Complex extreme learning machine applications in terahertz pulsed signals feature sets

Article

Accepted Version

Yin, X. -X., Hadjiloucas, S. ORCID: https://orcid.org/0000-0003-2380-6114 and Zhang, Y. (2014) Complex extreme learning machine applications in terahertz pulsed signals feature sets. Computer Methods and Programmes in Biomedicine, 117 (2). pp. 387-403. ISSN 0169-2607 doi: https://doi.org/10.1016/j.cmpb.2014.06.002 Available at https://centaur.reading.ac.uk/38006/

To link to this article DOI: http://dx.doi.org/10.1016/j.cmpb.2014.06.002

Publisher: Elsevier

Complex Extreme Learning Machine Applications in Terahertz Pulsed Signals Feature Sets

X.-X. Yin [a], S. Hadjiloucas [b], and Y. Zhang [a]

[a] *Center for Biomedical Engineering and School of Electrical & Electronic Engineering, The University of Adelaide, SA 5005, Australia.*
[b] *School of Systems Engineering, University of Reading, Reading RG6 6AY, U.K.*

Abstract

This paper presents a novel approach to the automatic classification of very large data sets composed of terahertz pulse transient signals, highlighting their potential use in biochemical, biomedical, pharmaceutical and security applications. Different types of THz spectra are considered in the classification process: firstly a binary classification study of poly-A and poly-C ribonucleic acid samples is performed, this is then contrasted with a difficult multi-class classification problem of spectra from six different powder samples that although have fairly indistinguishable features in the optical spectrum they also possess a few discernable spectral features in the terahertz part of the spectrum. Classification is performed using a complex-valued extreme learning machine algorithm that take into account features in both the amplitude as well as the phase of the recorded spectra. Classification speed and accuracy are contrasted with that achieved using a support vector machine classifier. The study systematically compares the classifier performance achieved after adopting different Gaussian kernels when separate amplitude and phase signatures are presented as feature vectors for both training and testing purposes. The study confirms the utility of complex-valued extreme learning machine algorithms for classification of the very large data sets generated with current terahertz imaging spectrometers. The classifier can take into consideration heterogeneous layers within an object as would be required within a tomographic setting and is sufficiently robust to detect patterns hidden inside noisy terahertz data sets. The proposed study opens up the opportunity for the establishment of complex-valued extreme learning machine algorithms as new chemometric tools that will assist the wider proliferation of terahertz sensing technology for chemical sensing, quality control, security screening and clinic diagnosis. Furthermore, the proposed algorithm should also be very useful in other applications requiring the classification of very large datasets.

Key words: THz, complex extreme learning machine, quaternary classification, Lagrangian, multiclass classification

---

1. Introduction

Terahertz (THz or T-ray) spectrometry and spectroradiometry have become increasingly popular sensing modalities over the past two decades due in part to recent advances in continuous wave terahertz sources and detectors, but mostly due to the wide proliferation of THz time domain spectrometers (TDS). The later, utilize ultrashort laser pulse sources to perform time-resolved studies of molecular dynamics as well as explore spectroscopic imaging applications at millimetre and sub-millimetre frequencies (also known as the far-infrared part of the spectrum shown in Fig. 1). The terahertz part of the spectrum lying between the millimetre wave and infrared (100 GHz-10 THz) is particularly rich in terms of spectral features because at these frequencies we observe molecular rotation in gases, van der Waals bond or hydrogen-bond stretches and torsional bond deformations in liquids, as well as low frequency bond vibrations and phonon vibrations in crystals. Furthermore, this is a frequency range where current state-of-the-art electron-spin-resonance systems are operating [1, 2], thus paving the way for better bio-molecule sensitivity on the basis of minute deviations in a sample's electron-spin resonance according to the physico-chemical state of the solvent. The higher frequencies associated with the THz spectrum correspond to the region where overtone and combination band spectroscopy can be performed; this is particularly interesting to environmental pollutants monitoring as well as in molecular astronomy. Infrared spectroscopy is unable to access lower frequency vibrational modes, making THz spectroscopy the only possible measurement modality in the above settings. A further attractive feature of THz spectroscopy as opposed to infrared spectroscopy is that samples have lower Rayleigh

---

* X.-X. Yin, Email: xiaoxia.yin@vu.edu.au

scattering at this part of the spectrum thus making non-invasive classification tests of samples while *in situ* (e.g. anthrax spores within an envelope) more reliable.

Terahertz transient spectrometers and imaging systems differ from their optical or infrared counterparts in that the signal-to-noise ratios in the acquired spectra is low due to a very inefficient process in the generation of the THz transients (lower by a factor of 10^5 compared to infrared time-domain spectroscopy systems centered at 800 nm). This introduces significant problems in the analysis and interpretation of spectra as well as the classification of samples. A further important difference of THz spectroscopy compared to its optical or infrared counterpart is that the longer wavelengths used enable the reliable recording of the phase delay across each frequency when the radiation is transmitted or reflected through the sample. In the case of time-domain spectrometry with a TDS system, one directly obtains reliable measurements of attenuation phase delay or dispersion at each spectral bin. Complementary information to traditional spectroscopic measurements may be therefore obtained. The current consensus in the bio-medical community is that advanced classification algorithms still need to be developed to assist screening, expert diagnosis, and subsequent treatment in an automated fashion.

It must be stressed that because of the significant cost associated with the installation and operation of THz transient spectrometers many of these systems are found mostly in national labs or most well-funded Physics, Chemistry or Biology research labs worldwide. The usual mistake by managers in these facilities is that optical experts are usually employed to run these systems. This decision may be partly justified on the basis that such systems are rather complex to operate requiring good understanding of optoelectronics as well as frequent alignment of the optical components in the system before performing an experiment. This practice, however, does not address the issue that a major bottle-neck resides in the analysis of the recorded spectra. It is not uncommon for users to have a relative lack of experience in the science of Chemometrics or the management of the associated very large data sets generated by these spectrometers. As a consequence THz-transient Chemometrics and sample classification is still at its infancy. The current work addresses this shortcoming by proposing novel classification modalities as chemometric tools specifically for these systems. The challenge for any automated THz pattern recognition systems is to explore the available spectral features in the input layer of the designed classifier as these were generated directly on the basis of the sample's THz response. Most molecules show rather complex THz absorption spectra with a multitude of absorption lines. In liquid or gaseous samples, those absorption lines are subject to thermal or pressure broadening at room temperature and within an imaging setting there may also be the result of several electromagnetic interactions [3] or pseudo-coherence errors [4] that would be associated with a thickness variation of the sample across its aperture when placed in the imaging system.

In previous works [5], we have shown that using signal processing techniques, it is possible to apodize [6] and de-noise the corresponding time-domain signatures [7] or spectra or alternatively model them using a range of modelling techniques adopted from the systems identification literature [8, 9]. We have further demonstrated that such analysis may be directly performed in the time, frequency or even the wavelet domains [10–13]. The current study is concerned with the classification of T-ray measurements on the basis of extracted features from their spectral signatures only. In this respect, the work follows directly to that performed by [14] where from the spectrum, a set of values were extracted as features, to be used as inputs to the classifier. In contrast, in the current work, we use directly the complex values associated with the Fourier transform of the time domain signatures, after taking into consideration separately the real and imaginary parts of the transform. Our goal is to explore the use of a complex value Extreme Learning Machine (ELM) to classify the complex-valued THz datasets using complex valued labels.

The procedure is in many respects analogous to quaternary classification, where complex coupled hyper-planes are defined to accommodate the output of the classifier. The formulation uses a complex input space for the spectral signatures as well as optimisation variables that are all complex valued. In contrast to classic Support Vector machine (SVM) algorithms, complex-valued ELMs address the complex valued hyper-planes through the calculation of the smallest norm of output weights with the smallest training error. In this respect, the operation of this algorithm is similar to its real-valued ELM counterpart [15]. The algorithm discards the normal threshold of b found in SVM-s, without calculating support vectors. As a consequence, the complex extreme learning machine (CELM) is specifically developed to address complex valued problems for multi-class classification with a dramatically reduced computational complexity and significantly improved computational speed. A further feature of the proposed algorithm that will become apparent in the performed study is that the label for multi-class implementation of the algorithm is critical in the complex-valued classification process. An additional requirement in the real and imaginary datasets to belong in the same class is imposed for the correct operation of the classifier. This approach avoids over fitting problems. Because inter-relations of the data at the input space are retained, the proposed approach is expected to lead to improved classifier performance compared with its real valued ELM counterpart.

For illustration purposes, different types of THz spectra are considered in the classification process: firstly a binary classification study of poly-A and poly-C Ribo-Nucleic Acid (RNA) samples is performed. This will then be contrasted with a difficult multi-class classification task of THz spectra belonging to six different powder samples. These samples, although have fairly indistinguishable features in the optical spectrum they also possess a few discernable spectral features in the THz part of the spectrum so that their classification can be performed.

The paper is organized as follows: Section II places the work within a THz bio-medical imaging context. Section III provides a description of a traditional terahertz imaging system enabling the reader to understand the general methodology in this sensing modality and the origin of the complex valued data-sets. In Section IV, after reviewing the basics of a quaternary classification scheme, we describe the CELM classifier. In Section V, we present THz measurements for the RNA and powdered samples. Section VI discusses binary and multiclass classification of the RNA and powder samples respectively, performed using CELM, contrasting this modality to that of real-valued ELM and SVM. The study systematically compares the classifier performance achieved after adopting different Gaussian kernels when separate amplitude and phase signatures are presented as feature vectors for both training and testing purposes. Section IV provides a conclusion of the work and some closing remarks.

2. Placement of the proposed family of classifiers within a THz bio-medical imaging context

The THz part of the spectrum is of direct relevance to the biomedical sciences, because complementary information to traditional spectroscopic measurements may be obtained. The vibrational spectral characteristics of many bio-molecules, lie in this range (wavenumbers between 3.3-333 cm^{-1}). Since THz photons, (or T-rays), have significantly lower energies (e.g., only 0.04 meV at 100 GHz) than X-rays, they have been considered by many as non-invasive. The interest in adopting THz radiation to perform imaging in a biomedical setting stems from the fact that it is considered as non-invasive. When THz pulses interact with biological tissue, the Gibbs free energy conveyed in the THz beam is insufficient to drive chemical reactions. The molar energy at a frequency f of 100 GHz is given from $E = Nhf$ where $N = 6.023 \times 10^{23}$ mol^{-1}, (Avogadro's number), and $h = 6.626 \times 10^{-34}$ Js (Planck's constant), and the calculated value of $E = 0.04$ 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 one may assume that any interference with biochemical processes should be minimal. Therefore, T-ray spectrometry is a very promising new sensing modality rapidly gaining momentum for clinical diagnosis.

A further advantage of performing imaging based on the optical properties of biological tissue with THz radiation is the associated lower scattering compared to infrared light. 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 spill-over around most tissues and organs. From a technological point of view, THz imaging needs to compete with positron emission tomography (PET) imaging that has pico-molar sensitivity but poor spatial resolution and magnetic resonance imaging (MRI) that offers milli-molar sensitivity with high spatial resolution. Indeed, a diffraction limited imaging system operating at 1 THz would have a spatial resolution of 300 μm, which should be considered sufficient for many biomedical applications. Meanwhile, since 70% of the human body is composed of water, a large part of the energy in the excitation pulse is attenuated. As a consequence, biomedical samples may only be identified through the application of advanced signal processing techniques for feature extraction and pattern classification.

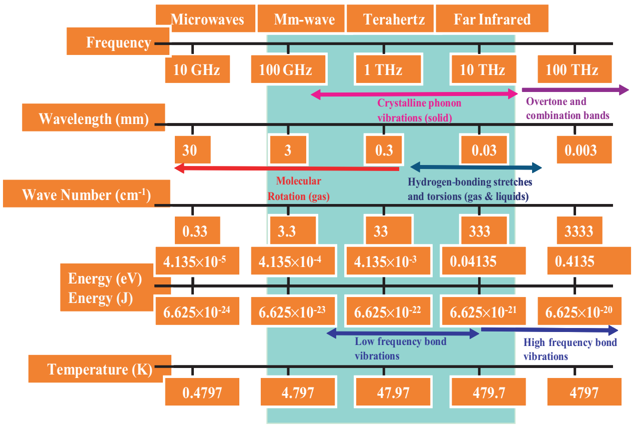


Fig. 1. Illustration of the THz part of the spectrum as related to other parts of the spectrum.

From a biomedical perspective, THz contrast imaging is also becoming an increasingly important modality as it can differentiate between samples on the basis of their water content, this is particularly important in applications where other imaging modalities may not be used (e.g. as a substitute for X-ray mammography diagnosis and cancer patient screening on the basis of breast tissue vascularization when the subjects are pregnant or lactating women). THz imaging has also shown significant potential for applications in both *in vivo* and *ex vivo* environments (e.g. the diagnosis of skin cancers). Although non-linear interactions between biological tissue and coherent THz radiation have been predicted by [16] and experimentally verified by the careful work of Grundler and the analysis of Kaiser [17] in the '90s, and more recently observed in yeast cells [10] the widely held view at the moment is that any measurement technique that operates within acceptable specific absorption rates is currently deemed as safe for bio-medical investigations. This motivates the current work, placing the proposed classifier algorithms in a quality-control or diagnostics setting.

3. Dataset generation using a THz imaging spectrometer in transmission configuration

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 ZnTe 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 ZnTe electro-optic (EO) detector crystal. 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 though the zero path difference of the two arms of the interferometer, the cross correlation of the optical and THz signal is obtained. The EO 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. The signal is measured with the use of a balanced optical photo-detection scheme operating on the two orthogonally-polarized spatially-separated signals coming out of the Wollaston prism. A lock-in amplifier (LOI) 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 direction, while keeping it at the focal plane of the parabolic mirrors. A diagram of the setup used for the measurements is shown in Fig. 2. Terahertz-transient images are composed of several pixels, where each pixel contains information of the attenuation, phase delay and dispersion of a sample relative to a background signal as recorded when a THz pulse is propagating in free space through the system (with the measurement port of the spectrometer evacuated). The Fourier transformed signal of the sample interferogram is normally ratioed with that of the background interferogram to obtain the complex insertion loss of the sample, this provides a measure of attenuation, phase delay and dispersion of the sample across each frequency. Although data acquisition of the time-domain signatures and processing at each pixel is usually performed in real time during the image acquisition process, further processing such as Fourier transformation and insertion loss measurements of the associated spectra is normally performed off-line[18]. The additional calculation steps of sample refractive index and absorption coefficients are also commonly performed off-line. Phase information is retrieved by varying the time delay between the THz wave and the probe beam [18]. For materials sufficiently transparent to terahertz radiation, one measures the transmitted responses and acquires spectral information to produce contrast images. For highly attenuating or opaque samples a reflection set-up may be used instead.

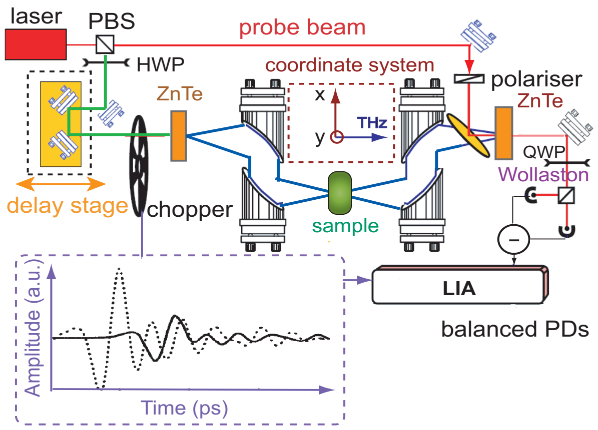


Fig. 2. Diagram of the electro-optic transmission THz-imaging set-up with ZnTe crystals used for both EO generation and detection, as illuminated by the femtosecond laser. The inset plot depicts time-domain THz-transient waveforms corresponding to sample and background interferograms (denoted by solid and dotted curves, respectively).

The solid curve in the inset of Fig. 2 depicts typical attenuation, phase delay and dispersion (pulse broadening) of the sample when compared to the background signature (dotted curve) for a single pixel of the imaging system. The observed phase delay is a measure of the average refraction index of the sample, whereas the broadening is caused by dispersion and frequency-dependent attenuation of the sample.

4. Classification Methodology

Kernel based learning and support vector machine (SVM) methodologies reside at the core of a range of interdisciplinary challenges. Their formulation shares concepts from different disciplines such as: linear algebra, mathematics, statistics, signal processing, systems and control theory, optimization, machine learning, pattern recognition, data mining and neural networks. This paper extends the formulation of a very important class of recently developed classifiers called Extreme Learning Machines (ELMs) to complex valued problems [15, 19]. The motivation for the proposed extension stems from the fact that the real valued EML has shown some of the lowest training errors among machine learning algorithms and in particular support vector machines classifiers (SVMs) [5, 20–22]. By extending ELMs to complex inputs, their applications domain can dramatically increase, encompassing all types of research associated to the study of the interaction of matter with waves, and in particular spectroscopy (acoustic, dielectric, optical, terahertz, infrared, electron-spin resonance, nuclear magnetic or paramagnetic resonance, etc.) as well as imaging and tomography modalities encountered across the Physical, Chemical and Biomedical disciplines. As a consequence, the proposed extension is fundamental both from a Machine Learning as well as from a Chemometrics perspective [23]. Because the above relations are also analogous to the blurring function (relating amplitude and phase) developed by Bode [24] to describe the dynamics of physical systems, such extension has a wide range of applications across all physical sciences. In this context, this study focuses on the use of CELM to perform binary and multi-class classification of RNA and powder samples respectively on the basis of images acquired by a THz-transient imaging spectrometer. The analysis is performed on large data sets as would be the case in a typical bio-medical or quality control setting. Classification is performed on the basis of discernable features in the measured THz spectra.

4.1. *Complex Valued ELM*

CELM adopts induced complex RKHS kernels [25] to map inputs from complex-valued non-linear spaces to other real valued higher dimensional linear spaces. This permits us to classify the inputs with linear complex valued feature vectors. It involves the aspects of quaternary classification, through the introduction of two complex-coupled hyperplanes [26]. For 2D inputs, the complex-coupled hyperplanes are used together to divide the input space into four partitions. Such approach is further supported by the work by Bouboulis et. al. [25], where within a SVM context, they showed that a derived 2D complex kernel is equivalent to an induced real kernel, formed as a linear combination of two identical 2D real value kernels. A widely linear estimation processing approach is adopted and the argument composed of the sum of the two parts (real and imaginary) is employed to model the output weights connecting the hidden layer with the feature mapping of the input into the hidden-layer feature space.

4.1.1. *A quaternary classification problem*

An important step of machine learning is to find hyperplanes that separate the space in relation to different target classes. According to ELM [15], in any real Hilbert space $\mathcal{H}$, a hyper-plane consists of all elements $\mathrm{f} \in \mathcal{H}$ that satisfy $\langle \mathrm{f}, \omega \rangle_{\mathcal{H}} = 0$, for some $\omega \in \mathcal{H}$. The approach differs from real valued supports vector machines (SVM), since in the real valued ELM, the offset b of the hyper-plane from the origin has been removed.

In order to be able to generalize the ELM rational to complex space, we adopt the method proposed by [25], and define a complex hyper-plane that divides the complex space $\hat{\mathcal{H}}$ into four parts through the introduction of a Hermitian matrix, label $*$. This enables us to classify objects into four classes (instead of two). This approach is also supported in Bouboulis's et. al. article [25] where it was postulated that $\langle \hat{\mathrm{f}}, \hat{\omega} \rangle_{\hat{\mathcal{H}}} = \langle \hat{\mathrm{f}}^{\Re}, \hat{\omega}^{\Re} \rangle_{\mathcal{H}} + \langle \hat{\mathrm{f}}^{\Im}, \hat{\omega}^{\Im} \rangle_{\mathcal{H}} + \mathcal{J}(\langle \hat{\mathrm{f}}^{\Im}, \hat{\omega}^{\Re} \rangle_{\mathcal{H}} - \langle \hat{\mathrm{f}}^{\Re}, \hat{\omega}^{\Im} \rangle_{\mathcal{H}})$ where $\hat{\mathcal{H}} = \mathcal{H}^2$, $\hat{\mathrm{f}}, \hat{\omega}$ indicates a complex decision function and the corresponding margin of the complex hyper-plane $\hat{\mathcal{H}}$. Here, symbol $\langle \cdot \rangle_{\mathcal{H}}$ is used to denote inner product in the corresponding real valued input space. Symbols of $\Re$ and $\Im$ indicate the real and imaginary parts. This is a real valued kernel function: $\langle \kappa(\cdot, \mathbf{X}), \kappa(\cdot, \mathbf{Y}) \rangle_{\mathcal{H}}$. The kernel $\kappa(\cdot, \mathbf{X})$ is used for a feature map of real valued input space $\mathcal{H}$, labeled by $\psi(\mathbf{X})$. According to [25], the corresponding complex Gaussian kernel is defined as:

$$\widehat{\kappa}^{j}_{\sigma_j, \mathcal{C}^d}(\widehat{\mathbf{Z}}, \widehat{\omega}) := \exp\left(- \frac{\sum_{k=1}^{d} (\widehat{z}_k - \widehat{\omega}^*_k)^2}{\sigma_j^2} \right), \tag{1}$$

where $\widehat{\kappa}$ denotes a complex valued kernel, $\widehat{\mathbf{Z}}, \widehat{\omega} \in \mathcal{C}^d$, $d \in \mathcal{N}$ or infinite, and $\widehat{\omega}$ denotes a complex weight (margin), with $*$ for a Hermitian matrix, $\widehat{z}_k$ denotes the k-th component of the complex vector $\widehat{\mathbf{Z}} \in \mathcal{C}^d$ and $\exp(\cdot)$ is the extended exponential function in the complex domain. Here, $\widehat{\kappa}^j$ indicates the j-th complex kernel function, depending on the value of kernel parameter σ_j, which is varied due to a different input for normal machine learning procedure, and therefore time consuming. A proposed method is to fix the value of kernel parameter σ_j to σ for all kernels in order to simplify computation [27].

We use symbols $\hat{\psi}(\hat{\mathbf{Z}})$ to denote complex feature mapping in this context. In this section and following, in addition to the boldfaced symbols for the vector and matrix valued quantities, the complex valued quantities are related to matrix quantities, with specific subscripts to describe the row and/or column of the complex valued matrix.

Definition 1. The complex machine learning task (i.e. CELM) is equivalent to two real machine learning tasks, i.e. (ELM) employing the two real kernels 2κ.

Definition 2. Let $\hat{\mathcal{H}}$ be a complex Hilbert space. The complex couple of hyper-planes is defined as the set of all $\mathrm{f} \in \hat{\mathcal{H}}$ that satisfy one of the following relations:

$$\Re(\langle \hat{\mathrm{f}}_L, \hat{\omega} \rangle_{\hat{\mathcal{H}}} + \langle \hat{\mathrm{f}}_{L^*}, \hat{\nu} \rangle_{\hat{\mathcal{H}}}) = 0 \tag{2}$$

$$\Im(\langle \hat{\mathrm{f}}_L, \hat{\omega} \rangle_{\hat{\mathcal{H}}} + \langle \hat{\mathrm{f}}_{L^*}, \hat{\nu} \rangle_{\hat{\mathcal{H}}}) = 0 \tag{3}$$

for some $\hat{\omega}, \hat{\nu} \in \hat{\mathcal{H}}$, where $\hat{\mathrm{f}}_L \in \hat{\mathcal{H}}$ represents two hyperplanes of the doubled real space, $\mathcal{H}^2$.

The input space is divided into four partitions after defining the complex-couple hyper-planes as defined above, on the basis of the positive and negative values of the two hyper-planes indicated by the left sides of the expressions in (2). These are: $H_{++} = \{\Re > 0, \Im > 0\}$; $H_{+-} = \{\Re > 0, \Im < 0\}$; $H_{-+} = \{\Re < 0, \Im > 0\}$; $H_{--} = \{\Re < 0, \Im < 0\}$.

4.1.2. *Multiclass Classification by a CELM*

Similar to ELM, the complex valued extreme meaning learning is an extension of single-hidden-layer feed-forward networks (SLFNs), where the hidden layer need not be tuned. Training of the classifier from the available data $\hat{\mathbf{Z}}$ is performed from complex-valued input space to complex-valued feature space through a feature map $\psi(\hat{\mathbf{Z}})$. In this context, we use the symbol ˆ to indicate the complex valued parameters. The goal of the complex machine learning task is to estimate a complex couple of maximum margin hyperplanes. According to the work in [25], for a 2D simple case, we aim to minimize:

$$\left\|\begin{matrix}\hat{\omega}^r + \hat{\nu}^r \\ \hat{\omega}^j - \hat{\omega}^j\end{matrix}\right\|^2_{\mathcal{H}^2} + \left\|\begin{matrix}-(\hat{\omega}^j + \hat{\nu}^j) \\ \hat{\omega}^r - \hat{\nu}^r\end{matrix}\right\|^2_{\mathcal{H}^2} = 2(\|\hat{\omega}\|^2_{\mathcal{H}} + \|\hat{\nu}\|^2_{\mathcal{H}}).$$

Given a training data set $(\hat{z}_n, \hat{\vartheta}_n)$, and $\hat{\theta} = [\hat{\vartheta}_1, ..., \hat{\vartheta}_N]^T$ with $\hat{\vartheta}_n$ $(n \in 1, ..., N)$ known complex labels with m classes, CELM aims to simultaneously minimize the training error $\|\psi\hat{\omega} + \psi^*\hat{\nu} - \hat{\theta}\|^2$ and the norm of the output weights $\|\hat{\omega}\|_{\mathcal{H}} + \|\hat{\nu}\|_{\mathcal{H}}$. The hidden-layer feature mapping matrix ψ is represented as:

$$\psi = \begin{bmatrix} \varphi_1(\hat{z}_1) & \cdots & \varphi_L(\hat{z}_1) \\ \vdots & \vdots & \vdots \\ \varphi_1(\hat{z}_N) & \cdots & \varphi_L(\hat{z}_N) \end{bmatrix} \tag{4}$$

where the dimension of ψ is set by the numbers of training samples N and the number or hidden nodes L, irrespective of the number of output nodes (number of classes), and $\{\hat{z}_1, ..., \hat{z}_N\} \in \hat{\mathbf{Z}}$.

For an m class classier with m output nodes where $m > 1$, the classification problem (denoted by $\hbar_P$) for CELM can be formulated as:

$$\min_{(\hat{\omega},\hat{\nu},C)} : \hbar_P = \left(\frac{1}{2} \parallel \hat{\omega} \parallel^2_{\hat{\mathcal{H}}} + \frac{1}{2} \parallel \hat{\nu} \parallel^2_{\hat{\mathcal{H}}} + \frac{C}{N}\sum_{n=1}^{N}(\hat{\delta}^2)\right), \tag{5}$$

Subject to:

$$\begin{cases} \Re(\langle\psi_{\hat{\mathcal{H}}}(\hat{z}_n), \hat{\omega}\rangle + \langle\psi^*_{\hat{\mathcal{H}}}(\hat{z}_n), \hat{\nu}\rangle) \geq \hat{\theta}^r_n - \hat{\delta}^r_n \\ \Im(\langle\psi_{\hat{\mathcal{H}}}(\hat{z}_n), \hat{\omega}\rangle + \langle\psi^*_{\hat{\mathcal{H}}}(\hat{z}_n), \hat{\nu}\rangle) \geq \hat{\theta}^j_n - \hat{\delta}^j_n \end{cases} \tag{6}$$

where C is a parameter given by the user. There is a trade-off between the distance in relation to the separating margin and the training error.

Using positive Lagrangian multiplies **a** and **b**, the associated Lagrangian function becomes

$$\mathrm{L}(\hat{\omega}, \hat{\nu}, \mathbf{a}, \mathbf{b}) = \frac{1}{2} \parallel \hat{\omega} \parallel^2_{\hat{\mathcal{H}}} + \frac{1}{2} \parallel \hat{\nu} \parallel^2_{\hat{\mathcal{H}}} + \frac{C}{N} \parallel \hat{\delta}^r_{n,\rho} + \hat{\delta}^j_{n,\rho} \parallel^2$$
$$- \sum_{n=1}^{N}\sum_{\rho=1}^{m} a_{n,\rho}\left(\Re\left(\langle\psi_{\hat{\mathcal{H}}}(\hat{z}_n), \hat{\omega}_\rho\rangle + \langle\psi^*_{\hat{\mathcal{H}}}(\hat{z}_n), \hat{\nu}_\rho\rangle\right) - \hat{\vartheta}^r_{n,\rho} + \hat{\delta}^r_{n,\rho}\right)$$
$$- \sum_{n=1}^{N}\sum_{\rho=1}^{m} b_{n,\rho}\left(\Im\left(\langle\psi_{\hat{\mathcal{H}}}(\hat{z}_n), \hat{\omega}_\rho\rangle + \langle\psi^*_{\hat{\mathcal{H}}}(\hat{z}_n), \hat{\nu}_\rho\rangle\right) - \hat{\vartheta}^j_{n,\rho} + \hat{\delta}^j_{n,\rho}\right)$$

Here, $\hat{\theta}_n = \{\hat{\vartheta}_{n,\rho}\}$ with $\rho = 1, ..., m$, where $\{\hat{\vartheta}_{n,\rho}\}$ denotes the output value of the ρ output node for the training data $\hat{z}_n$ and m labels the number of the classes of the output. When both the real and imaginary parts of the Sth element $\hat{\vartheta} n, s$ are one and the remaining of $\hat{\vartheta}_n$ are zero, we attribute to this class the designation $S + \mathcal{J}S$. Using Wirtinger's calculus to compute the respective gradients, we have:

$$\frac{\partial \mathrm{L}}{\hat{\omega}^*_\rho} = \frac{1}{2}\hat{\omega}_\rho - \frac{1}{2}\sum_{n=1}^{N} a_{n,\rho}\psi^T_{\hat{\mathcal{H}}}(\hat{z}_n) + \frac{\mathcal{J}}{2}\sum_{n=1}^{N} b_{n,\rho}\psi^T_{\hat{\mathcal{H}}}(\hat{z}_n) = 0$$
$$\Rightarrow \qquad \hat{\omega}_\rho = \sum_{n=1}^{N}\left(\alpha_{n,\rho} - \mathcal{J}b_{n,\rho}\right)\psi^T_{\hat{\mathcal{H}}}(\hat{z}_n) \tag{7}$$

$$\frac{\partial \mathrm{L}}{\hat{\nu}^*_\rho} = \frac{1}{2}\hat{\nu}_\rho - \frac{1}{2}\sum_{n=1}^{N} a_{n,\rho}{\psi^*_{\hat{\mathcal{H}}}}^T(\hat{z}_n) + \frac{\mathcal{J}}{2}\sum_{n=1}^{N} b_{n,\rho}{\psi^*_{\hat{\mathcal{H}}}}^T(\hat{z}_n) = 0$$
$$\Rightarrow \qquad \hat{\nu}_\rho = \sum_{n=1}^{N}\left(a_{n,\rho} - \mathcal{J}b_{n,\rho}\right)\psi^{*T}_{\hat{\mathcal{H}}}(\hat{z}_n) \tag{8}$$

$$\begin{aligned} \frac{\partial \mathrm{L}}{\hat{\delta}^r_{n,\rho}} &= \frac{2C}{N}\hat{\delta}^r_{n,\rho} - a_{n,\rho} = 0 \Rightarrow \hat{\delta}^r_{n,\rho} = \frac{N}{2C}a_{n,\rho} \\ \frac{\partial \mathrm{L}}{\hat{\delta}^j_{n,\rho}} &= \frac{2C}{N}\hat{\delta}^j_{n,\rho} - b_{n,\rho} = 0 \Rightarrow \hat{\delta}^j_{n,\rho} = \frac{N}{2C}b_{n,\rho} \end{aligned} \tag{9}$$

$$\frac{\partial \mathrm{L}}{a_{n,\rho}} = -\frac{1}{2}\left(\Re\left(\langle\psi_{\hat{\mathcal{H}}}(\hat{z}_n), \hat{\omega}_\rho\rangle + \langle\psi^*_{\hat{\mathcal{H}}}(\hat{z}_n), \hat{\nu}_\rho\rangle\right)\right)$$
$$-\frac{1}{2}\left(-\hat{\vartheta}^r_{n,\rho} + \hat{\delta}^r_{n,\rho}\right) = 0 \tag{10}$$

$$\frac{\partial \mathrm{L}}{b_{n,\rho}} = -\frac{1}{2}\left(\Im\left(\langle\psi_{\hat{\mathcal{H}}}(\hat{z}_n), \hat{\omega}_\rho\rangle + \langle\psi^*_{\hat{\mathcal{H}}}(\hat{z}_n), \hat{\nu}_\rho\rangle\right)\right)$$
$$-\frac{1}{2}\left(-\hat{\vartheta}^j_{n,\rho} + \hat{\delta}^j_{n,\rho}\right) = 0 \tag{11}$$

According to the last two equations,

$$\langle\psi_{\hat{\mathcal{H}}}(\hat{z}_n), \hat{\omega}_\rho\rangle + \langle\psi^*_{\hat{\mathcal{H}}}(\hat{z}_n), \hat{\nu}_\rho\rangle - \hat{\vartheta}_{n,\rho} + \hat{\delta}_{n,\rho} = 0 \tag{12}$$

By substituting Eqn. 7, Eqn. 8 and Eqn. 9, Eqn. 12 can be written as:

$$(\mathbf{a} - \mathcal{J}\mathbf{b})\left(\psi_{\hat{\mathcal{H}}}(\hat{z}_n)\psi^T_{\hat{\mathcal{H}}}(\hat{z}_n) + \psi^*_{\hat{\mathcal{H}}}(\hat{z}_n)\psi^{*T}_{\hat{\mathcal{H}}}(\hat{z}_n)\right)$$
$$+\frac{N}{2C}(\mathbf{a} + \mathcal{J}\mathbf{b}) = \hat{\theta} \tag{13}$$

The real part of the output is:

$$\begin{aligned} \theta^{\mathbf{r}} &= \mathbf{a}\Re\left(\psi_{\hat{\mathcal{H}}}\psi^T_{\hat{\mathcal{H}}} + \psi^*_{\hat{\mathcal{H}}}\psi^{*T}_{\hat{\mathcal{H}}}\right) + \frac{N}{2C}\mathbf{a} \\ &= \mathbf{a}\left(\Re\left(\psi_{\hat{\mathcal{H}}}\psi^T_{\hat{\mathcal{H}}} + \psi^*_{\hat{\mathcal{H}}}\psi^{*T}_{\hat{\mathcal{H}}}\right) + \frac{N}{2C}\mathbf{I}\right) \end{aligned} \tag{14}$$

whereas the imaginary part of the output is:

$$\begin{aligned} \theta^{\mathbf{j}} &= -\mathbf{b}\Im\left(\psi_{\hat{\mathcal{H}}}\psi^T_{\hat{\mathcal{H}}} + \psi^*_{\hat{\mathcal{H}}}\psi^{*T}_{\hat{\mathcal{H}}}\right) + \frac{N}{2C}\mathbf{b} \\ &= \mathbf{b}\left(-\Im\left(\psi_{\hat{\mathcal{H}}}\psi^T_{\hat{\mathcal{H}}} + \psi^*_{\hat{\mathcal{H}}}\psi^{*T}_{\hat{\mathcal{H}}}\right) + \frac{N}{2C}\mathbf{I}\right) \end{aligned} \tag{15}$$

By substituting Eqn. 14 and Eqn. 15 to Eqn. 7 and Eqn. 8, the real and imaginary parts of the output weights are written as:

$$\omega^{\mathbf{r}} = \Re\left(\psi_{\hat{\mathcal{H}}}^{T}\right)\mathbf{a}$$

$$= \Re\left(\psi_{\hat{\mathcal{H}}}^{T}\right)\left(\Re\left(\psi_{\hat{\mathcal{H}}}\psi_{\hat{\mathcal{H}}}^{T} + \psi_{\hat{\mathcal{H}}}^{*}\psi_{\hat{\mathcal{H}}}^{*T}\right) + \frac{N}{2C}\mathbf{I}\right)^{-1}\theta^{\mathbf{r}} \tag{16}$$

$$\omega^{\mathbf{j}} = -\Im\left(\psi_{\hat{\mathcal{H}}}^{T}\right)\mathbf{b}$$

$$= \Im\left(\psi_{\hat{\mathcal{H}}}^{T}\right)\left(\Im\left(\psi_{\hat{\mathcal{H}}}\psi_{\hat{\mathcal{H}}}^{T} + \psi_{\hat{\mathcal{H}}}^{*}\psi_{\hat{\mathcal{H}}}^{*T}\right) - \frac{N}{2C}\mathbf{I}\right)^{-1}\theta^{\mathbf{j}} \tag{17}$$

$$\nu^{\mathbf{r}} = \Re\left(\psi_{\hat{\mathcal{H}}}^{*T}\right)\mathbf{a}$$

$$= \Re\left(\psi_{\hat{\mathcal{H}}}^{*T}\right)\left(\Re\left(\psi_{\hat{\mathcal{H}}}\psi_{\hat{\mathcal{H}}}^{T} + \psi_{\hat{\mathcal{H}}}^{*}\psi_{\hat{\mathcal{H}}}^{*T}\right) + \frac{N}{2C}\mathbf{I}\right)^{-1}\theta^{\mathbf{r}} \tag{18}$$

$$\nu^{\mathbf{i}} = -\Im\left(\psi_{\hat{\mathcal{H}}}^{*T}\right)\mathbf{b}$$

$$= \Im\left(\psi_{\hat{\mathcal{H}}}^{*T}\right)\left(\Im\left(\psi_{\hat{\mathcal{H}}}\psi_{\hat{\mathcal{H}}}^{T} + \psi_{\hat{\mathcal{H}}}^{*}\psi_{\hat{\mathcal{H}}}^{*T}\right) - \frac{N}{2C}\mathbf{I}\right)^{-1}\theta^{\mathbf{j}} \tag{19}$$

The output decision functions of the CELM classifier are:

$$\Re(\hat{\mathrm{f}}_{L}(\hat{\mathbf{Z}})) = \Re\left(\psi_{\hat{\mathcal{H}}}(\hat{\mathbf{Z}}), \hat{\omega}\right) = \Re$$

$$\left(\psi_{\hat{\mathcal{H}}}(\hat{\mathbf{Z}})\psi_{\hat{\mathcal{H}}}^{T}\left(\left(\psi_{\hat{\mathcal{H}}}\psi_{\hat{\mathcal{H}}}^{T} + \psi_{\hat{\mathcal{H}}}^{*}\psi_{\hat{\mathcal{H}}}^{*T}\right) + \frac{N}{2C}\mathbf{I}\right)^{-1}\hat{\theta}\right) \tag{20}$$

$$\Im(\hat{\mathrm{f}}_{L}(\hat{\mathbf{Z}})) = \Im\left(\psi_{\hat{\mathcal{H}}}(\hat{\mathbf{Z}}), \hat{\omega}\right) = -\Im$$

$$\left(\psi_{\hat{\mathcal{H}}}(\hat{Z})\psi_{\hat{\mathcal{H}}}^{T}\left(\left(\psi_{\hat{\mathcal{H}}}\psi_{\hat{\mathcal{H}}}^{T} + \psi_{\hat{\mathcal{H}}}^{*}\psi_{\hat{\mathcal{H}}}^{*T}\right) + \frac{N}{2C}\mathbf{I}\right)^{-1}\hat{\theta}\right) \tag{21}$$

$$\Re(\hat{\mathrm{f}}_{L^*}(\hat{\mathbf{Z}})) = \Re\left(\psi_{\hat{\mathcal{H}}}^{*}(\hat{\mathbf{Z}}), \hat{\nu}\right) = \Re$$

$$\left(\psi_{\hat{\mathcal{H}}}^{*}(\hat{\mathbf{Z}})\psi_{\hat{\mathcal{H}}}^{*T}\left(\left(\psi_{\hat{\mathcal{H}}}\psi_{\hat{\mathcal{H}}}^{T} + \psi_{\hat{\mathcal{H}}}^{*}\psi_{\hat{\mathcal{H}}}^{*T}\right) + \frac{N}{2C}\mathbf{I}\right)^{-1}\hat{\theta}\right) \tag{22}$$

$$\Im(\hat{\mathrm{f}}_{L^*}(\hat{Z})) = \Im\left(\psi_{\hat{\mathcal{H}}}^{*}(\hat{Z}), \hat{\nu}\right) = -\Im$$

$$\left(\psi_{\hat{\mathcal{H}}}^{*}(\hat{\mathbf{Z}})\psi_{\hat{\mathcal{H}}}^{*T}\left(\left(\psi_{\hat{\mathcal{H}}}\psi_{\hat{\mathcal{H}}}^{T} + \psi_{\hat{\mathcal{H}}}^{*}\psi_{\hat{\mathcal{H}}}^{*T}\right) + \frac{N}{2C}\mathbf{I}\right)^{-1}\hat{\theta}\right) \tag{23}$$

where $\psi_{\hat{\mathcal{H}}}\psi_{\hat{\mathcal{H}}}^{T}$ and $\psi_{\hat{\mathcal{H}}}^{*}\psi_{\hat{\mathcal{H}}}^{*T}$ are $N \times N$ or $L \times L$ matrices, according to the size of the inputs. When $m = 1$, the predicted class label of sample $\hat{\mathbf{Z}}$ is:

$$\mathrm{label}(\hat{\mathbf{Z}}) = \mathrm{sign}(\langle\hat{\psi}_{\hat{\mathcal{H}}}(\hat{\mathbf{Z}}), \hat{\omega}\rangle + \langle\hat{\psi}_{\hat{\mathcal{H}}}^{*}(\hat{\mathbf{Z}}), \hat{\nu}\rangle) \tag{24}$$

When $m > 1$, the predicted class label of sample $\hat{\mathbf{Z}}$ is:

$$\mathrm{label}(\hat{\mathbf{Z}}) = \arg\max_{n=1,..,m}(\langle\psi_{\hat{\mathcal{H}}}(\hat{\mathbf{Z}}), \hat{\omega}\rangle + \langle\psi_{\hat{\mathcal{H}}}^{*}(\hat{Z}), \hat{\nu}\rangle) \tag{25}$$

where $\mathrm{label}(\hat{\mathbf{Z}}) = \mathrm{label}(\Re(\hat{Z})) + \mathcal{J}\mathrm{label}(\Im(\hat{\mathbf{Z}}))$. Here, we employ the induced real kernel $2\hat{\kappa}^{r}$ instead of the complex kernel $\hat{\kappa}$ for the solution of the complex labelling function.

4.2. *Classifier design for the identification of T-ray spectra*

4.2.1. *Binary classifier design for the identification of poly-A and poly-C T-ray spectra*

Currently, the identification of the binding state of DNA is an emergent interdisciplinary research topic within the THz community, because it promises a label-free modality for the determination of the four base pairs; furthermore it has the potential to eliminate the polymerase chain reaction (PCR) step commonly associated with a sequencing process. In spite of the lack of characteristic absorption features in the T-ray region, it is possible to discriminate un-hybridized from hybridized DNA strands on the basis of observed loading (scattering parameters) of samples deposited on planar micro-fabricated T-ray resonators [28–31]. Furthermore, there have been suggestions that proteins can be detected by T-ray circular dichroism (TCD) spectroscopy, because many bio-molecules in crystalline form exhibit strong and specific features in their dielectric spectra [28] different from their phonon resonances.

In the current study, spectra from two different RNA polymer strands, polyadenylic acid (poly-A), and polycytidylic acid (poly-C) are used as inputs for the binary classification task. Commercially available poly-A and poly-C potassium salts are used for the experiment (Sigma–Aldrich, product numbers P9403 and P4903). The experimental data sets are generated under guidance with personnel in the time-domain THz facility at the University of Freiburg, Germany. Details regarding the sample crystallization protocol are descried in detail in Fischer et al. [28]. The THz image illustrated in Fig. 3, is created using a THz time-domain spectroscopy imaging system based on free-space propagation and aperture-less focusing of the T-ray beam. Each pixel in the image represents the normalized peak values corresponding to Poly-A and Poly-C. The sample consists of a 4 × 4 array of spots. Two of the spots are removed from the substrate in order to identify the orientation of the substrate in the image. The spot of Poly-A is shown at the top left corner of the image, it shows weak transmission, compared with the spots of poly-C. The positions of poly-A and poly-C sub-images are labelled in the diagram to the right of the picture. Based on the positions of poly-A and poly-C within the terahertz image, we select eight neighbouring pixels around a center pixel position from each spot for signal post-processing and classification. The pixels lying on the boundaries of each class are excluded from training and test vectors.

4.2.2. *Multi-class classifier design for the identification of powder sample spectra*

In the second example, a multi-class classification problem is considered. 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 powder data classification is of interest to the pharmaceutical industries

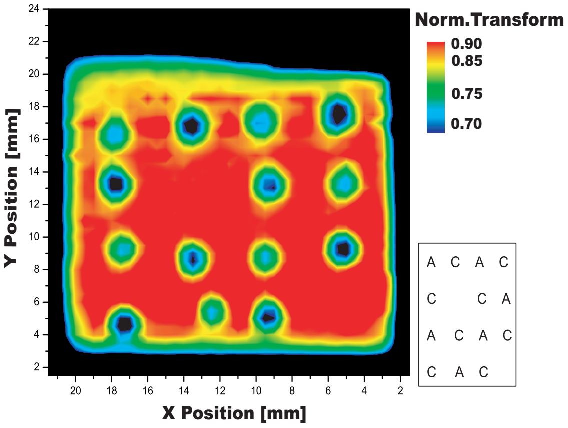


Fig. 3. T-ray transmission image of the poly-A and poly-C, showing stronger absorption in poly-C compared with poly-A. Each spot contained 200 μg of either poly-A or poly-C in alternating order, as indicated in the diagram on the right. The colour scale indicates the normalized peak values of the two RNA samples.

(for the detection of drug polymorphs and isomorphs [32–34]. Such investigations also have applications in security (e.g. fingerprinting of explosives and illicit drug detection [5, 35, 36]). Our goal is to demonstrate a generic feature extraction methodology that may be used across different THz data sets. This is of significant importance to the THz community as current data driven classifiers prohibit proper inter-comparison between results obtained in different labs. Current practice precludes the development of standards, guidelines and specifications that could be adopted by the biomedical, pharmaceutical as well as security industries, which are envisaged to become emerging markets for THz-transient spectrometers [37]. In this sense, the requirement that our proposed algorithm should perform well in two very different classification tasks (binary as well as multi-class) represents a departure from previous THz works presented in the literature. Furthermore, a universal approach to the management of the associated large data sets generated through this measurement modality can be developed.

In the current work, multi-class classification is performed for the following samples: sand, talcum, salt, powdered sugar, wheat flour, and baking soda on the basis of their recorded THz spectra. Absorbance, phase delay and dispersion of the THz pulses are directly related to sample density, concentration of absorbers as well as thickness. All samples have a 4 mm thickness and are held in a specially made sample holder with two Teflon windows. A traditional T-ray imaging system is used to detect the T-ray responses. Differential absorption is measured for each pixel with the empty cuvette providing the background signal. Images constructed from 50 pixel responses (with a pixel spacing of 100 m) can be acquired in under 30 min. Extraction of the complex insertion loss is straight-forward once these data sets are obtained [38, 39]

4.2.3. *T-ray feature extraction from frequency domain data*

Both types of classification task are performed to assess the potential of CELMs, ELMs and SVMs in T-ray pulsed classification. RBF kernels (both real and complex-valued) are applied for statistical feature mapping. Signal processing is applied to track the key features of training vectors for different classes of signals. The Fourier transform of the time-domain signatures produces complex-valued spectra, containing both phase and amplitude information. The amplitude and phase at certain key frequency components constitute pairs of feature subsets on which the classification is based. An important advantage of this approach is the small dimensionality of feature vectors. This allows the features to be directly extracted from pulsed responses with relatively low computational complexity.

4.3. *Performance Assessment of Classification*

Cross-validation methods [40] and a leave one out (LOO) [41, 42] estimator of the de-convolved T-ray data set are utilized to provide a nearly unbiased estimate of the prediction error rate. The performance of classifying the RNA samples is evaluated using eight-fold cross-validation, while the powdered material classification is validated using LOO. The RNA dataset is divided into eight subsets of approximately equal size.

For CELMs, the real and imaginary parts of each subset are tested using the classifier trained on the remaining subsets consisting of both the real and imaginary parts. The real and imaginary parts of the complex valued labels associated in the input matrix are used for training the classifier to calculate the real and imaginary parts of the complex valued output weights, respectively. The results from the 8 runs (50 runs in the case of powder samples) for each class of RNA samples (powder samples), corresponding to real and imaginary parts, respectively, are averaged to provide a statistical estimate of the complex valued classifier performances. Therefore, the test elapsed time indicates the 8 runs (300 runs) required to perform classification as testing time. For real valued SVMs, in order to achieve maximum classification accuracy, we use both phase and amplitude as training and testing feature vectors. This approach serves as a useful comparison with the classification accuracy using CELMs. In order to compare the classifier performances between complex valued EML and real valued EML, we extent the same classifier design from CEML to real valued EML. In real-valued classification, only part of the complex valued inputs, (either the real or phase part), is used as an input to train the classifier.

4.3.1. *Binary classification of poly-A and poly-C datasets T-ray spectra*

Currently, the identification of the binding state of DNA is an emergent interdisciplinary research topic within the THz community, because it promises a label-free modality for the determination of the four base pairs; furthermore it has the potential to eliminate the polymerase chain reaction (PCR) step commonly associated with a sequencing process. In spite of the lack of characteristic absorption features in the T-ray region, it is possible to discriminate un-hybridized from hybridized DNA strands on the basis of observed loading (scattering parameters) of samples deposited on planar micro-fabricated T-ray resonators [28]. Furthermore, there have been suggestions that proteins can be detected by T-ray circular dichroism (TCD) spectroscopy, because many bio-molecules in crystalline form exhibit strong and specific features in their dielectric spectra [28] different from their phonon resonances. In the current case, we illustrate the resultant real valued classification performance using only the complex part (phase), as the feature vector, to train and test the classifier. This choice is made on the basis of the large observed differences of the spectral phase depicted in Fig. 7(b). This approach facilities classification. To tune the algorithm, we used small-and-separate validation sets drawn from the test subsets, with the remainder of the test subsets used for testing the classification performance. In the approach outlined above, each RNA pixel instance is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified.

4.3.2. *Multiclass classification of powder datasets T-ray spectra*

Similarly, LOO is used to evaluate each unknown feature vector, and then is used as a basis to evaluate classifier designs for powder classification [7, 41]. Therefore, LOO accuracy depicts also the percentage of correctly classified data sets. The reason why LOO is used instead of eight-fold cross-validation for the powder experiment is due to the relatively small number of measurements for the different powders. With such a restriction, LOO is preferred as the overall classification experiment is averaged over more runs. In order to evaluate the classification performance in relation to the two types of THz experiments presented in the paper, accuracy of classification is used as the quantity for assessing the performance of all the classification tasks. This is equal to the number of correct classified test vectors N_ι^{true} for all-class samples $\iota = 1, ..., m$ divided by the total number of vectors to be tested N^{total}.

$$\text{accuracy} = \frac{\sum_{\iota=1}^{m} N_\iota^{\text{true}}}{N^{\text{total}}} \tag{26}$$

5. Experimental Results

5.1. *The Fourier Spectrum Analysis*

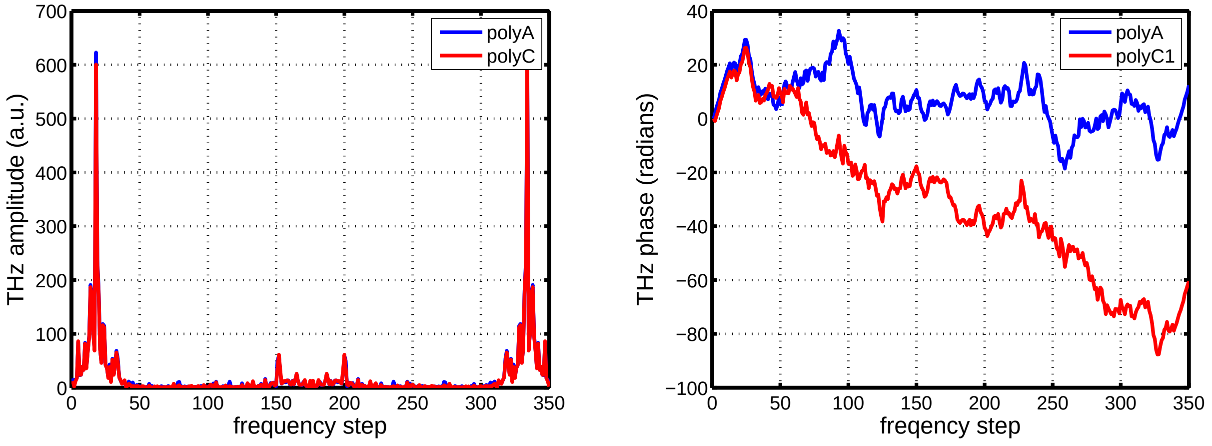
5.1.1. *The Fourier Spectrum Analysis for the Classification of Poly-A and Poly-C T-Ray Pulses*

In the RNA study, each pixel is composed of 350 time-domain points. Upon Fourier transformation, one generates 350 frequency bins spanning from DC to 8 THz. The 3000 pixel data set consists of pairs of background and sample data. The population of pixels belonging to the poly-A and poly-C classes is 48 for both classes; this number excludes background pixels. In order to obtain reduced dimension feature subsets and make them discriminable for the different class, as the amplitude and phase values of the pulse responses are first calculated, and then those values corresponding to the frequency with the greatest amplitude (i.e., strongest response) are used as the input features to the classifier. This process extracts a 2-D feature vector from the full spectral data with 350 non-redundant dimensions.

Fig. 4 displays the corresponding amplitude and phase spectra of poly-A and poly-C, which are obtained by fast Fourier transform (FFT) of the T-ray pulse responses. Linear extrapolation of the phase to DC is performed with phase de-branching.

Fig. 4 displays the corresponding amplitude and phase spectra of poly-A and poly-C, which are obtained by fast Fourier transform (FFT) of the T-ray pulse responses. Linear extrapolation of the phase to DC is performed with phase de-branching. The amplitude and phase values at 2 frequency bins are used as inputs to all the classifiers. For CELMs, the amplitude and phase are combined to form a complex valued input. For real valued EML, since the output needs to be associated to a real valued parameter, only part of the complex valued inputs are used in the training process of the classifier, (either the real or the imaginary part). In the current case, we prefer to use the imaginary part, phase, as the feature vector, to train and test the classifier. This is deemed acceptable as both the real and imaginary parts are related to each other through the Kramers-Kronig relation. We use phase as the input feature space, because it shows better separation than amplitude curves. For CELMs, two real Gaussian kernels are used for the final feature mapping from two nonlinear feature spaces to two linear ones, as two real ELMs tasks are conducted to realise complex valued learning. This differs from, real machine learning, i.e. ELMs and SVMs, where only one real Gaussian kernel is used for mapping. Accordingly, we apply the classifiers training algorithm to produce CELM- and SVM-associated learning vector patterns in a 2-D feature space (consisting of amplitude and phase). These are illustrated in Fig. 5(a) and (b), respectively, with the penalty parameter $C = 0.5$ and $\sigma = 1$ for CELMs, and the penalty parameter C of infinity and the width parameter of Gaussian kernel σ set equal to $1 \times e - 5$.

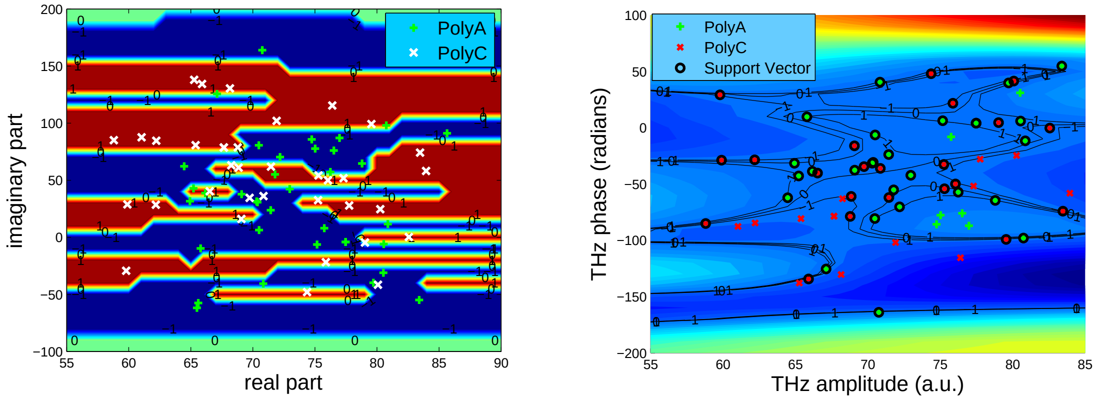
Fig. 5(a) and (b) depict 36 training vectors for illustration purposes. The background colour shows the shape of the decision surface. In Fig. 5(a), red regions represent the class belonging to the poly-C sample labelled by 1, and blue regions indicate the class related to poly-A sample labelled by -1. Contrary to real-valued machine learning, the labels of CELMs are complex valued. The numbered labels to be output are shown in Fig. 5(a) are calculated as the addition of doubled value of the real part $(\mathbf{R})$ and the value of imaginary part (J), in relation to the complex valued labels, with zero indicating non-classified data. That is, the num-



(a)

(b)

Fig. 4. Illustration of (a) amplitude and (b) phase of poly-A and poly-C T-ray RNA spectra as a function of frequency. In order to show an full phase variation throughout all frequency, we keep the whole frequency bins, though the amplitude plot has symmetry with center at 175th frequency bin.



(a)

(b)

Fig. 5. Illustration of binary classification for the recognition of RNA samples consisting of 36 training vectors for each. (a) Illustration of a CELM classification scheme, using two real Gaussian kernels to map the training vectors to a 2-D complex-valued feature space, with the penalty parameter $C = 0.5$ and $\sigma = 1$. (b) Illustration of a real SVM classification scheme, using a real Gaussian kernel to map the training vectors to a 2-D complex-valued feature space. The penalty parameter C is set to infinity and the width parameter of the Gaussian kernel σ is set to $1 \times e^{-5}$.

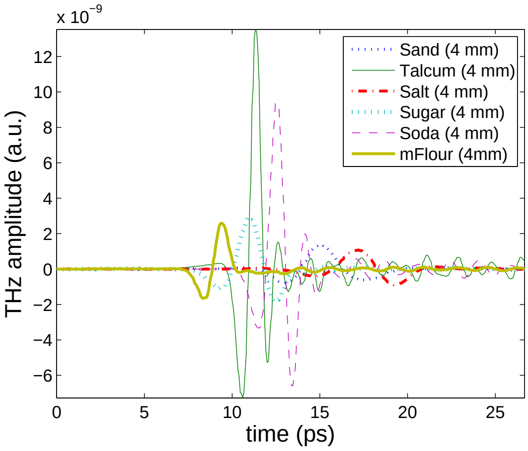


Fig. 6. Single pixel plots of complex valued learning vectors for the six powder samples measured via T-rays illustrating the linear decision function among each classes by applying induced real RKHS kernels to map the complex input data into a 2D complex-valued feature space. There are 49 pixels selected randomly from each of six classes of powder samples. The labels are complex valued and produce 12 classes.

bered labels ($\mathbf{Y}$) satisfies the equation: $\mathbf{Y} = 2 \times \mathbf{R} + \mathbf{J}(\Im)^2$ with $(\Im)^2 = -1$. Specifically, we set the classification label belonging to poly-A as $\mathbf{I} + \Im(\mathbf{I})$, and belonging to poly-C as $\textbf{-I} + \Im(\textbf{-I})$. The $\mathbf{I}$ indicates an identity matrix. Fig. 5(b) illustrates a SVMs classification scheme, where dark blue regions represent the class belonging to the poly-C sample labelled by 1, and light blue regions indicate the class related to poly-A sample labelled by -1. Separating hyper-planes for two classes are indicated by 0. The circles represent the calculated support vectors. Compared with the training vectors, the number of support vectors are reduced, which takes on an important role in achieving the ideal shape of hyper-planes and facilitating computation of the classification algorithm. In both cases, the machine learning for two-class samples — poly-A and poly-C denoted by white "+" and black "×" are approximately separated by their own boundary lines though there is a little overlapping. More detailed results on classification accuracy are described in the next section, where 200 random selections of training vectors are fed to the classifiers.

5.1.2. *The Fourier Spectrum Analysis for Multi-class Classification of THz spectra*

The images of powder samples consist of $6 \times 50 = 300$ pixels. For each pixel, the number of samples associated to a pulse time transient is 400. Fig. 6 shows the corresponding time domain signals for a single pixel taken from each of the powder sample image datasets.

Fig. 7 shows the phase and amplitude plots in the frequency domain for six different powder samples. Each curve is associated with a single pixel sampled from the image data. The spectrum has a cut-off frequency at 3 THz. Sharp changes of amplitude at the second frequency bin may be observed in Fig. 7(a). Good separations of curves of T-ray phase are illustrated in Fig. 7(b). We produce the learning vector patterns for multiclass recognition via CELM, shown in Fig. 8(a) after Fourier transformation of the time-domain signatures and extraction of the corresponding complex valued features associated with the second frequency bin. We use 49 input vectors related to each powder sample for training the classifier. Two real RKHS kernels are used for mapping. The optimal Gaussian parameter of σ is set to 100 and the penalty parameter C is set to 0.1. The labels are complex-valued and produce 12 output classes. Back ground colour shows the contour shape of the decision surface, (these are numbered from 2-12), these correspond to the amplitude calculations derived from the sum of real and imaginary values of the respective complex labels. It can be observed that THz measurements of powder samples regarding salt, sand, talcum, are grouped more tightly than the powder samples of flour, soda and sugar.

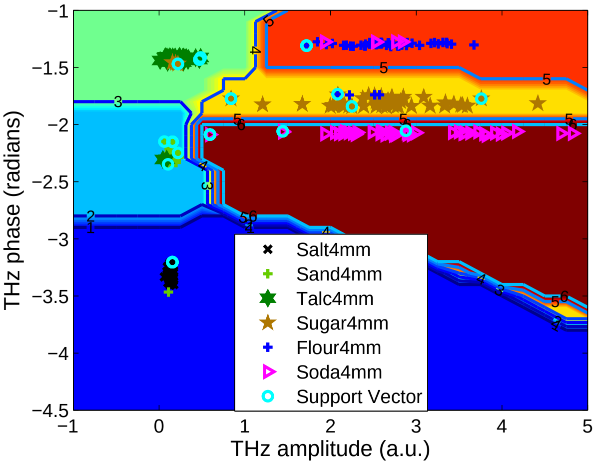
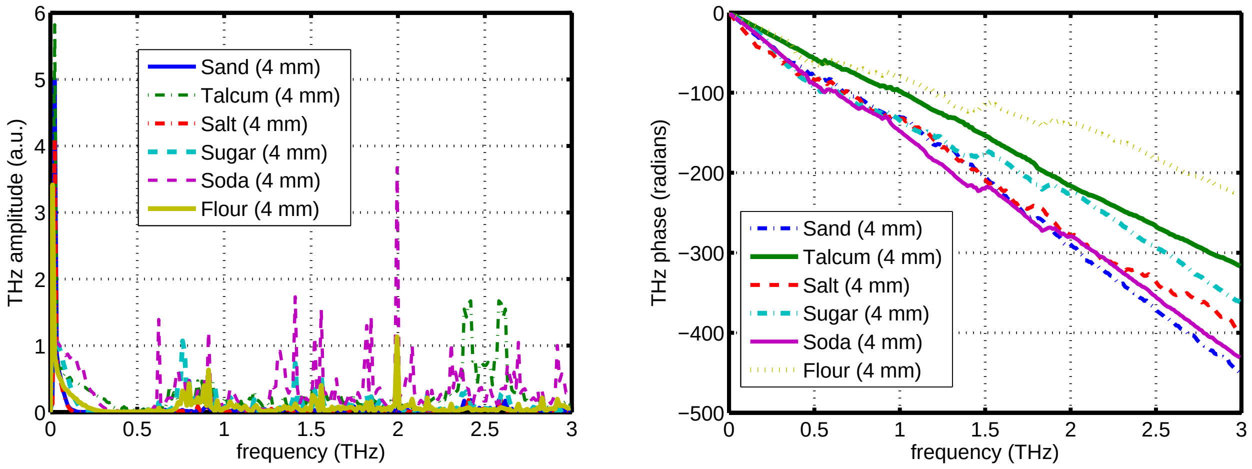


Fig. 9. Learning vectors for the powder data sets plotted to illustrate the linear decision function between the pairs of classes after applying a Gaussian kernel for mapping. There are 49 pixels selected randomly from each of the six powder samples. Background colour shows clearly the contour shape of the decision surface. The small yellow region on the bottom of the right hand side denotes undecided classification.

Only the labels consisting of the same real and imaginary parts (both parts label the same class) are validated for final power identification. The labelled contours that correspond to different real and imaginary parts (the real and imaginary parts label the different classes) are illustrated in Fig. 8(b). These regions are undecided in the classification process and are therefore excluded to avoid over-fitting problems.

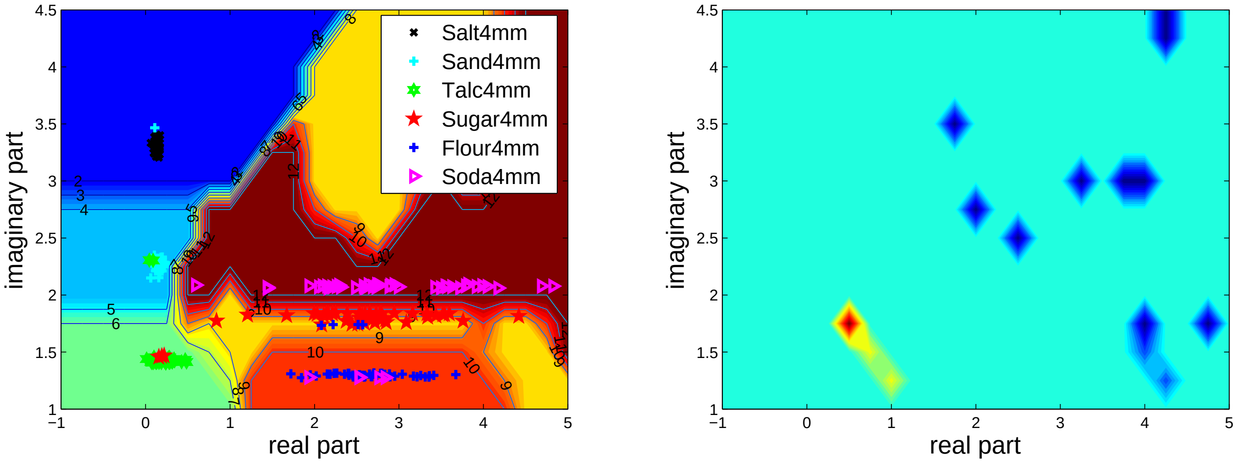
Fig. 9 illustrates the multi-class separation for the six types of powder substances using SVMs. SVMs are designed according to a pair wise-strategy. One real Gaussian kernel with $C = 1000$ and $\sigma = 1 \times e^{-7}$ is used to map the input data into a 2D Fourier feature space for visualisation purposes. The support vectors indicated by cyan circles are subsets of the training data sets and are used to construct a two-dimensional hyper-plane in feature space, which acts as a boundary separating each class of different powder materials.



(a)

(b)

Fig. 7. Illustration of Fourier spectrum. (a) shows the amplitude (attenuation) as a function of terahertz frequency, whereas (b) shows corresponding phase delay (equivalent to chromatic dispersion) as a function of terahertz frequency.



(a)

(b)

Fig. 8. Illustration of CELM multi-class classification scheme. (a) Complex valued learning vectors for the six samples plotted to illustrate the linear decision function among each classes by applying induced real RKHS kernels to map the complex input data into 2D complex valued feature space. There are 49 pixels selected randomly from each of the six powder samples. The labels are complex valued, generating 12 classes. (b) Illustration of the colour coded regions with non-zeros indicated by the colour bar. The colour regions with non-zero value indicate that the multi-class powder sample classification process remains undecided by CEML as the real and imaginary parts are not equal to each other.

5.2. *Resultant Classification Performance*

5.2.1. *CELM Classification performance of RNA sample Spectra*

For classification of RNA samples, two real Gaussian kernels are applied to generate complex valued RKHS. All the classification runs are performed in MATLAB version R2013a on a personal computer running Windows 7 with an Intel(R) Core(TM) i5-3470 CPU (3.20 GHz) and 8 GB of memory. Using CELMs, the average time spent classifying the two classes of RNA samples is 0.1293 seconds after 200 classification runs using 36 datasets for each class of RNA sample. To evaluate the effect of the Gaussian kernels for the RNA sample classification with complex valued feature, suitable values of C and σ are considered via a parametric search using separate validation sets. After training, the final classification accuracy is compared. In the training phase, the training vectors are randomly selected from a given proportion, (varying from 1/8 to 6/8), of the input population of 48 pixel responses from each RNA class. The highest classification performance was obtained for the penalty parameter $C = 1$ and $\sigma = 1$, with a classification accuracy of 72%.

It was found that the classification accuracy varies throughout the range of values of σ, which was varied from 0.1 to 100, in steps of 1 in log scale. The classification accuracy increased with an increased number of training vectors, according to our expectation. The classification accuracies are varied according to the penalty parameter values C. Fig. 10(a)—(c) show classification performance using our algorithm as a function of different sizes of testing data sets, related to $C = 0.1$, $C = 0.5$, and $C = 5$, respectively. The various value of σ are all plotted according to the different values of C for direct comparison. It is clear that the three figures show the different trends, which implies that the different value of penalty parameter C leads to different classification characteristics. In Fig. 10(b), the curve, related to σ of 10, leads to best performance, especially when the number of training vectors is in the range from 48 to 84, with classification performance over 70%, while in Fig. 10(c), the curve with σ of 100, shows second best performance of classification accuracy compared with the others. In general, the classification performances shown in Fig. 10(a) are inferior when compared with Fig. 10(b) and Fig. 10(c), but not by very much.

For comparison purposes, we also use SVMs to classify the object samples. Similar to the classification of RNA samples via CELM, real Gaussian kernels are applied with both phase and amplitude as feature vectors for training and testing of the SVM classifiers. The time spent classifying the RNA samples is 4.14 seconds for 200 runs of measurements when using 72 input data sets.

To evaluate the effect of the Gaussian kernel on the RNA classification problem via SVMs, the values of C are tuned. After training, the final error rate, the number of support vectors and the elapsed time are compared. The training phase follows the same procedure as CELMs. Considering the classification performance to be less sensitive on the choice of λ, we illustrate the tuning of C briefly. The highest classification performance was obtained for the penalty parameter $C = 1$ and $\lambda = 0.003$, with a classification accuracy of 72%. This is the same as the classification accuracy achieved by CELM, but the elapsed time is 30 times slower than using CELM for the two-class classification.

It was found that the classification accuracy using SVMs is similar to CELM across a range of values for C, from 0.001 to 10^4, when this was varied in steps of 1 on a log scale. The classification accuracy is improved with an increased number of training vectors, according to our expectations. Fig. 11 shows the classification performance using SVMs for different sizes of test data sets, respectively. The various values of C are all plotted for direct comparison. In this figure, the curve, related to a C value of 1, shows that best performance is achieved when the number of training vectors is in the range of 48 to 84.

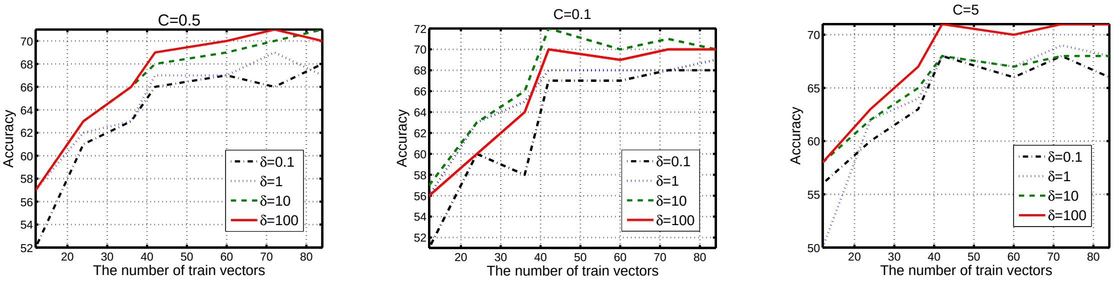
The number of the computed support vectors is roughly one-third fewer than the number of training vectors. A small number of SVs is desirable for implementation since their number directly relates to the computational complexity of the automatic classification task. Fig. 12 shows the variation of the number of SVs with the increased number of the input training feature sets for binary classification of RNA measurements. The number of the input training RNA samples is varied from 12 to 84, with a step size of 12. It can be seen that the corresponding number of SVs, the accuracy and the elapsed time increased monotonically and almost linearly with the number of training vectors.

5.2.2. *CELM Classification performance of multi-class powder sample spectra*

Tables 1–3 show the achieved multi-class classification accuracy on the THz Fourier spectral features after applying three types of machine learning algorithms: CELM, ELM and SVMs, respectively, as a function of varying penalty parameter C, optimal Gaussian kernel parameter σ and elapsed time. A leave-one-out (LOO) estimator is used for both training and testing purposes. For training, pixels from all the classes are presented to the three classifiers. The remaining 1 pixel from each class is used for testing. The classification experiments are repeated over 50 runs. Therefore, the test elapsed time indicates the 300 runs required to perform classification as testing time. All the powder sample classification runs were performed using MATLAB version R2013a on a personal computer running Windows 7 with an Intel(R) Core(TM) i5-3470 CPU (3.20 GHz) and 8 GB of memory.

We used real valued Gaussian kernels for all classifiers. To evaluate both real and complex valued ELMs, we varied the optimal parameter σ from 0.1 to 1000 and the penalty parameter C from 10 to 1000, in steps of 1 in log scale. For SVMs, we also set a similar optimal parameter σ, but did not include results from such setting because it does not lead to any meaningful classification. Similarly, the penalty parameter C of SVMs is varied from 0.1 to 100, in steps of 1 in log scale for analysis. Setting these parameter values allows comparable classification performance among different classifiers.

The classification performance for CELM is listed in Table. 1, where both real and imaginary parts are used as input features for training and testing for the classifier. The known real and imaginary parts of complex valued labels associated matrix are applied further to calculate the real and imaginary parts of output weights, respectively, for training purposes. For real valued machine learning, i.e. SVMs, we use both phase and amplitude as training and testing feature vectors, to find the maximum possible classification accuracy, with only real valued labels associated matrix as an input label matrix. This way, one can make a useful comparison with the classification accuracy using CELMs, as these use both real and imaginary portions of the input complex valued labels each time for training, separately. The classification performance for CELM is listed in Table. 2. The elapsed time of the CELM classification scheme is around $3 \times e^{-1}$ second, which is nearly hundred times faster than in real SVM classification (which requires



(a) (b) (c)

Fig. 10. Illustration of classification performance using our algorithm versus different sizes of testing data sets, with the range of values for σ, from 0.1 to 100, in steps of 1 in log scale, related to $C = 0.1$ in (a), $C = 0.5$ in (b), and $C = 5$ in (c), respectively.

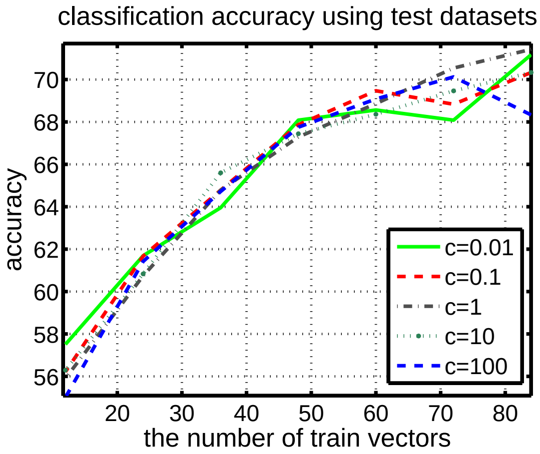


Fig. 11. Illustration of the validation of classification accuracy, via the plot of classification performance versus the number of input validation vectors, corresponding to the different value of parameter C.

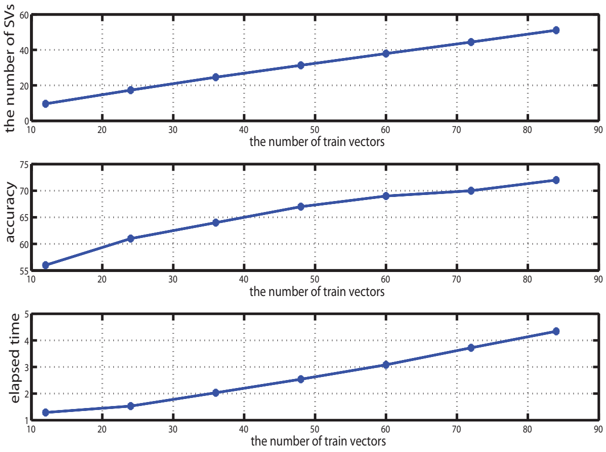


Fig. 12. Illustration of the variation of the number of SVs, classification accuracy and elapsed time with the increment of the number of the input training feature sets in the classification of RNA data, with $C = 1$ and $\sigma = 0.003$.

an elapsed time of over 30 seconds). Furthermore, the total classification accuracy of CELM is increased in tandem with the penalty parameter C. This can be observed by finding the classification accuracy with the same value of the optimal parameter σ. It is worth noting that the total classification accuracy of CELM, however, is reduced for an increased value of the optimal parameter σ. This means, that a bigger σ, value results in a lower classification accuracy for CELM. Among these T-ray measurements, powder samples of salt and talcum are easiest to be separated, with classification accuracy of 100% under all the cases, whereas the powder sample of sugar is more difficult to identify.

To evaluate the classification performance of SVMs, in addition to the classification accuracy and elapsed time, we list the number of support vectors (SVs) used to calculate the boundaries of each powder class, these are illustrated in Table. 2. In contrast to CELMs, the classification accuracy of SVMs is increased as a result of increasing the value of parameter C, only when $\sigma = 0.1$. In this case, the total classification accuracy is increased rapidly from 55.44% at $C = 0.1$ to 99.66% at $C = 10$ and $C = 100$ respectively, with the associated number of support vectors needed reduced from 993 to 516. The elapsed time is reduced dramatically from over 24 seconds to around 11 seconds, due to the reduced number of SVs. For the case where σ is changed from 10 to 1000, the classification accuracy does not significantly change as a function of parameters C and σ. The associated classification accuracy is around 87%–88%. The number of SVs is 1500, which results in much longer elapsed time (38 seconds), due to the increased computation load from the increased number of support vectors. Among these T-ray measurements, powder samples of salt are easiest to be separated with classification accuracy of 100% under all the cases, whereas the powder sample of sugar and soda are more difficult to identify.

Similarly to the binary classification example, CELM performance is compared to ELM performance also for the multi-class case. The classification performance achieved is listed in Table. 3. An improved classification accuracy is observed with the increased value of the C variable. In

Table 1
Classification accuracy (%) and elapsed time are illustrated for powder classification using CELM with varying penalty parameter C and optimal Gaussian kernel parameter σ. The Fourier spectral features are extracted as real and imaginary parts.

C	σ	each class name and classification accuracy						total classification accuracy	elapsed time (s)
		salt	sand	talcum	sugar	flour	soda		
10	0.1	100	100	100	93.88	100	100	98.98	0.7236
	1	100	100	100	95.92	100	100	99.32	0.8651
	10	100	100	100	0	100	100	83.33	0.6716
	100	100	100	100	63.27	100	100	89.46	0.6935
	1000	100	0	100	0	83.67	93.88	48.30	0.6661
100	0.1	100	100	100	95.92	100	100	99.32	0.6901
	1	100	100	100	95.92	100	100	99.32	0.6786
	10	100	100	100	0.6327	100	100	89.46	0.6734
	100	100	100	100	0	100	93.88	82.31	0.6669
	10000	100	0	100	0	65.31	93.88	59.86	0.6756
1000	0.1	100	100	100	95.92	100	100	99.32	0.6880
	1	100	100	100	95.92	100	100	99.32	0.6758
	10	100	100	100	97.96	100	100	99.66	0.6877
	100	100	100	100	0	100	93.88	82.31	0.8549
	1000	100	0	100	0	97.96	83.67	63.61	0.6817

Table 2
Classification accuracy (%) and elapsed time are illustrated for powder classification using SVMs with varying penalty parameter C and optimal Gaussian kernel parameter σ. The Fourier spectral features are extracted as input features for classification.

C	σ	each class name and classification accuracy						total classification accuracy	♯ of SVs	elapsed time (s)
		salt	sand	talcum	sugar	flour	soda			
0.1	0.1	100	20.41	79.59	24.49	75.51	32.65	55.44	993	24.5390
	10	100	100	100	85.71	100	40.82	87.76	1500	35.7086
	100	100	100	100	91.84	100	36.73	88.10	1500	33.6962
	1000	100	100	100	95.92	100	26.53	87.07	1500	34.8194
1	0.1	100	100	100	91.84	100	53.06	90.82	546	14.9137
	10	100	100	100	81.63	100	40.82	87.07	1500	33.2700
	100	100	100	100	91.84	100	36.73	88.10	1500	38.7038
	1000	100	100	100	95.92	100	26.53	87.07	1500	33.6806
10	0.1	100	100	100	97.96	100	100	99.66	516	10.9045
	10	100	100	100	81.63	100	40.82	87.07	1500	38.1266
	100	100	100	100	91.84	100	36.73	88.10	1500	38.0018
	1000	100	100	100	95.92	100	26.53	87.07	1500	38.0486
100	0.1	100	100	100	97.96	100	100	99.66	516	11.2789
	10	100	100	100	81.63	100	40.82	87.07	1500	33.3998
	100	100	100	100	91.84	100	36.73	88.10	1500	33.1502
	1000	100	100	100	95.92	100	26.53	87.07	1500	37.6274

addition, a reduction in classification accuracy is observed when the value of the σ variable is increased. The maximum classification accuracy is found to be 99.32%. The elapsed time is half of the time that required when using CELM. However, when compared with CELMs, the total classification accuracy of real ELM is reduced rapidly when the value of the σ variable is decreased. In contrast to CELM, among these T-ray measurements, powder samples of flour and soda are easiest to be separated for real ELM, with classification accuracy of 100% in all the cases, whereas the

Table 3
Classification accuracy (%) and elapsed time for powder classification using real ELM with varying penalty parameter C and optimal Gaussian kernel parameter σ. The same validation procedure to that used in CEML validation is followed. One of the Fourier spectral phase features, is extracted as training and testing to generate feature vectors for all classes of powder datasets.

C	σ	each class name and classification accuracy						total classification accuracy	elapsed time (s)
		salt	sand	talcum	sugar	flour	soda		
10	0.1	100	95.92	100	100	100	100	99.32	0.3580
	1	100	44.90	73.47	100	100	100	86.39	0.4136
	10	100	0	0	0	100	100	50	0.3500
	100	0	0	0	100	100	100	33.33	0.3488
	1000	0	0	0	0	100	100	33.33	0.3537
100	0.1	100	95.92	100	100	100	100	99.32	0.4902
	1	100	0.6122	100	100	100	100	93.54	0.3400
	10	100	0	0	14.29	100	100	52.38	0.3475
	100	100	0	0	0	100	100	50	0.3461
	1000	0	0	0	0	100	100	33.33	0.3481
1000	0.1	100	95.92	100	100	100	100	99.32	0.3502
	1	100	65.31	100	100	100	100	94.22	0.3459
	10	100	0	0	100	100	100	66.67	0.7800
	100	100	0	0	0	100	100	50	0.3513
	1000	0	0	0	0	100	100	33.33	0.3559

sand powder sample is more difficult to identify.

6. Conclusions

The widening proliferation of THz transient imaging as well as tomographic systems in the biomedical and pharmaceutical industries as well as the security sector have led to the generation of very large datasets requiring novel methods for comparing and classifying the acquired spectra. Typical THz images can be composed of more than 512x512 pixels each, and to each pixel output corresponds a time-domain THz transient signal usually recorded over a span of several picoseconds, oversampled at 200 to 3000 instances. Furthermore, the resolution of recording these time domain transients is usually 12 bits. Spectral resolution is a direct consequence of the duration of the time domain signal recorded as well as the sampling rate attained, with the finer resolution achieved when a larger number of samples is recorded. This leads to management problems that can only be addressed within a very large database context. This paper describes a CELM-based classifier based on selected spectral features of these Fourier transformed T-ray pulsed signals. The work assesses the suitability of CELMs in performing binary classification tasks of RNA samples, as well as its applicability in problems requiring a multi-class separation, the later problem being explored using datasets from six powder samples of different composition. The work also contrasts this performance to that using SVMs as well as EMLs. Since the observed spectral phase variations of the recorded signals between samples are generally larger than their amplitude counterparts they are more appropriate to be used as inputs to the classifier. CELM has an advantage over SVM classifiers in that it generates an output much faster without significantly sacrificing classification accuracy. The advantage of using only selected spectral features as inputs to the classifier is that the number of feature vectors required is minimized, this has computational advantages and avoids over-fitting problems. This work, therefore, establishes CELM with Gaussian kernels as a viable alternative to EML and SVM for this type of data sets.

7. Acknowledgments

The authors would like to thank B. Ferguson of the ARC National T-ray Facility, University of Adelaide, and B. Fischer, University of Freiburg, Germany, for providing the data.

References

[1] R. I. Hunter, P. A.-S. Cruickshank, D. R. Bolton, P. C. Riedi, and G. M. Smith, “High power pulsed dynamic nuclear polarisation at 94 GHz,” *Physical Chemistry Chemical Physics* **12**(22), pp. 5752–5756, 2010.

[2] G. W. Reginsson, R. I. Hunter, P. A.-S. Cruickshank, D. R. Bolton, S. T. Sigurdsson, G. M. Smith, and O. Schiemann, “W-band PELDOR with 1 kW microwave power: molecular geometry, flexibility and ex-

change coupling," *Journal of Magnetic Resonance* **216**, pp. 175–182, 2012.

[3] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light, Seventh Edition*, Cambridge University Press, Cambridge, U.K., 2002.

[4] S. Hadjiloucas, G. C. Walker, J. W. Bowen, and A. Zafiropoulos, "Propagation of errors from a null balance terahertz reflectometer to a sample's relative water content," *Journal of Physics: Conference Series* **178**(1), p. Art. No. 012012, 2009.

[5] X. Yin, B. Ng, B. Fischer, B. Ferguson, and D. Abbott, "Support vector machine applications in terahertz pulsed signals feature sets," *IEEE Sensors Journal* **7**(12), pp. 1597–1608, 2007.

[6] R. Galvão, S. Hadjiloucas, A. Zafiropoulos, G. Walker, J. Bowen, and R. Dudley, "Optimization of apodization functions in THz transient spectrometry," *Optics Letters* **32**(20), pp. 3008–3010, 2007.

[7] X.-X. Yin, B. W.-H. Ng, B. Fischer, B. Ferguson, and D. Abbott, "Application of autoregressive models and wavelet sub-bands for classifying terahertz pulse measurements," *Journal of Biological Systems* **15**(4), pp. 551–571, 2007.

[8] S. Hadjiloucas, R. Galvão, J. Bowen, R. Martini, M. Brucherseifer, H. Pellemans, P. H. Bolivar, H. Kurz, J. Digby, G. Parkhurst, and J. Chamberlain, "Measurement of propagation constant in waveguides using wideband coherent THz spectroscopy," *Journal of the Optical Society of America B: optical physics* **20**(2), pp. 391–401, 2003.

[9] R. Galvão, S. Hadjiloucas, V. Becerra, and J. Bowen, "Subspace system identification framework for the analysis of multimoded propagation of THz-transient signals," *Measurement Science and Technolgy* **16**(5), pp. 1037–1053, 2005.

[10] S. Hadjiloucas, M. S. Chahal, and J. W. Bowen, "Preliminary results on the non-thermal effects of 200-350 *ghz* radiation on the growth rate of *s.* cerevisiae cells in micro-colonies," *Physics in Medicine and Biology* **47**(21), pp. 3831–3839, 2002.

[11] R. Galvão, S. Hadjiloucas, and J. Bowen, "Use of the statistical properties of the wavelet transform coefficients for the optimisation of integration time in fourier transform spectrometry," *Optics Letters* **27**(8), pp. 643–645, 2002.

[12] R. Galvão, S. Hadjiloucas, J. Bowen, and C. Coelho, "Optimal discrimination and classification of THz spectra in the wavelet domain," *Optics Express* **11**(12), pp. 1462–1473, 2003.

[13] X. X. Yin, K. M. Kong, J. W. Lim, B. W.-H. Ng, B. Ferguson, S. P. Mickan, and D. Abbott, "Enhanced t-ray signal classification using wavelet preprocessing," *Medical and Biological Engineering and Computing* **45**(6), pp. 611–616, 2007.

[14] K. Wang and D. M. Mittleman, "Metal wires for terahertz wave guiding," *Nature* **432**, pp. 376–379, 2004.

[15] G. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **42**(2), pp. 513–529, 2011.

[16] H. Fröhlich, "The biological effects of microwaves and related questions," *Advanced Electronics and Electron Physics* **53**, pp. 85–152, 1980.

[17] W. Grundler and F. Kaiser, "Experimental evidence for coherent excitations correlated with cell growth nanobiology," *Nanobiology* **1**, pp. 163–176, 1992.

[18] D. M. Mittleman, R. H. Jacobsen, and M. C. Nuss, "T-ray imaging," *IEEE Journal of Selected Topics in Quantum Electronics* **2**(3), pp. 679–692, 1996.

[19] G.-B. Huang, D. Wang, and Y. Lan, "Extreme learning machines: a survey," *International Journal of Machine Learning and Cybernetics* **2**(2), pp. 107–122, 2011.

[20] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery* **2**, pp. 121–167, 1998.

[21] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, USA, 1995.

[22] X. Yin, B. Ng, and D. Abbott, *Terahertz imaging for biomedical applications: pattern recognition and tomographic reconstruction*, Springer-Verlag, New York, USA, 2012.

[23] A. El-Gindy and G. Hadad, "Nonparametric bayes error estimation using unclassified samples," *Journal of AOAC International* **95**(3), pp. 609–623, 2012.

[24] M. E. Van-Valkenburg, "In memoriam: Hendrik W. Bode (1905-1982)," *IEEE Transactions on Automatic Control* **AC-29**(3), pp. 193–194, 1984.

[25] P. Bouboulis, K. Slavakis, and S. Theodoridis, "Adaptive learning in complex reproducing kernel hilbert spaces employing wirtinger's subgradients," *IEEE Transactions on Neural Networks and Learning Systems* **2**(99), pp. 260–276, 2012.

[26] F. A. Tobar, A. Kuh, and D. P. Mandic, "A novel augmented complex valued kernel LMS," in *the 7th IEEE Sensor Array and Multichannel Workshop 2012*, pp. 473–476, 2012.

[27] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **42**(2), pp. 513–529, 2012.

[28] B. Fischer, M. Hoffmann, H. Helm, R. Wilk, F. Rutz, T. Kleine-Ostmann, M. Koch, and P. Jepsen, "Terahertz time-domain spectroscopy and imaging of artificial RNA," *Optics Express* **13**(14), pp. 5205–5215, 2005.

[29] M. Brucherseifer, M. Nagel, P. Haring-Bolívar, H. Kurz, A. Bosserhoff, and R. Büttner, "Propagation of errors from a null balance terahertz reflectometer to a sample's relative water content," *Applied Physics Letters* **77**(24), pp. 4049–4051, 2000.

[30] M. Nagel, P. Haring-Bolívar, M. Brucherseifer,

H. Kurz, A. Bosserhoff, and R. Büttner, “Integrated planar terahertz resonators for femtomolar sensitivity label-free detection of DNA hybridization,” *Applied Optics* **41**(10), pp. 2074–2078, 2002.

[31] M. Nagel, F. Richter, P. Haring-Bolívar, and H. Kurzr, “A functionalized THz sensor for marker-free DNA analysis,” *Physics in Medicine and Biology* **48**(22), pp. 3625–3636, 2003.

[32] C. Strachan, P. Taday, D. Newnham, K. Gordon, J. Zeitler, M. Pepper, and T. Rades, “Using terahertz pulsed spectroscopy to quantify pharmaceutical polymorphism and crystallinity,” *Optics Express* **94**(4), pp. 837–846, 2005.

[33] J. Zeitler, P. Taday, D. Newnham, M. Pepper, K. Gordon, and T. Rades, “Terahertz pulsed spectroscopy and imaging in the pharmaceutical setting - a review,” *Journal of Pharmacy and Pharmacology* **59**(2), pp. 209–223, 2007.

[34] Y. Watanabe, K. Kawase, T. Ikari, H. Ito, Y. Ishikawa, and H. Minamide, “Spatial pattern separation of chemicals and frequency-independent components by terahertz spectroscopic imaging,” *Applied Optics* **42**(28), pp. 5744–5748, 2003.

[35] J. Federici, B. Schulkin, F. Huang, D. Gary, R. Barat, F. Oliveira, and D. Zimdars, “Thz imaging and sensing for security applications-explosives, weapons and drugs,” *Semiconductor Science and Technology* **20**(7), pp. S266–S280, 2005.

[36] K. Kawase, Y. Ogawa, Y. Watanabe, and H. Inoue, “Non-destructive terahertz imaging of illicit drugs using spectral fingerprints,” *Optics Express* **11**(20), pp. 2549–2554, 2003.

[37] B. Ferguson, S. Wang, H. Zhong, D. Abbott, and X.-C. Zhang, “Powder retection with t-ray imaging,” in *Proceeding of SPIE Terahertz for Military and Security Application, 2003*, **5070**, pp. 7–16, 2003.

[38] B. Ferguson and D. Abbott, “Denoising techniques for terahertz responses of biological samples,” *Microelectronics Journal* **32**(12), pp. 943–953, 2001.

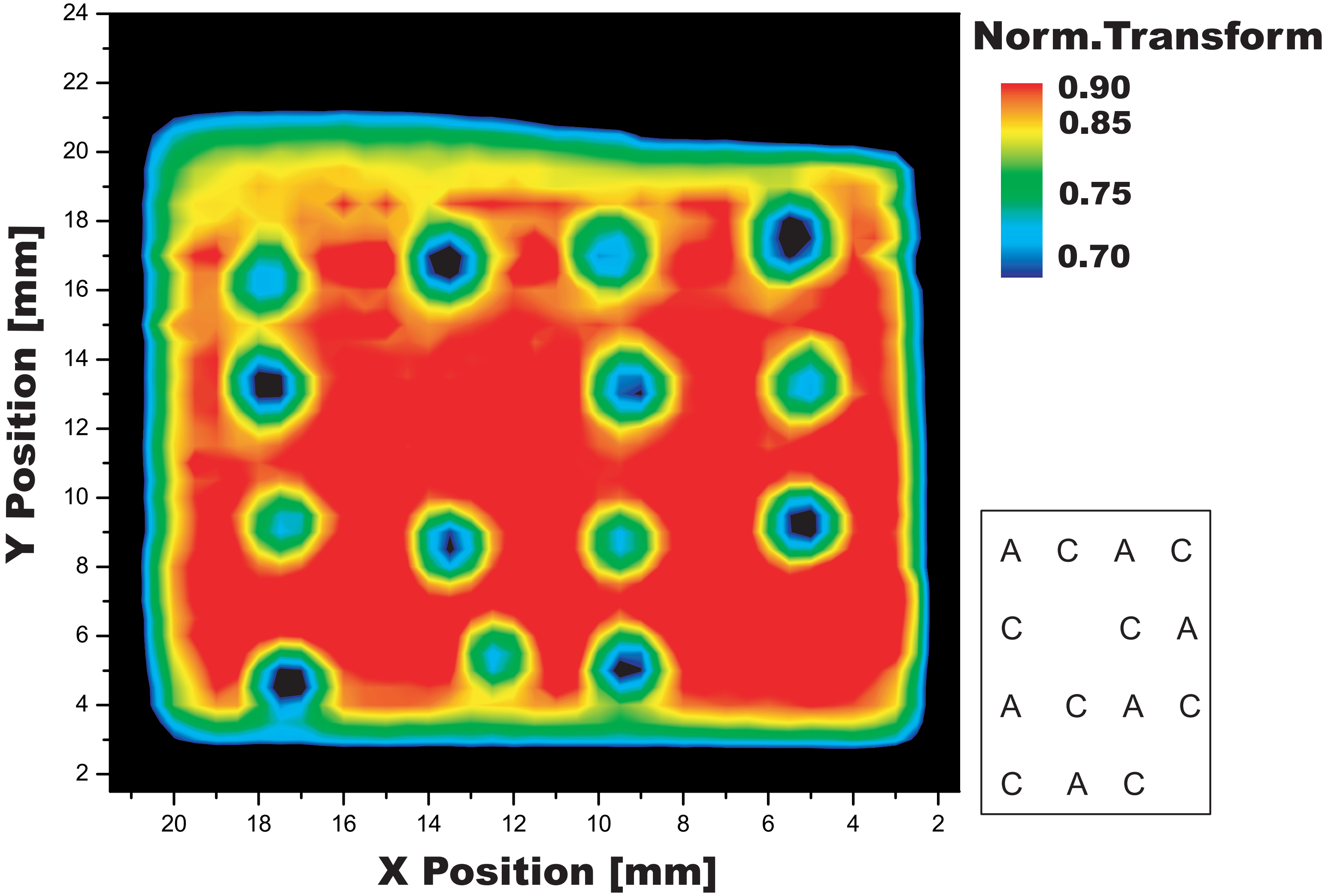
[39] L. Duvillaret, F. Garet, and L. Coutaz, “Areliable method for extraction of material parameters in terahertz time-domain spectroscopy,” *IEEE Journal on Selected Topics in Quantum Electronics* **2**(3), pp. 739–746, 1996.

[40] Y. Bengio and Y. Grandvalet, “No unbiased estimator of the variance of k-fold cross-validation,” *Journal of Machine Learning Research* **5**, pp. 1089–1105, 2003.

[41] K. Fukunaga and D. M. Hummels, “Leave-one-out procedures for nonparametric error estimates,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**(4), pp. 421–423, 1989.

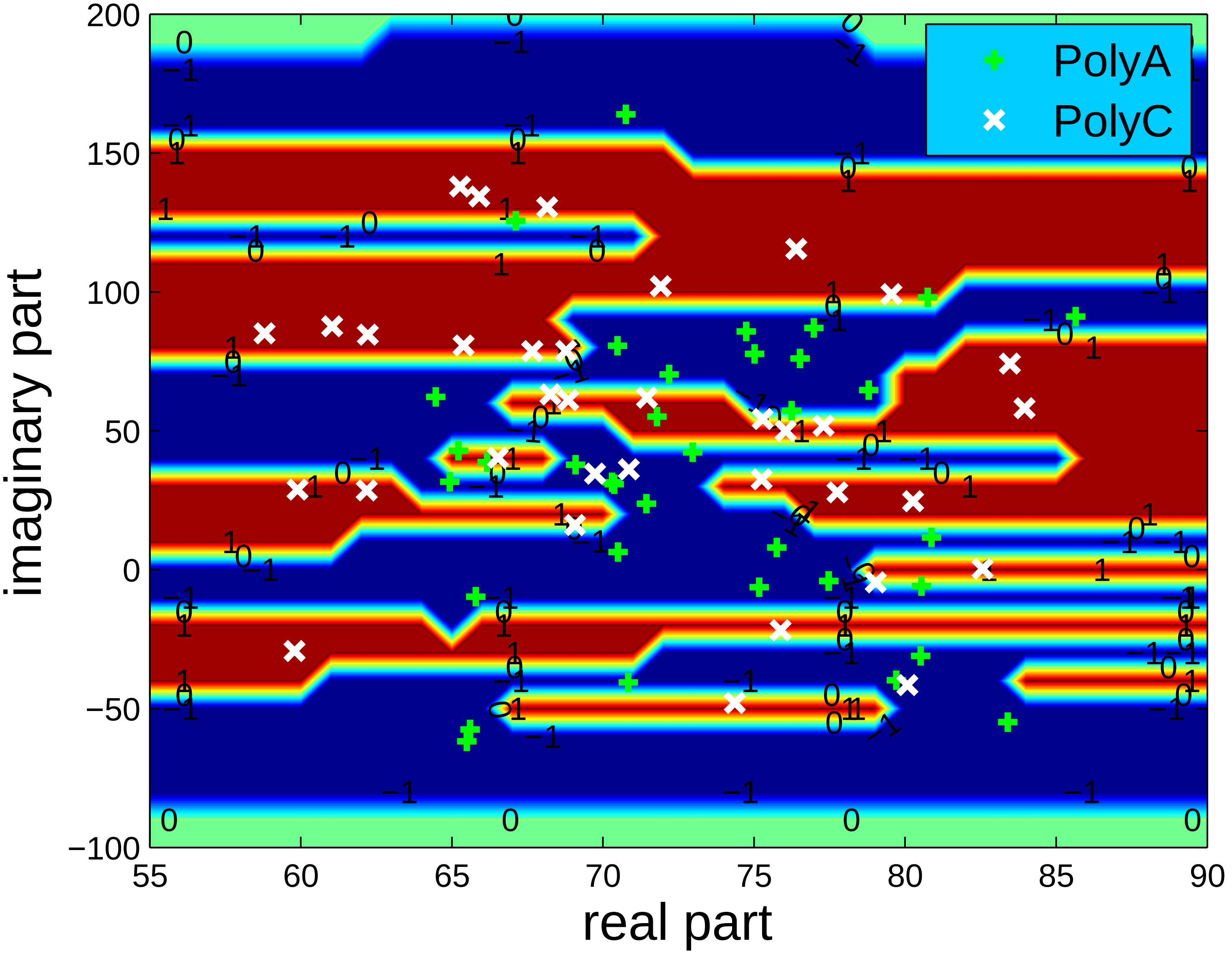
[42] K. Fukunaga and D. L. Kessell, “Chemometrics in pharmaceutical analysis: an introduction, review, and future perspectives,” *IEEE Transactions on Information Theory* **IT-19**(4), pp. 434–440, 1973.

biochip11.eps



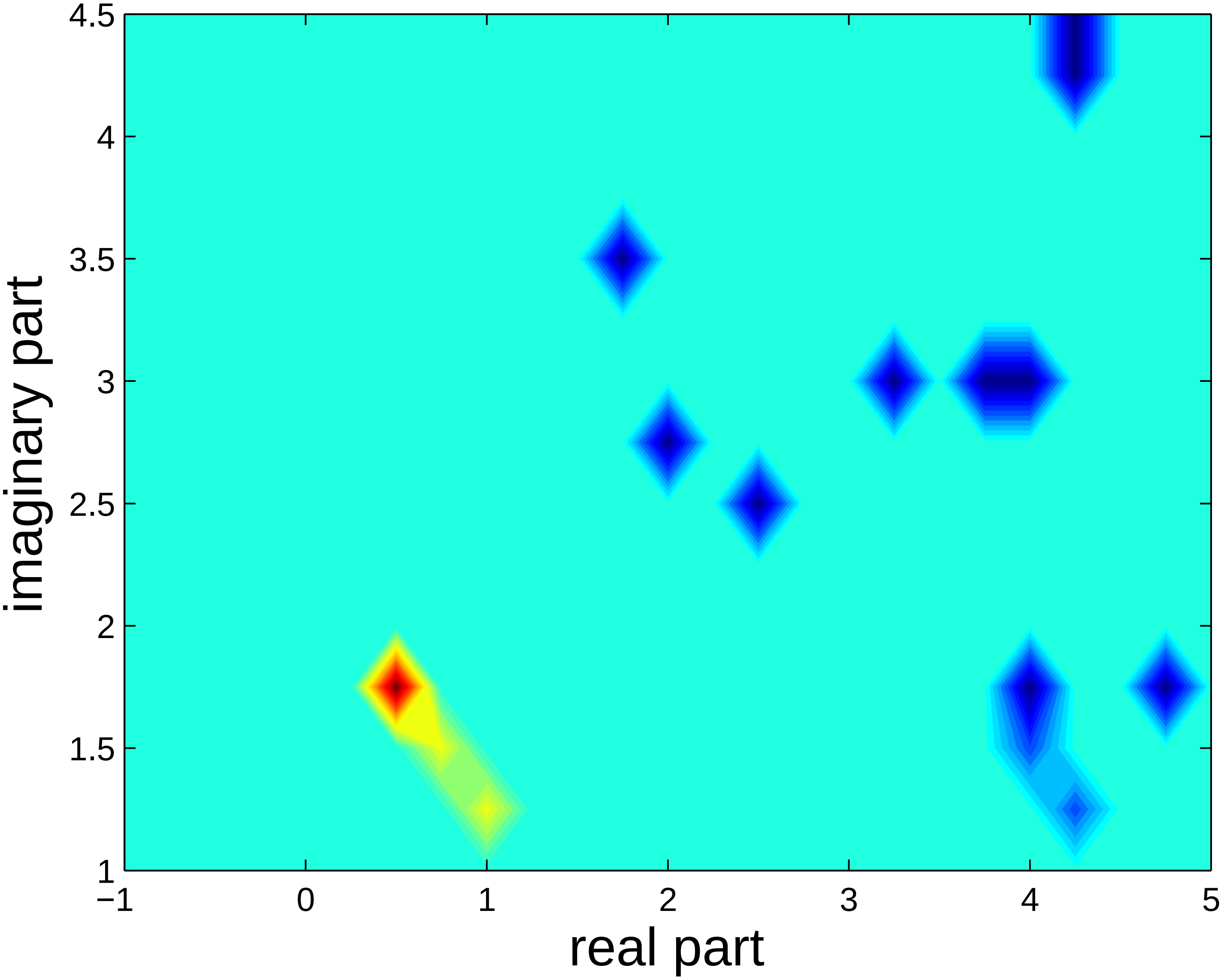
A	C	A	C
C		C	A
A	C	A	C
C	A	C	

celm_rna.eps

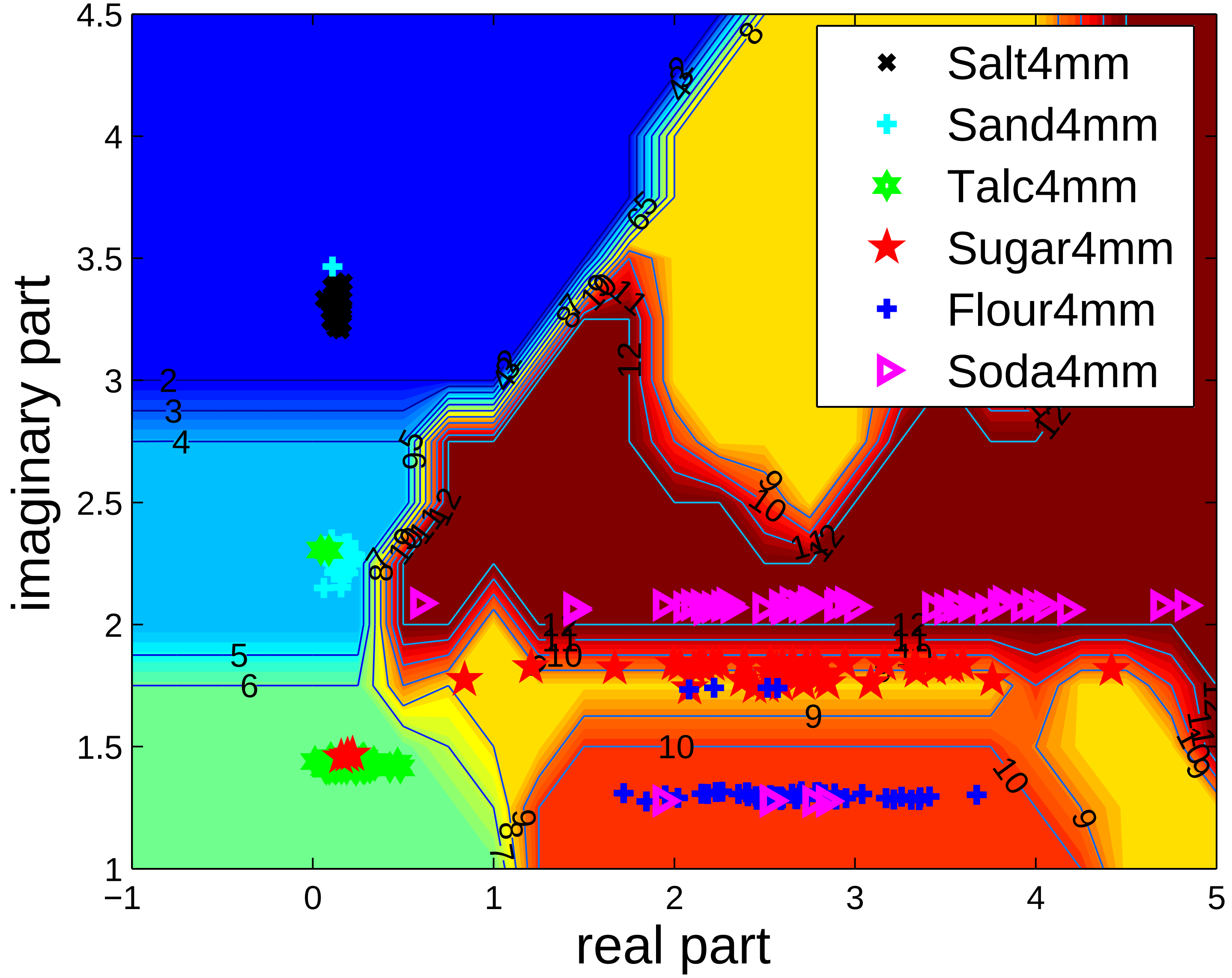


PolyA
PolyC
imaginary part
real part
200
150
100
50
0
−50
−100
55
60
65
70
75
80
85
90

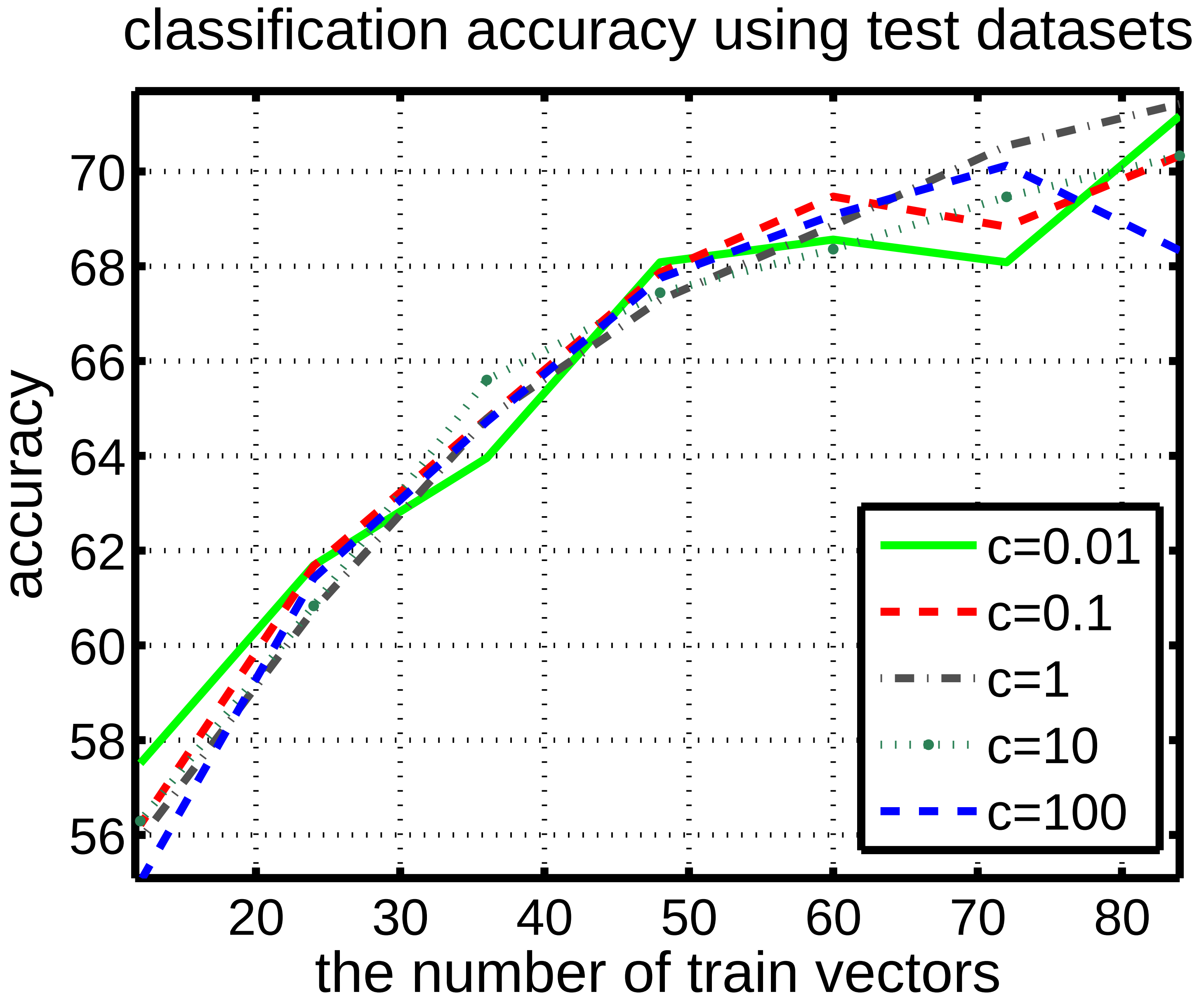
ceml1-nonclass.eps



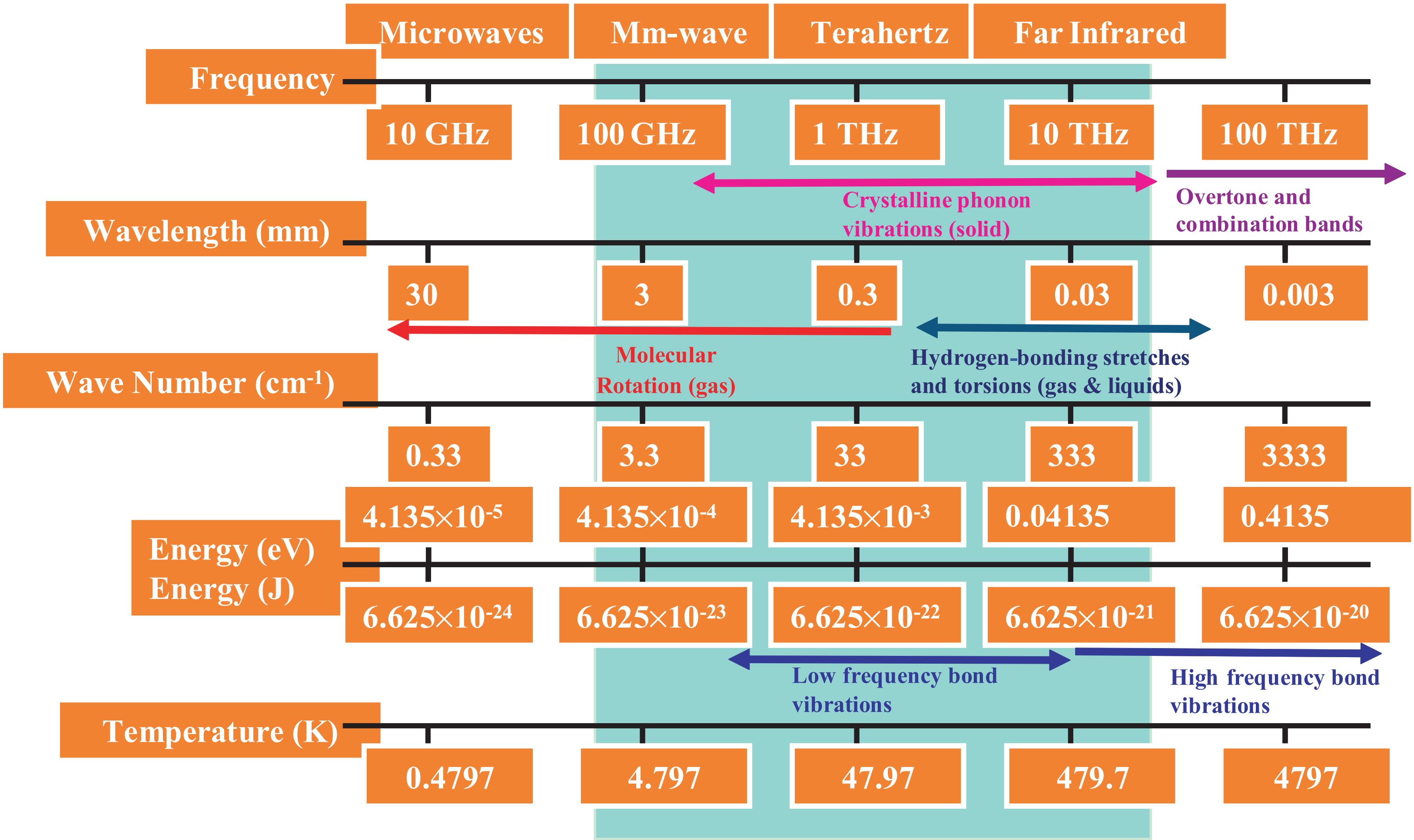
ceml1.eps


Salt4mm
Sand4mm
Talc4mm
Sugar4mm
Flour4mm
Soda4mm
imaginary part
real part
4.5
4
3.5
3
2.5
2
1.5
1
−1
0
1
2
3
4
5
2
3
4
5
6
7
8
9
10
11
12

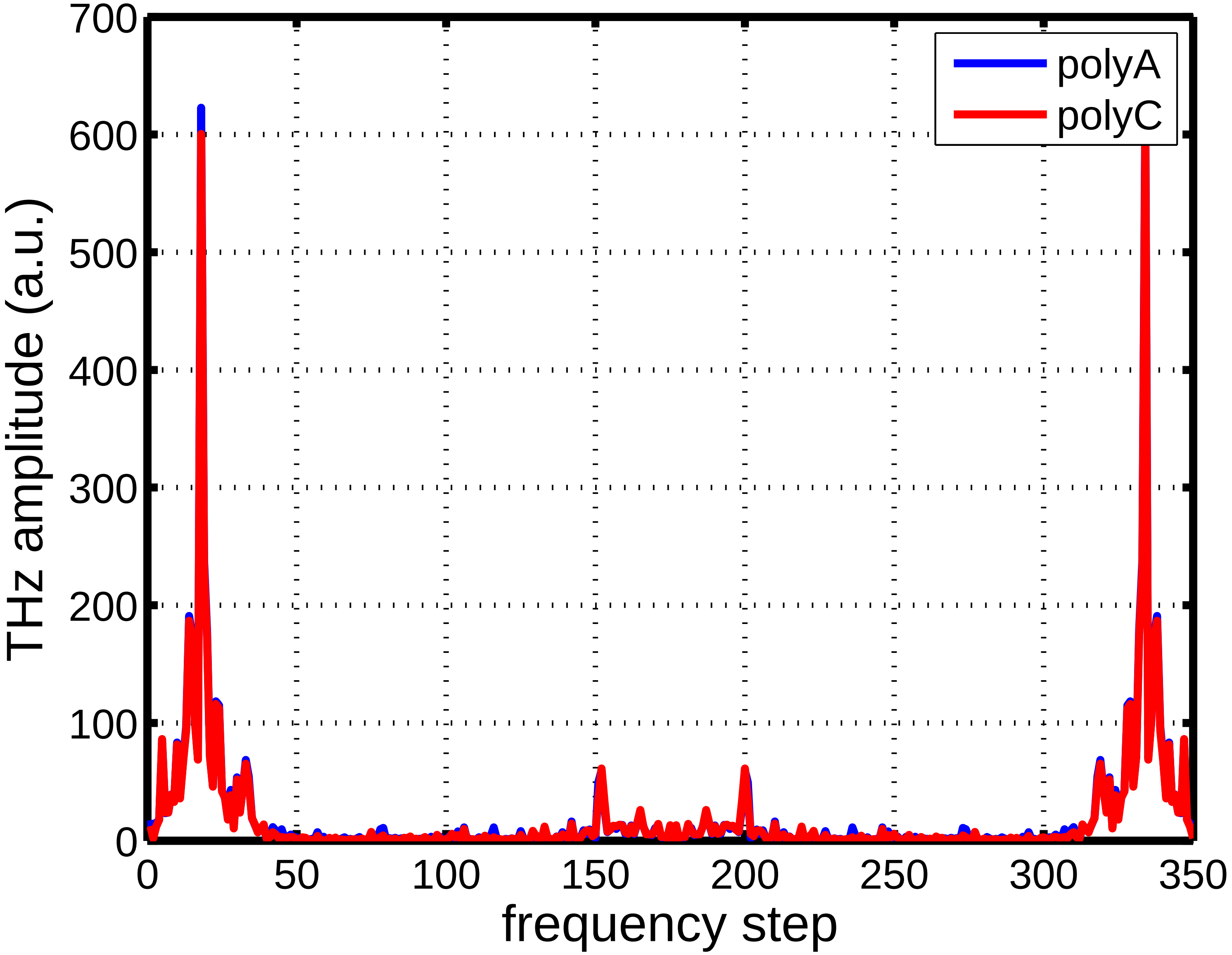
c_accuracy_test.eps



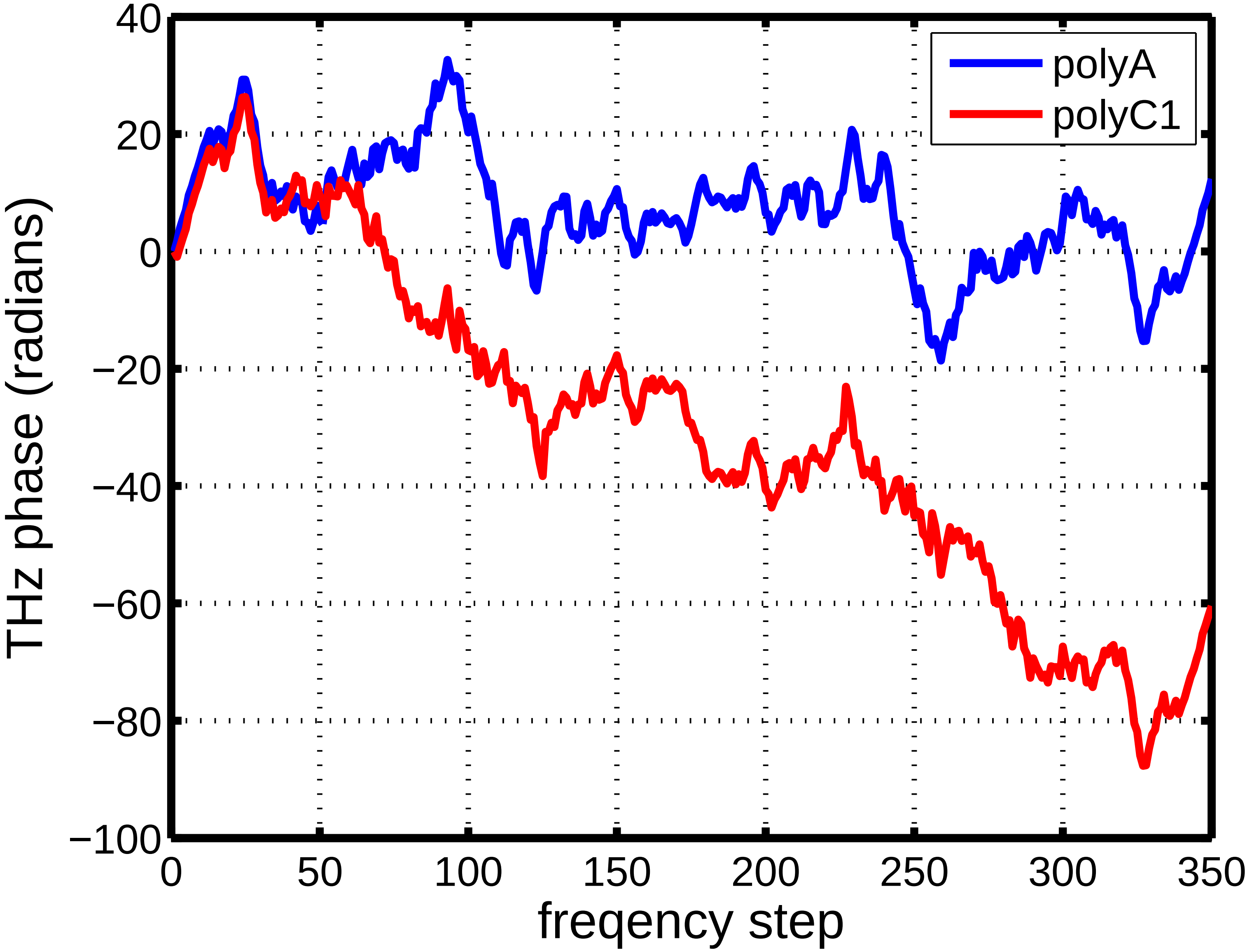
electrospectrum.eps



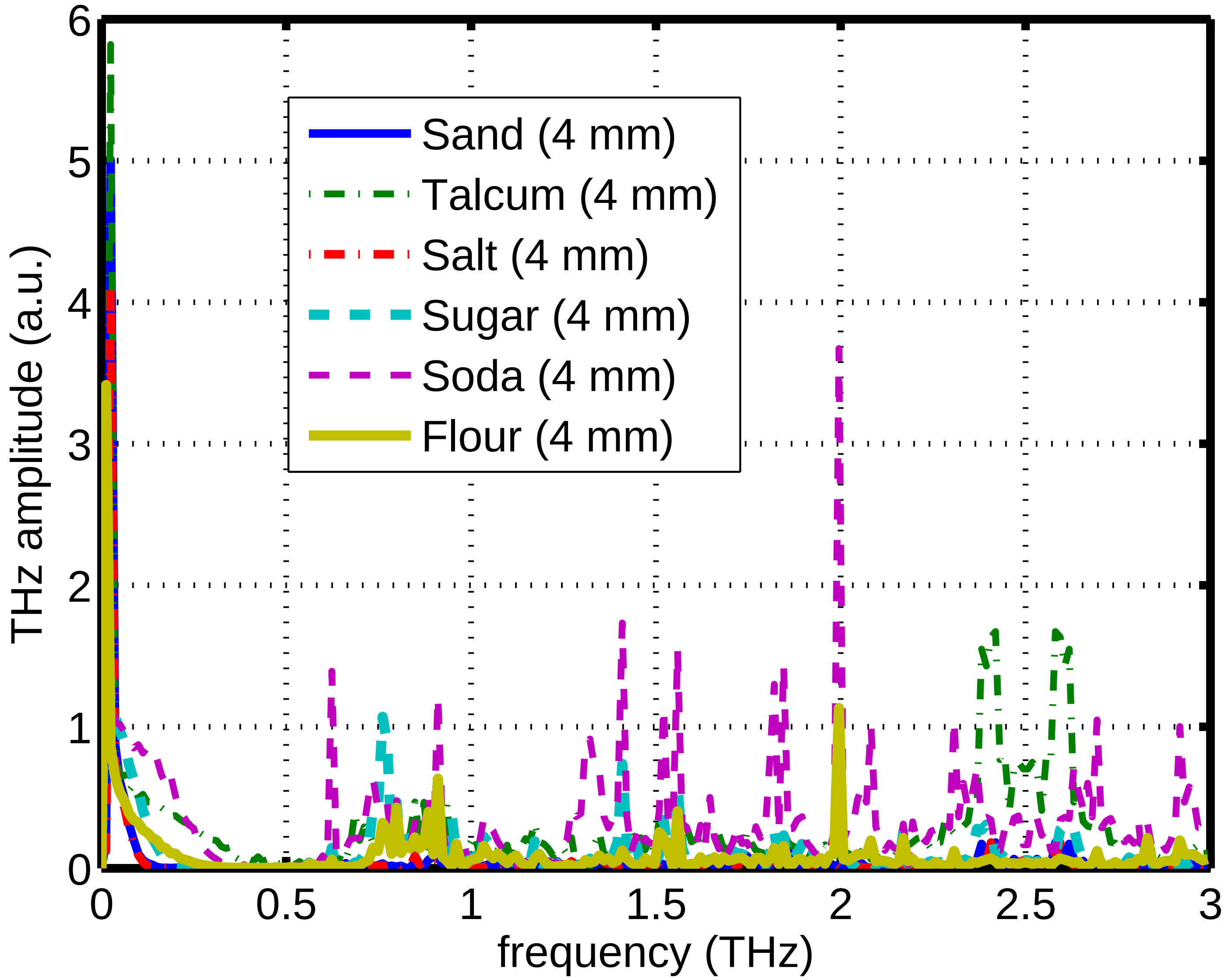
magitude_rna.eps



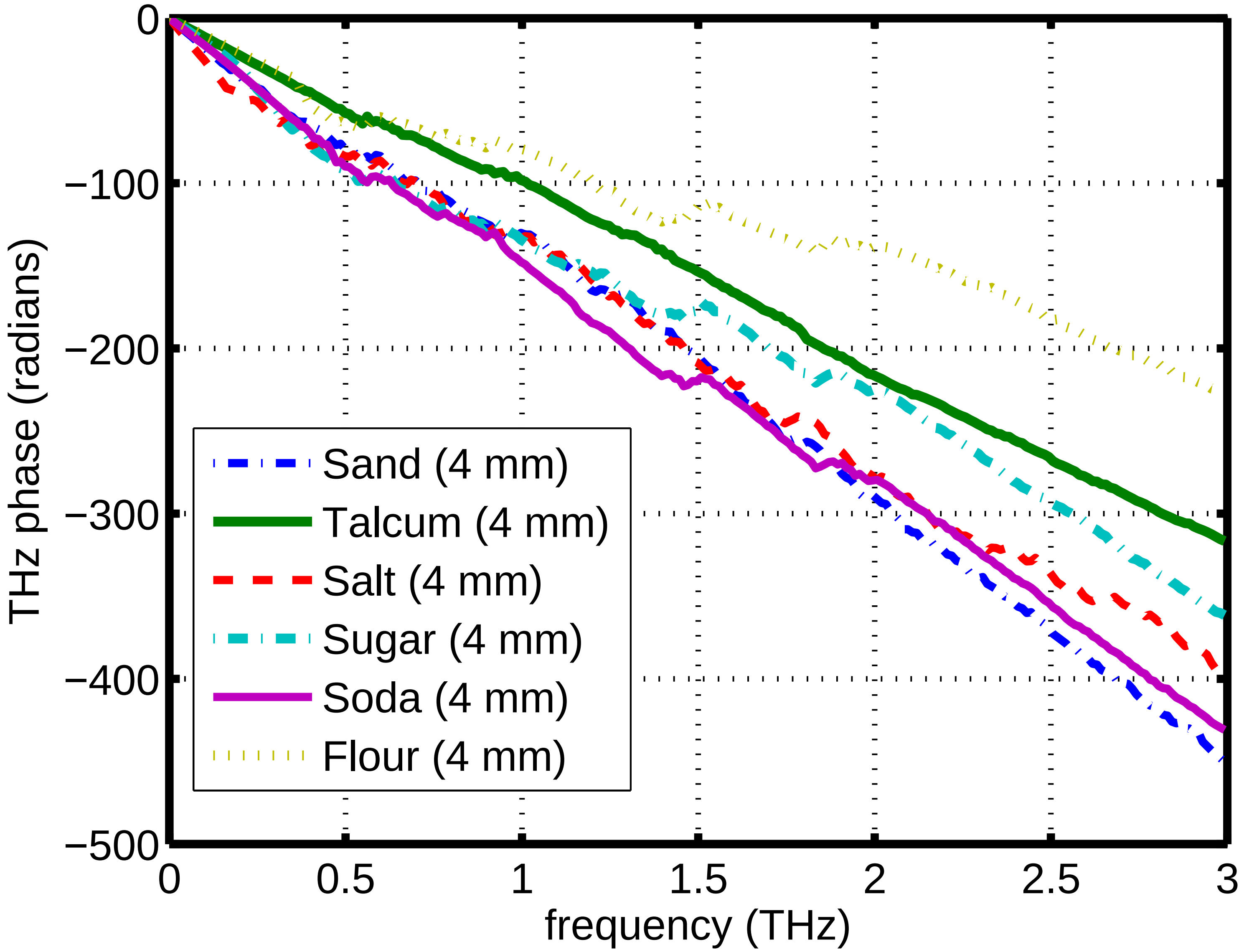
phase_rna.eps



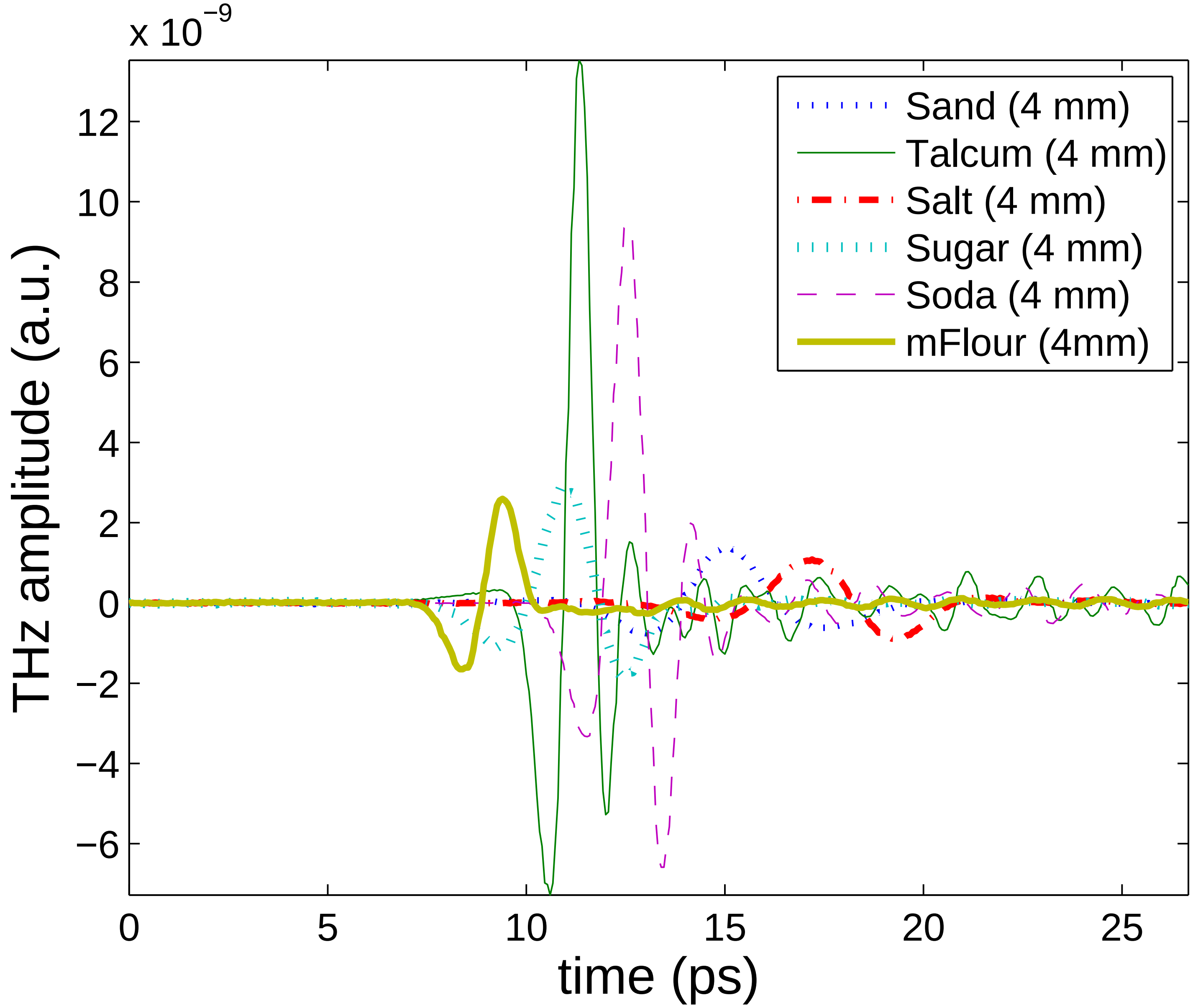
powder_magnitude.eps



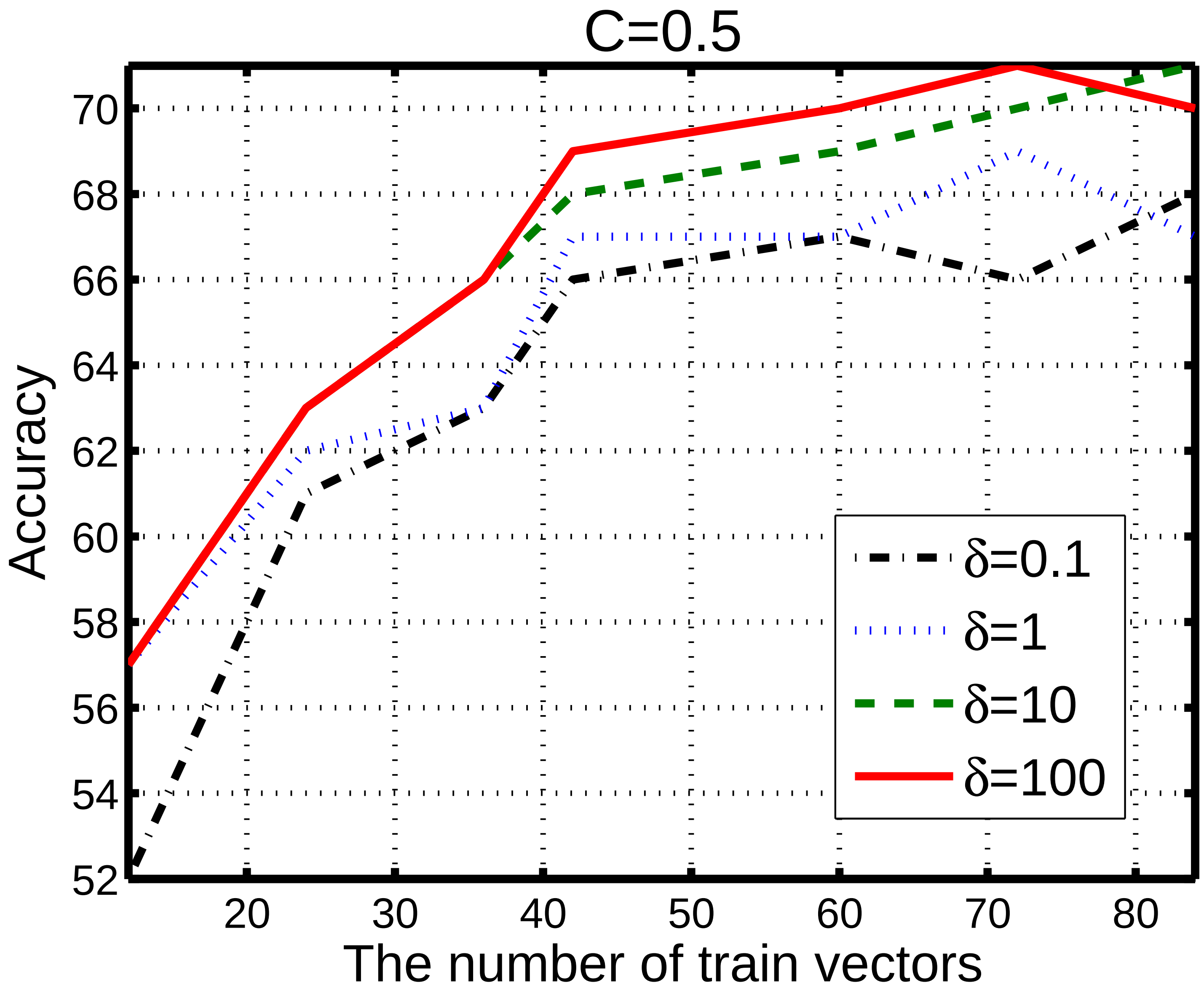
powder_phase.eps



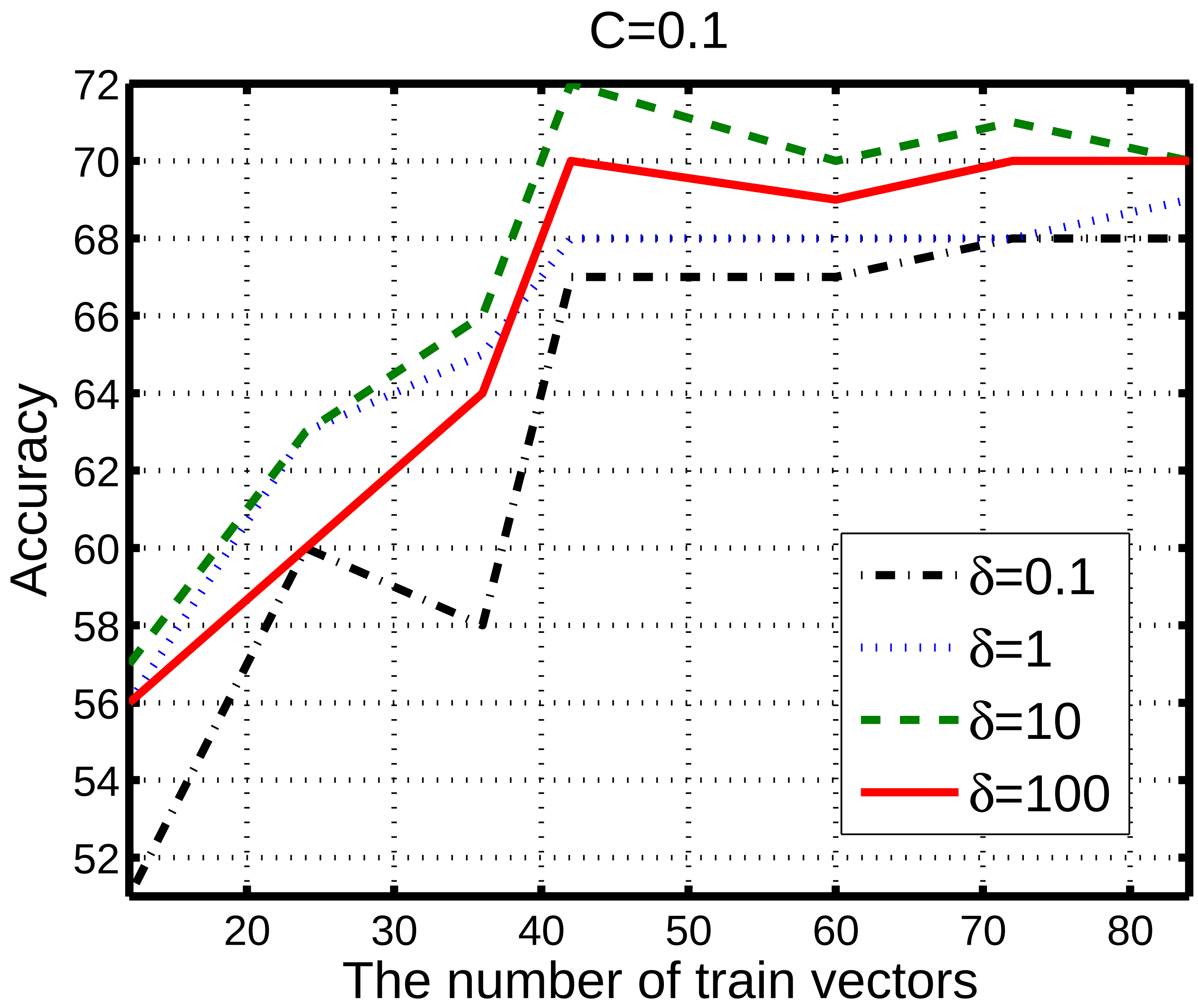
power_plot.eps



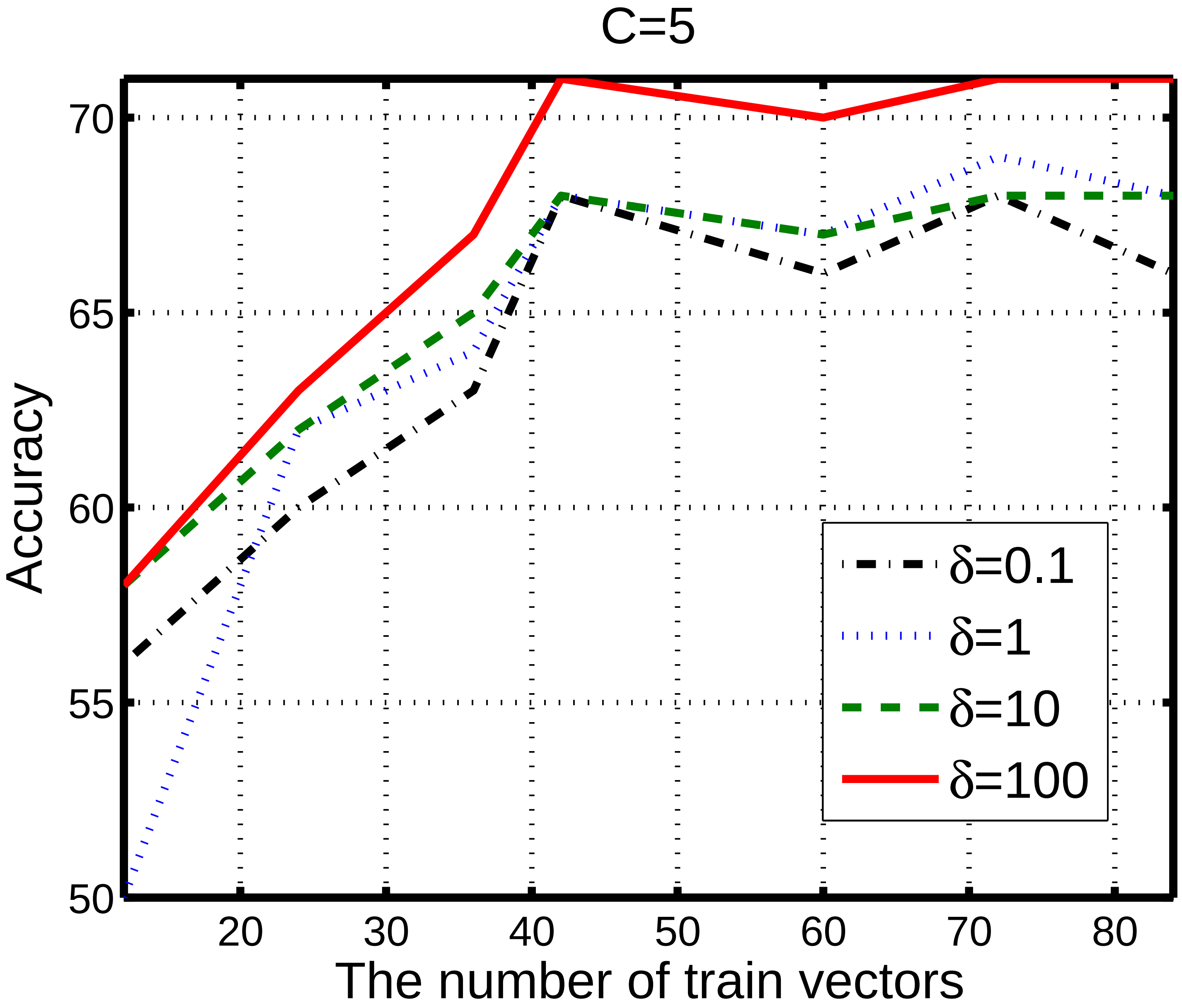
statisticaccuracy1.eps



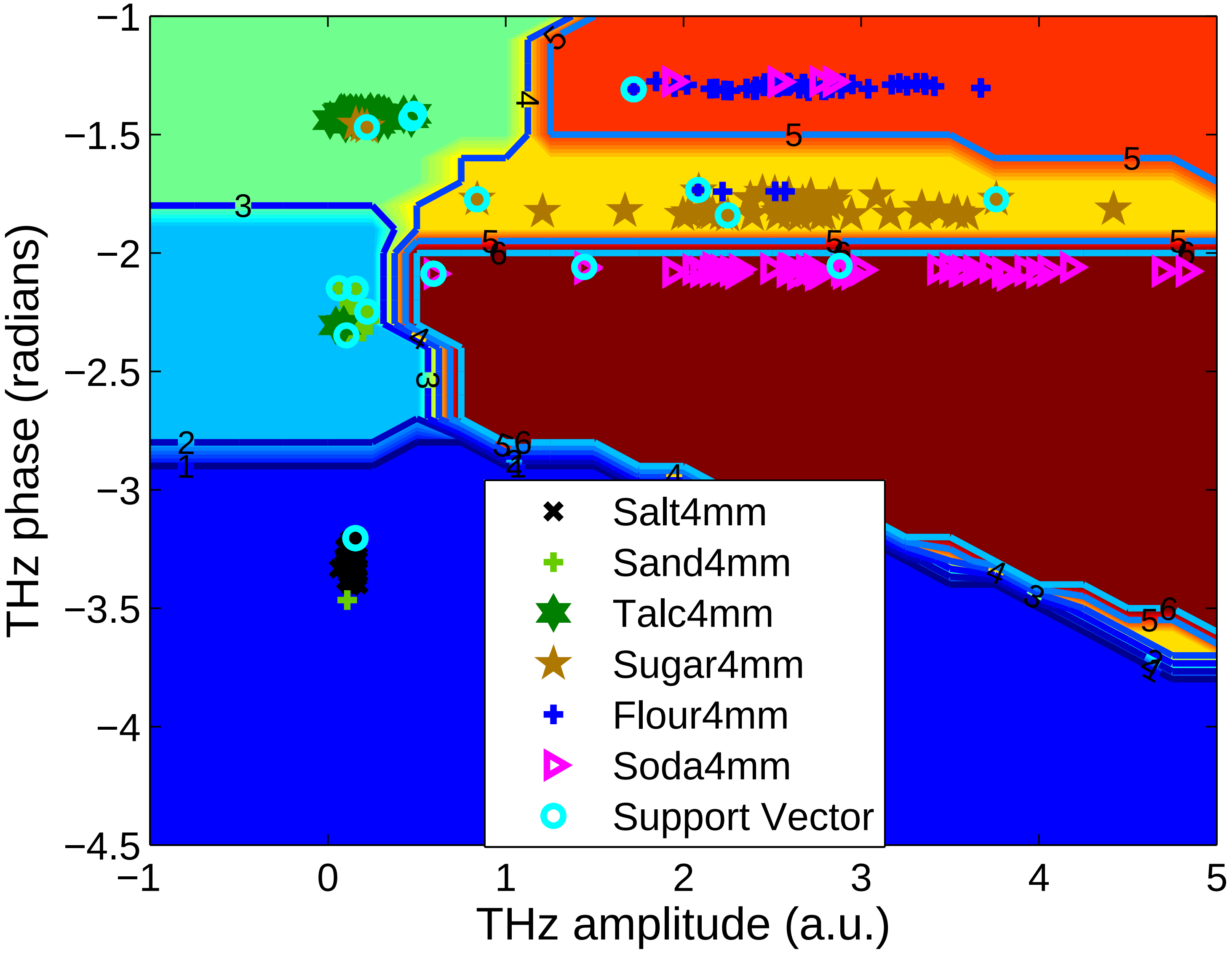
statisticaccuracy2.eps



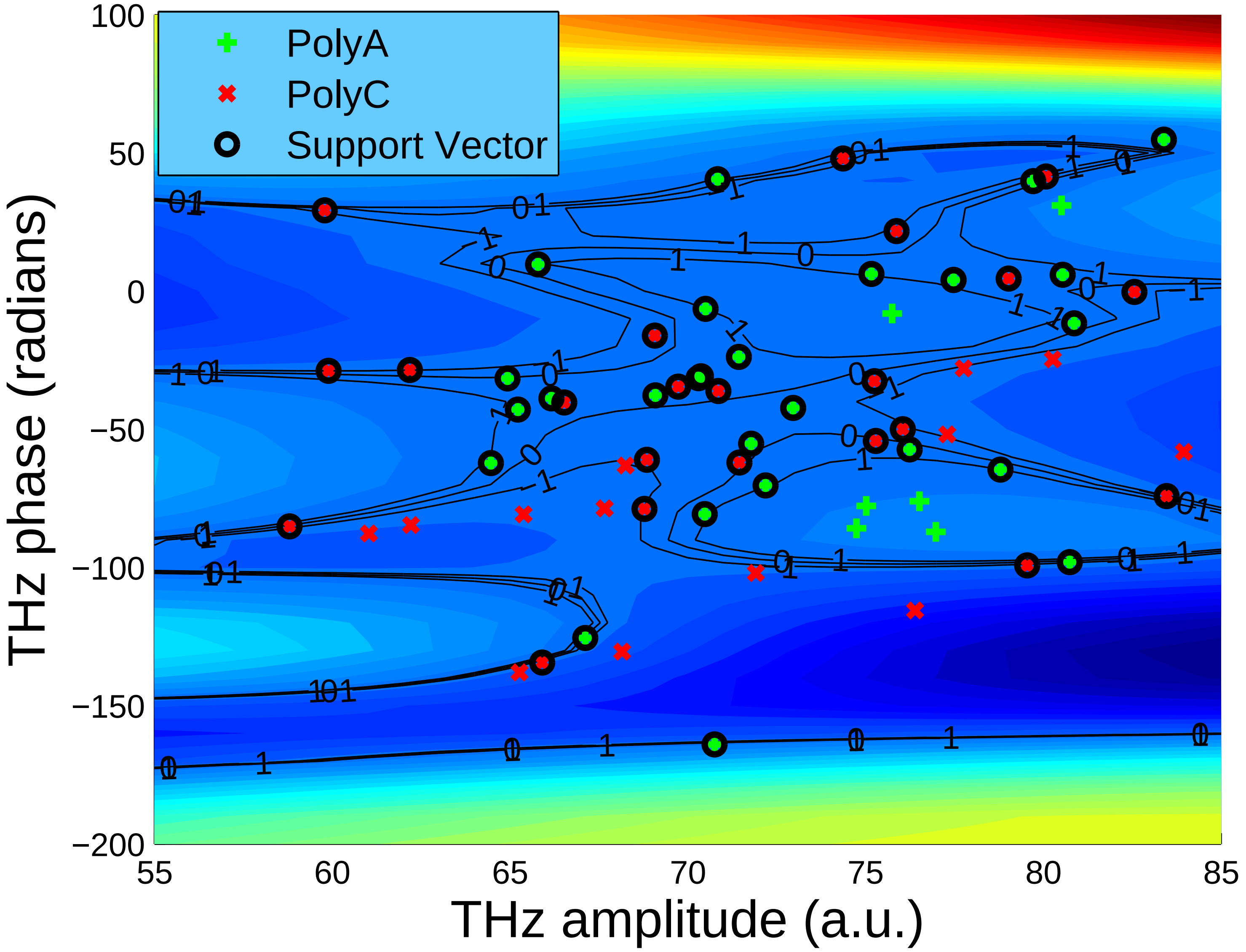
statisticaccuracy3.eps



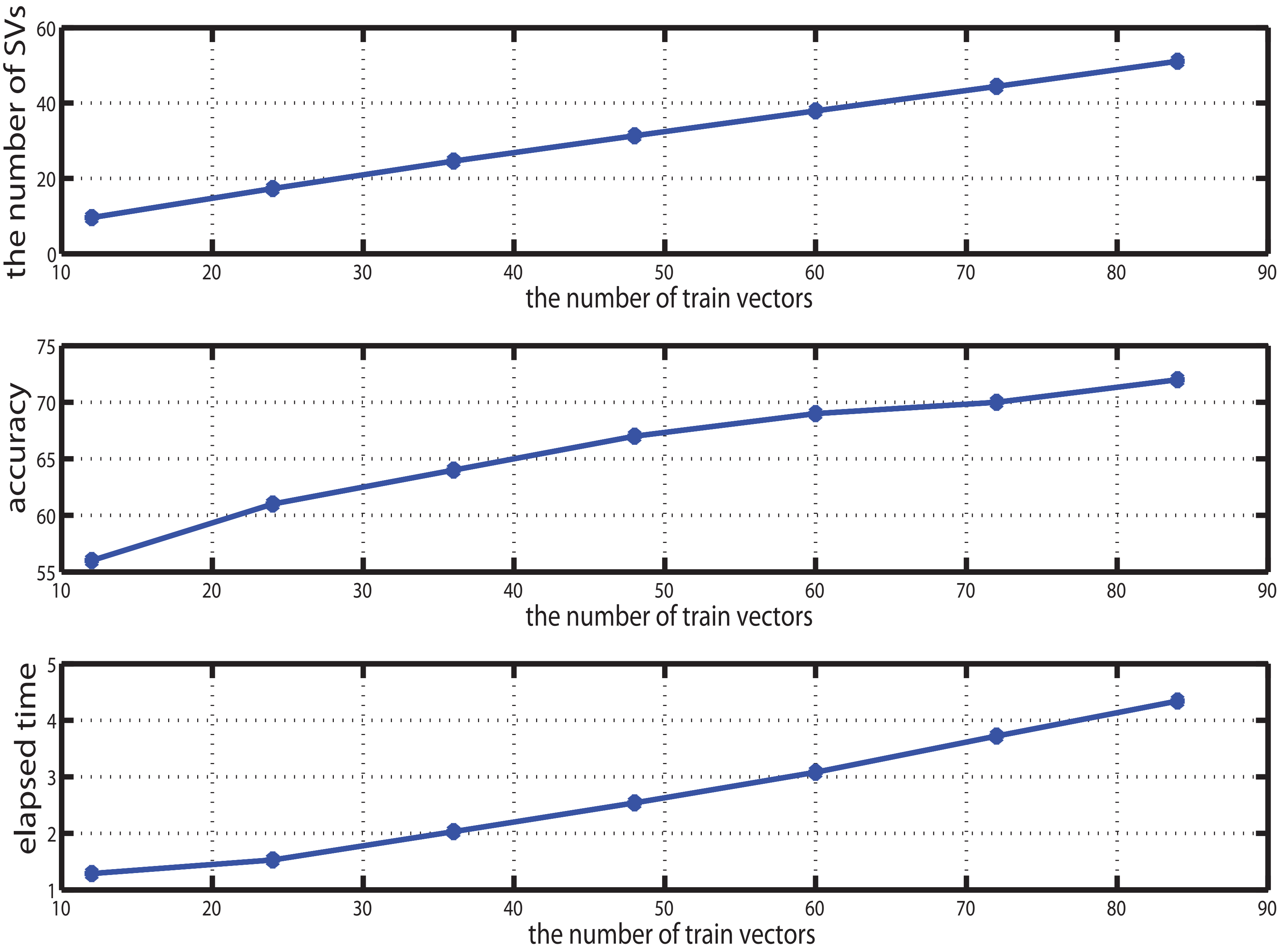
svmclassification.eps



svms_rna.eps


PolyA
PolyC
Support Vector
100
50
0
−50
−100
−150
−200
55
60
65
70
75
80
85
THz phase (radians)
THz amplitude (a.u.)

svsaccuracy3.eps



transmissionmode.eps