurements of five different chemical compounds. Examples for each of these compounds can be seen in Fig. 9 (a) on the top level. Fig. 9 (b) shows how the measurements are put

together and what the ground truth looks like. The goal of a classification is to classify these compounds apart. The aim of image processing is to smooth within compound areas and to preserve the differences.

## 5.1 Clustering

Clustering is used to combine information from various feature channels. A successful feature selection should hence yield the same clustering result as beforehand.

We use two different clustering techniques. One of them is classical complete link clustering which has a good noise resistance and can be easily used with an arbitrary distance measure. Therefore, we do not use the classical Euclidean distance here but a cosine distance measure that is invariant to scalar multiplications. This is advisable due to the vertical variance of the spectra. This phenomenon can be seen in the bottom plot of Fig. 9 (a). The other clustering method we use is Ward's link clustering, which is very popular amongst researchers as a general homogeneity of the data partitioning is used as a clustering criteria. This homogeneity measure is based on the squared error, hence, the Euclidean distance measure is necessary.

In Fig. 10 the result of the Ward's link clustering is shown while Fig. 11 shows the cosine distance clustering. The first 7 clusters are visualized, each one represented by one color, i.e. if two pixels have the same color, the algorithm regards them as being in the same cluster. In Fig. 10 using 16 features yields the best result. While using 32 coefficients leads to over-classification within tartaric acid and under-classification within the other compounds, using the full spectra leads to misclassification altogether, e.g. within Glucose and ASS. In Fig. 11 all methods can differentiate between the clusters well. Nevertheless the cosine distance measure in combination with the complete linkage clustering leads to a result where the full spectra perform slightly better than the reduced feature sets.

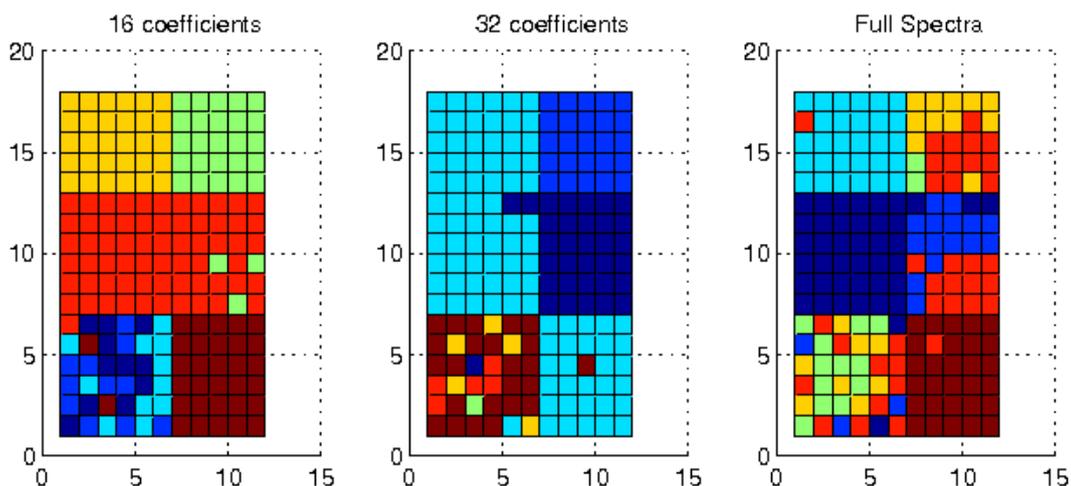


Fig.10 Euclidean distance clustering with Ward's linkage function.

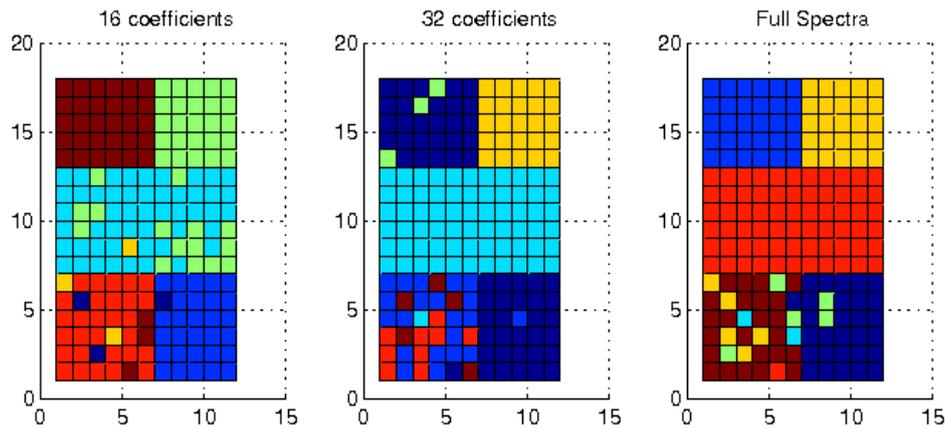


Fig.11 Cosine distance clustering with complete linkage function.

## 5.2 Image Processing

One reason for feature selection is the improvement of the clustering result by avoiding the curse of dimensionality. Another is that with such a reduced set of channels, slice-wise image processing is possible. The clustering applied beforehand did only include spectral information into an image segmentation. But as we are coping with an imaging method, there is also spatial information that should be taken into consideration. With more than 500 channels this is difficult whereas with only 16 an interpretable result can still be produced. In Fig. 12 the first 12 wavelet coefficients without any image processing are shown. In Fig 13 the same channels are visualized after a Perona-Malik diffusion filter was applied to them. This filter is especially well applicable when smoothing on the one hand and edge preserving on the other hand is necessary. The result shows that this aim is served well in most channels and thereby the spatial information is well included. Further image processing is sensible and should be applied in future research steps. However by using such a general form of feature selection the possibility to do that is given.

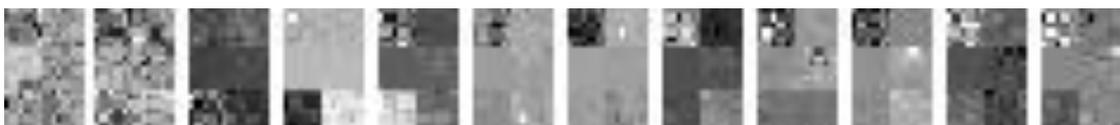


Fig.12 of the 16 wavelet channels of the example image

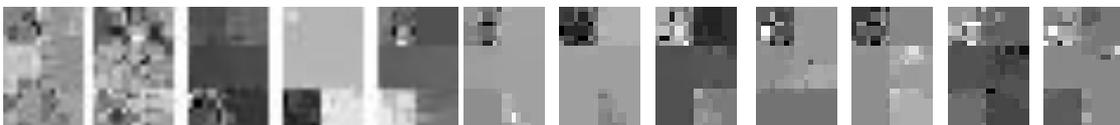


Fig.13 Result of the edge-preserving smoothing of the channels

## 6. Conclusions

A wavelet-based feature selection was successfully applied on real-world hyperspectral data. The validity of this approach was verified by an extensive simulation of spectra of the same basic form and characteristics as THz transmittance spectra. Using a certain level of Daubechies10 wavelets has furthermore proven to outperform the usage of other wavelet

basis functions. The number of features could be reduced by 95%.

To illustrate these findings, the method was applied to a number of THz imaging measurements of different chemical compounds. Hierarchical clustering was used to classify these compounds. The classification on the basis of the reduced feature set lead to a similar or even more clearly differentiating result as the classification based on the full spectra. The findings of a correlation analysis based on the simulated spectra are thereby confirmed. Furthermore, the advantage of the dimensionality reduction was illustrated by the application of a Perona-Malik smoothing filter to the reduced feature set.

We conclude that using wavelet coefficients instead of the whole spectrum is an adequate method for dimensionality reduction in hyperspectral THz imaging. In addition the simulation of THz spectra is well applicable for the evaluation of feature selection methods and should be further used to improve the algorithms used for that purpose.

## Acknowledgement

We thank the colleagues of the Department of Knowledge-Based Mathematical Systems at the JKU, especially Erich Peter Klement for discussion and support as well as Karin Wiesauer and Stefan Katletz from RECENDT GmbH and the Image Processing Department of the Fraunhofer ITWM, Kaiserslautern. Part of this work was supported by the BMWi German Federal Ministry of Economics and Technology (THESEUS program, use case ORDO) and the BMBF German Federal Ministry of Education and Research (TEKZAS program).

## References

- [1] H. Hoshina, Y. Sasaki, A. Hayashi, C. Otani, and K. Kawase, "Non Invasive Mail Inspection System with Terahertz Radiation", *Applied Spectroscopy*, 63(1), 81–86, (2009).
- [2] C.I. Chang. *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Springer, (2003).
- [3] K. Masood, N. Rajpoot, K. Rajpoot, and H. Qureshi, "Hyperspectral Colon Tissue Classification using Morphological Analysis", In *International Conference on Emerging Technologies, ICET'06*, 735–741, (2006).
- [4] R. Salomon, S. Dolberg, and SR Rotman, "Automatic Clustering of Hyperspectral Data", In *2006 IEEE 24th Convention of Electrical and Electronics Engineers in Israel*, 330–333, (2006).
- [5] T.N. Tran, R. Wehrens, and L.M.C. Buydens. SpaRef, "a Clustering Algorithm for Multispectral Images", *Analytica Chimica Acta*, 490(1-2), 303–312, (2003).
- [6] T. Warren Liao, "Clustering of Time Series Data - a Survey", *Pattern Recognition*, 38: 1857– 874, (2005).
- [7] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering, "A review. *ACM Comput. Surv*", 31(3): 264–323, (1999).
- [8] P. Berkhin, "Grouping Multidimensional Data, chapter: A Survey of Clustering Data Mining Techniques, Springer Berlin Heidelberg, 25–71 (2006).
- [9] J.G. Dy and C.E. Brodley, "Feature Selection for Unsupervised Learning", *Journal of Machine Learning Research*, 5:845–889, (2004).
- [10] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection", *Journal of Machine Learning Research*, 3, 1157–1182, (2003).
- [11] Y. Chen, S. Huang, and E. Pickwell-MacPherson, "Frequency-Wavelet Domain Deconvolution for Terahertz Reflection Imaging and Spectroscopy", *Opt. Express*, 18:1177–1190, (2010).
- [12] R. Wang, L. Lihua, W. Hong, and N. Yang, "A THz Image Edge Detection Method Based on Wavelet and Neural Network", In *Ninth International Conference on Hybrid Intelligent Systems*, 420–424, (2009).

- [13] R. Galvao and T. Yoneyama, "A Competitive Wavelet Network for Signal Clustering", *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 34(2):1282–1288, (2004).
- [14] H. Zhang, T.B. Ho, Y. Zhang, and M.S. Lin, "Unsupervised Feature Extraction for Time Series Clustering Using Orthogonal Wavelet Transform", *Informatica*, 30:305–319, (2006).
- [15] C. Valens. A Really Friendly Guide to Wavelets,. Available in: <http://perso.wanadoo.fr/polyvalens/clemens/wavelets/wavelets.html>.
- [16] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, (1999).
- [17] A. Bonvalet and M. Joffre, "Terahertz Femtosecond Pulses", *Femtosecond laser pulses: principles and experiments*, (1995).
- [18] L. Duvillaret, F. Garet, and J.L. Coutaz, "Influence of Noise on the Characterization of Materials by Terahertz Time-Domain Spectroscopy", *Journal of the Optical Society of America B*, 17(3):452–461, (2000).
- [19] B.M. Fischer, "Chemische Analytik und Bildgebung mit gepulster Terahertz-Strahlung", [www.analytik-news.de](http://www.analytik-news.de), (2009).
- [20] J.L. Rodgers and W.A. Nicewander. Thirteen Ways to Look at the Correlation Coefficient", *American Statistician*, 59–66, (1988).