Database of frequency-domain terahertz reflection spectra for the DETRIS project

Paweł Piotr Cielecki, Mathias Hedegaard Kristensen and Esben Skovsen

Department of Materials and Production, Section for Physics and Mechanics, Aalborg University, Skjernvej 4A, DK-9220 Aalborg East, Denmark

Hereby, we present a database of frequency-domain terahertz reflection spectra for the project Detection of explosives using terahertz radiation at improved standoff-distances (DETRIS) conducted at Aalborg University in collaboration with MyDefence and with help from Danish Defence and Danish Ministry of Defence Acquisition and Logistics Organisation, and financial support from The Innovation Fund Denmark.

1.1 Samples

Terahertz (THz) reflection spectra contained in the database were acquired from solid-state samples in form of truncated cylinders with a diameter of 25 mm and weight of 7 g shown in the inset of Fig. 1. The angle of 15° between the back and the front surface of the sample was introduced to prevent interference. We selected six compounds with discernible spectral features within the investigated THz range. This includes:

Compound	***Name in the Database***
4-Aminobenzoic acid	PABA
Galactitol	Galactitol
Hexogen	RDX
L-tartaric acid	LTA
Theophylline	Theophylline
α-Lactose monohydrate	Lactose

We mixed each of the materials with a spectrally inert polyethylene (PE) powder, which functions as a binder, at weight percentages of the active material of 80%, 50% and 20%, respectively. PE was purchased from Micro Powders, LTA from MERCK, and RDX was supplied by the Danish Ministry of Defence Acquisition and Logistics Organisation. The remaining materials were purchased from Sigma Aldrich. The mixtures were compressed into samples under the pressure of approximately 4 tons using a hydraulic press. Galactitol and LTA were ground into a fine powder using mortar and pestle before mixing with PE. Otherwise, the obtained samples were brittle and had a tendency to break during removal from the press due to relatively large crystal size of these compounds. The obtained samples were used as is without any additional surface treatment like, e.g. polishing. In total, we fabricated 38 samples: Two samples for each material composition and two samples made of a pure PE, as summarized in following table:

Material	*Weight percentage of the active material*			
	20%	50%	80%	100%
Galactitol	● ●	● ●	● ●	
Lactose	● ●	● ●	● ●	
L-TA	● ●	● ●	● ●	
PABA	● ●	● ●	● ●	
RDX	● ●	● ●	● ●	
Theophylline	● ●	● ●	● ●	
PE				● ●

1.2 Terahertz setup and measurements

The setup used for THz characterization is based on the commercially available frequency-domain spectrometer TeraScan 1550 (Toptica Photonics) operating in a coherent detection scheme. The setup was arranged using 1" optics in the reflection configuration as shown in Fig. 1. The incidence angle was approx. 11° and the total path length of the THz beam was approximately 1 m. We performed all measurements using the parameters listed below:

- frequency scanning range – 0.09 – 1.19 THz,
- frequency step – 80 MHz,
- integration time – 3ms.

After each measurement, a computer-controlled two-axis translation stage with a fixed step of 0.5 mm moved the sample to a new (random) position within a 7 x 7 mm scanning area allowing measurements of different spots on the sample.

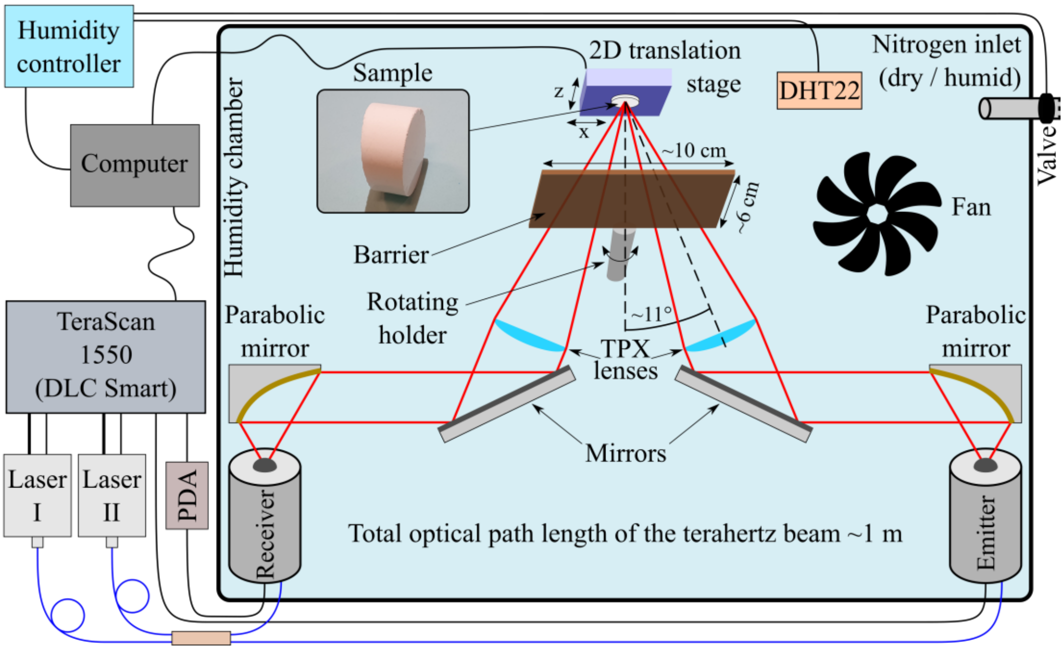

Fig. 1 Schematic drawing of the THz setup.

We conducted THz measurements under various experimental conditions divided into three datasets, as described below:

Datasheet I: Measurements performed under ambient conditions. We measured each sample 80 times in the series of 20 measurement. A single reference was recorded before each series of measurements. For that purpose, we replaced the sample with an aluminum mirror.

Datasheet II: Measurements performed under controlled humidity conditions. To this end, a custom-built humidity chamber that encloses the THz setup was purged with either dry or water vapor saturated nitrogen to lower or increase the humidity, respectively. A custom-built humidity controller opens and closes the nitrogen flow based on the readout from a DHT22 sensor, which allowed maintaining the relative humidity inside the chamber within ±0.3 percent points of the intended level. To ensure a uniform humidity distribution, a fan was installed next to the nitrogen inlet. We measure each sample 20 times at a relative humidity of 90%, 50% and 10%, respectively. Due to a long purging time, we measured the samples in groups of two and, at each relative humidity, we recorded a single reference measurement intended for both samples.

Datasheet III: Measurements performed under ambient conditions through various barrier materials described in the table below. The barrier was placed in a rotating holder with a clearance of ~10 cm by 6 cm located approximately 3.7 cm in front of the sample. In this configuration, the barrier blocks both the incoming and reflected THz beam, as it would be the case in real-world applications. Due to the large number of investigated barriers, these measurements were performed only on samples with the highest concentration of active material (80%) and PE samples. For each barrier material, we measured each of the considered samples 20 times. Each such measurement series was preceded by a reference measurement. The barrier was kept parallel to the sample surface for the first six measurements in the series. For the remaining measurements, a stepper motor coupled to the barrier holder rotated the barrier (around z-axis) by a fixed step in the range of approx. −32° to 32°.

Name of the barrier	***Description***
BubbleFoil	Two layers of packaging ‘bubble foil’
Cotton	Plain 100% cotton fabric obtained from a T-shirt
Duct	Plastic shopping bag (the same as in *PlasticBag*) with a single layer of a 0.23 mm thick duct tape (3M, Duct Tape 3939)
LDPE	Low-density polyethylene packaging foam with a thickness of approx. 5.3 mm
Paper	A sheet of office paper
PaperBag	Shopping bag made of recycled paper
PET	Polyethylene terephthalate (PET) obtained from a plastic bottle
PlasticBag	Plastic shopping bag made of low-density polyethylene
Polyamide	A plain fabric made of 96% polyamide and 4% elastane obtained from a T-shirt
Polystyrene	polystyrene sheets with thickness of approx. 4.5 mm and approx. 6.5 mm were used interchangeably

1.3 Database structure and file details

The database consists of raw .txt files obtained from a frequency-domain THz spectrometer. The files contain three tab-delimitated columns. The first column represents a set frequency expressed in GHz. The second column contains the values of the measured photocurrent in nA. The third column contains the actual frequency given in GHz. The set frequency represents an expected frequency value and may differ from the actual frequency. Therefore, it is recommended to use the actual frequency (v) listed in the third column. For more information, we refer to Toptica’s manual for TeraScan 1550 system.

The database contains 8467 files (8118 measurements and 349 references) arranged into three datasets described earlier. **Dataset I** contains 3190 files (3038 measurements and 152 references) categorized by the active material and in the next directory level by material concentration. As with all the datasets, the reference files are always in the same folder as the corresponding measurement files. Noteworthy, there are 160 measurements for each material composition except of Lactose80 and PABA80. For these material compositions, only 159 have been performed due to a processing error. **Dataset II** consists of 2337 files (2280 measurements and 57 references). As with **Dataset I**, the files are arranged by the active material in the first directory level and by material concentration in the second. Finally, **Dataset III** contains 2.940 files (2800 measurements and 140 references) arranged by the barrier material.

The measurement files in the database are named according to the following scheme:

$$\overbrace{\text{2020\_07\_17}}^{date}\_\underbrace{\text{Theophylline80}}_{sample\ details}\_\overbrace{\text{BubbleFoil}}^{conditions}\_\underbrace{0023}_{number}.\text{txt} \tag{1}$$

The filename starts with the *date* of the measurement followed by *sample details*, experimental *conditions*, and ends with the measurement's *number*. The *date* is in the format of YYYY_MM_DD. *Sample details* consist of two parts: The name of the active material and a numeric value corresponding to its weight percentage in the sample. For samples made of pure PE, which do not contain any active material, the second part of *sample details* is omitted. In this case, *sample details* are only 'PE'. Experimental *conditions* describe either relative humidity given in form of 'RHxx', where xx is the percentage value of the relative humidity (this applies to **Dataset II**) or a name of the barrier material used in the measurement (this applies to **Dataset III**). For **Dataset I**, which contains measurements under ambient conditions and without any barrier materials, the experimental *conditions* field is not used. Since we performed multiple measurements of each sample at a given set of experimental conditions, similar measurements were given consecutive *numbers*. The *number* has form of four digits padded with zeros before the value. Hence, the sample measurement filename from (1) corresponds to the 23rd measurement recorded on the sample with 80% weight percentage of Theophylline through BubbleFoil.

The reference files are named in a similar fashion:

$$\overbrace{\text{2020\_07\_17}}^{date}\text{\_reference\_for\_}\underbrace{\text{Theophylline80}}_{sample\ details}\_\overbrace{\text{BubbleFoil}}^{conditions}\_\underbrace{0021-0040}_{range}.\text{txt} \tag{2}$$

At the beginning of the filename, the *date* when the reference was recorded is given followed by '*reference_for*'. Subsequently, there is information about the measurement files, which the reference is intended for. This includes *sample details*, experimental *conditions*, and the *range* of measurements' numbers, which are analogous to the parts of measurement filenames. *Sample details* is the name of the active material and its weight percentage in the sample. For the measurements from **Dataset III,** experimental *conditions* correspond to the applied barrier material used, while for measurements from the **Dataset I**, this field is unused. Because we intended to use references and measurements recorded at different humidity levels with each other, reference files in **Dataset II** are named differently. We did not state the relative humidity during sample measurements in the experimental *conditions* field as the reference is not limited for use with measurements performed at a single relative humidity. Instead, we specified the relative humidity level during recording the reference. To this end, we replaced '*reference_for*' with '*reference_RHxx_for*', where xx is the percentage value of the relative humidity. Based on the above description, the sample reference filename from (2) is intended for measurements 21 to 40 performed on the sample containing 80% weight percentage of Theophylline with a BubbleFoil used as a barrier material. This includes the measurement with name (1).

1.4 Data processing and MATLAB repository

Due to characteristics of the coherent detection scheme, the measured photocurrent depends both on the amplitude of the THz electric field (E_{THz}) and the phase shift ($\Delta\varphi$) between the THz wave and the laser beat signal via $I_{ph} \propto E_{THz}\cos(\Delta\varphi = 2\pi\Delta L\nu/c)$. The phase difference depends on the THz frequency (ν) and the optical path difference $\Delta L = L_{THz} + L_E - L_R$, where L_{THz} is THz beam path, and L_E and L_R are optical beat paths to emitter and receiver, respectively. To obtain the THz spectrum, $E_{THz}(\nu)$, it is necessary to extract the amplitude of the detected THz signal from the interference pattern (also known as fringes) in the frequency scan [1–3]. A common approach is to find the fringe extrema, where the THz amplitude is directly proportional to the absolute value of the photocurrent ($\Delta\varphi$ is a multiple of π), and neglect the remaining data points. Consequently, the effective spectral resolution of the measurement is determined by the spacing between adjacent extrema, which in our case was around 0.9 GHz. Instead of analyzing the extrema separately, we calculated the THz amplitude and corresponding THz frequency based on the information contained in two adjacent extrema (maximum and minimum). This approach should compensate for a possible offset in the photocurrent. The obtained data points were then interpolated onto integer GHz frequencies to compensate for spectral shifts of the fringes. As the low frequency regime exhibited no material-related spectral features and was dominated by standing wave patterns, we cropped the spectra to range from 0.3 to 1.17 THz. All data processing was performed in MATLAB (version 2020b).

We saved the processed data together with the raw data in the .mat file named **Main**. There are two variables in the **Main** file: The structure array named **Database** and a vector named **Frequency**. The **Database** contains both raw and processed data from the THz measurements. Each row of the array corresponds to a single .txt file (either measurement or reference). The **Database** structure have the following fields:

Name of the field	*Description*
'FileName'	The name of the corresponding .txt file
'Folder'	Path to the folder containing the .txt file
'Frequency_raw'	Actual frequency, which corresponds to the third column in the .txt file (*raw data*)
'Photocurrent'	Measured photocurrent, which corresponds to the second column in the .txt file (*raw data*)
'E_THz'	Amplitude of the THz electric field extracted from the measured photocurrent (*processed data*)
'Reference'	The field indicates whether the file is a reference or the measurement. For reference this field is 'true'
'Dataset'	Indicates the dataset to which the file belongs, ∈{1, 2, 3}

The **Frequency** vector contains discrete frequencies (GHz) onto which the electric field amplitude THz has been interpolated. This vector is common for all the data. In order to obtain THz spectra one should plot E_THz of the considered measurement or reference versus the **Frequency** vector.

1.5 Contact information

In case of any questions related to the database, please contact us:

Paweł Piotr Cielecki pawel.cielecki@yahoo.com

Esben Skovsen es@mp.aau.dk

1.6 Acknowledgements

This work was financed by the Innovation Fund Denmark Grand Solutions program with grant no. IFD-7076-00017B. The project was performed in collaboration with MyDefence. The authors wish to express their gratitude to the Danish Defence, and Danish Ministry of Defence Acquisition and Logistics Organisation for knowledge and assistance in measurements of explosives.

1.7 References

[1] A. Roggenbuck, H. Schmitz, A. Deninger, I.C. Mayorga, J. Hemberger, R. Güsten, M. Grüninger, Coherent broadband continuous-wave terahertz spectroscopy on solid-state samples, New J. Phys. 12 (2010) 043017. doi:10.1088/1367-2630/12/4/043017.

[2] D.W. Vogt, R. Leonhardt, High resolution terahertz spectroscopy of a whispering gallery mode bubble resonator using Hilbert analysis, Opt. Express. 25 (2017) 16860. doi:10.1364/OE.25.016860.

[3] A. Roggenbuck, K. Thirunavukkuarasu, H. Schmitz, J. Marx, A. Deninger, I.C. Mayorga, R. Güsten, J. Hemberger, M. Grüninger, Using a fiber stretcher as a fast phase modulator in a continuous wave terahertz spectrometer, J. Opt. Soc. Am. B. 29 (2012) 614. doi:10.1364/JOSAB.29.000614.

Known issues:

- PABA20 measurements 101-160 in the Dataset I, the date in the filename is 2019_08_21 while it should be 2019_08_22