Fourier Transforms - New Analytical Approaches and FTIR Strategies

Edited by Prof. Goran Nikolic

ISBN 978-953-307-232-6
Hard cover, 520 pages
Publisher InTech
Published online 01, April, 2011
Published in print edition April, 2011

New analytical strategies and techniques are necessary to meet requirements of modern technologies and new materials. In this sense, this book provides a thorough review of current analytical approaches, industrial practices, and strategies in Fourier transform application.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Gür Emre Güraksın, Uçman Ergün and Ömer Deperlioğlu (2011). The Analysis of Heart Sounds and a Pocket Computer Application via Discrete Fourier Transform, Fourier Transforms - New Analytical Approaches and FTIR Strategies, Prof. Goran Nikolic (Ed.), ISBN: 978-953-307-232-6, InTech, Available from: http://www.intechopen.com/books/fourier-transforms-new-analytical-approaches-and-ftir-strategies/the-analysis-of-heart-sounds-and-a-pocket-computer-application-via-discrete-fourier-transform

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

1

Fourier Transform Infrared Imaging Spectroscopy in Biomedicine – Important Things to Consider When Planning a New Experiment

Rieppo J.[1], Rieppo L.[2], Saarakkala S.[3] and Jurvelin J.S.[2]
[1]*Department of Biomedicine, University of Eastern Finland, Kuopio,*
[2]*Department of Physics and Mathematics, University of Eastern Finland, Kuopio,*
[3]*Department of Diagnostic Radiology, University of Oulu and Oulu University Hospital, Oulu,*
Finland

1. Introduction

Fourier transform infrared imaging spectroscopy (FT-IRIS) offers unique possibilities to collect chemical information from biological samples with high spatial resolution (generally ~10 μm) [1]. The development of FTIR instruments has introduced this technique also for biomedicine. FT-IRIS is a demanding technique that requires a good understanding of the measurement and analysis principles. Application of the technique can be divided into three phases: 1) sample preparation 2) data collection and 3) data analysis. Each of the three steps is crucial for the outcome of the study and significant sources of error may exist in all three steps. Biologists and medical doctors are often not aware of the technical aspects of the measurement principle or the data analysis methods. Therefore, lack of information may impair successful application of this technique, and the full potential of the technique is not achieved.

This book chapter is written for an entry level user without the formal training for spectroscopy or data mining techniques. This chapter addresses important issues which should be considered when the FT-IRIS experiments are designed and data is analysed. We cover briefly: 1) the overall potential of FT-IRIS in qualitative and quantitative biomedical research, 2) essential issues to be taken into account in preparation and measurement of the biological samples and 3) different methods for analysis of acquired spectral data.

2. Planning a new experiment

FT-IRIS studies require careful planning of the study protocol beforehand. The first task is to define the aim of the study and to evaluate whether FT-IRIS is suitable for a planned experiment. FT-IRIS is a diverse technique that can be applied to numerous applications. Existing literature introduces numerous studies varying from simple qualitative applications to highly sophisticated applications of complex neural networks. FT-IR spectroscopy has been applied e.g. to cancer diagnostics [2-7], bone diseases [8-10], osteoarthritis [11-13], neurological diseases [14-16] and atherosclerosis [17,18]. Each and

every study is different and the used study protocols can be transferred only in rare cases directly to a new study. In general, sample preparation has the least variation, and also the measurement conditions can be standardized up to certain point, but the data-analysis has to by designed for each study individually.

2.1 Sample preparation

FT-IRIS studies are carried out either using thin histological sections (transmission measurement) or by using smooth, polished surfaces (reflection measurement). Histological tissue sections can be produced from: 1) fixated samples embedded into paraffin or other resins or 2) application of cryosectioning. Both techniques have their own advantages and limitations, which are discussed next.

2.1.1 Chemically fixed histological samples

Chemical fixatives have been used for decades in histology for preparing thin tissue sections for microscopy. Typical sample preparation involves formalin fixation and decalcification process. Samples are further dehydrated with ascending series of alcohol, and finally lipids are removed with xylene before sample is infiltrated with liquid paraffin [19]. Paraffin embedding is needed to create support for fragile sections in order to be able to cut 3-7 µm-thick histological samples. Embedded samples are easier to cut and, therefore, they have been used most often in histological studies. However, it should be realized that chemical fixatives introduce a few problems in FT-IRIS studies. Fixatives form chemical cross-links to protein structures preventing tissue deformation during the section processing. This is advantageous when only the morphology of the sample is considered, as is the case in traditional histology. However, since chemical fixatives alter the proteins and carbohydrates of the sample, they have potentially negative effects on spectroscopic analysis. It is known that the effect of chemical cross-linking can be seen in IR spectra [20, 21]. This clearly indicates that the fixation has permanently changed the tissue chemical properties. However, chemical fixation is typically done in a standardized manner within one study, and its effects can be assumed to be similar for each sample. Therefore, chemical fixation does not necessarily hinder the use of embedded sections.

Another possible source of error is that residual traces of paraffin are evident in paraffin-embedded sections even after chemical removal of paraffin [22]. Instead of chemical removal of paraffin, it is also possible to subtract paraffin after the measurements by using paraffin spectra [5]. Fortunately, a paraffin spectrum contains only a few narrow peaks. In order to minimize the contribution of paraffin residues to the analysis, one might also simply exclude the paraffin peak areas from the analysis.

In addition, the fact that lipids and part of the solubilized proteins are most likely lost from the sample due to the sample processing must be taken into account. Thus, information arising from these compounds cannot be studied from the fixated samples. Taken together, these limitations should be taken into account when the use of chemically fixed tissue sections are planned in new FT-IRIS studies.

2.1.2 Cryosections

When cryosections are used in FT-IRIS measurements samples are not treated with fixatives prior to sectioning, and it requires only minimal sample preparation. Tissue is prepared and embedded into cryo-embedding medium. Subsequently, sample is frozen with liquid nitrogen and sectioned with a cryotome. However, here one should note that the accuracy of

section thickness of cryotome is not at the level of paraffin or plastic embedded samples. On the other hand, since the method does not require chemical processing of the sample, the sample is maintained close to its original biological form. Embedding media used for cryosectioning often stains the samples, which causes problems for data analysis. After cutting tissue samples with a microtome, sections have to be rinsed with water, but the lack of chemical cross-linking makes the thin sections pliant, and consequently sections wrinkle easily during the preparation. In general, repeated sectioning needs to be done until the good quality section is found for the final measurements.

Advice: Chemical treatment of the sample is not needed with cryosections. Tissue is minimally altered compared to chemical fixation. Soluble proteins and lipids can also be studied when cryosections are used.

Conclusions: Chemical fixatives alter the collected IR data. Different fixatives cause different alterations and therefore sample processing must be standardized between the specimens. Fixative-related alterations are not a problem with simple univariate analysis techniques, but they may cause potential artefacts when sophisticated multivariate techniques are used. Chemical fixation and embedding is also poorly suited for research problems when solubilized proteins or lipids are main interests of the study. Furthermore, a significant section thickness variation caused by a microtome itself exists both with embedded sections and cryosections. This variation has to be kept in mind in quantitative analysis and compensated, if necessary, with repeated measurements or with the reference sample technique [23]. Preparation of the cryosections is fast but the overall time consumption is considerable since the quality of cryosection is inferior to embedded sections.

2.1.3 The effect of the variable section thickness

Section thickness varies significantly between histological sections due to the various uncertainties (e.g. temperature changes during cutting, inadequate embedding media support, microtome-related inaccuracy and sample-related variation). In general, the variation is smaller with embedded samples. Section thickness variation affects directly the measurement results, since according to Beer-Lambert law, absorption is directly related to section thickness. Normalization (e.g. vector normalization) of the spectral data is usually conducted before qualitative multivariate analysis. Therefore, section thickness variation is usually not a problem with qualitative analyses. However, quantitative analysis requires a strict control of the section thickness since the thickness variation is one of the main sources of errors in the FT-IRIS experiments. Variation is particularly harmful in sample-to-sample correlation analysis where FT-IRIS measurements are correlated with the reference technique. The nature of section thickness error is random, and therefore its distribution can be assumed as Gaussian. Consequently, the section thickness error is not as significant when group means of different sample groups are compared. However, it is essential to keep in mind that the error caused by the section thickness variation can be greater than the biological variation itself.

Advice: Evaluate whether the type of experiment you are conducting can be affected by the variable section thickness. If quantitative analysis from FT-IRIS spectra is carried out and compared with reference methods (correlation analysis), a good control of section thickness is essential. On the other hand, qualitative (multivariate) analysis is least affected by the section thickness variation. As a rule of thumb, section thickness artefact can be reduced by averaging multiple measurements from single specimen or by using homogenous reference material cut along with biological tissue, as described by Rieppo et al. (2004) [23].

2.2 Data collection

Type of the experiment sets the requirements for data collection. Acquisition of spectral data has to be optimized for signal quality and the measurement time. Measurement of a single specimen can be very time consuming when high quality data is measured and large areas are covered. A simple univariate data analysis can be carried out with low quality data, but when sophisticated multivariate analysis is carried out, the signal-to-noise ratio has to be adequate. The length of the data collection is determined by four factors: 1) measured spatial area, 2) spectral resolution, 3) spatial resolution and 4) number of averaged scans. Spectral resolution, spatial area and spatial resolution are directly proportional to the data collection time.

2.2.1 Spatial resolution

Modern imaging instruments allow the use of different spatial resolutions. Data collection time increases significantly when spatial resolution is increased due to the decreased signal-to-noise ratio and the increased amount of the measured pixels. Better spatial resolution requires more repeated scans to obtain equal signal-to-noise ratio. It is important to notice that spatial resolution is always limited by the diffraction due to the long wavelength used in mid-IR measurements. Spatial resolution can be approximated to be one-half of the used wavelength. Thus, best spatial resolution is achieved with the shortest wavelength of the measured spectrum (approx. 2.5 µm with 4000 cm^{-1}), and the resolution decreases gradually as the wavelength increases. However, optimal spatial resolution requires synchrotron operated devices due to the poor S/N-levels at diffraction limited resolutions. Consequently, in practice with standard thermal globar IR light sources the resolution is limited to ~7-15 µm depending on the used wavelength (1700-700 cm^{-1}). A true spatial resolution of the FT-IRIS device is defined as the minimum distance where two separate features can be fully separated from each other. Many FT-IRIS devices offer a better pixel resolution by magnification but that is only nominal resolution of the device (limited by sample stage movement or magnification), as actual signal, limited by the physical diffraction, arises from larger area. When planning FT-IRIS investigations one should carefully think whether there is a need for the smallest pixel sizes. For example, increase from 6.25 µm to 25 µm increases the number of measured pixels to 16-fold. Furthermore, if an equal signal-to-noise ratio is wanted, the measurement time is increased even more.

Advice: High spatial resolution tremendously increases the measurement time. A very critical evaluation has to be carried out whether the diffraction limited resolution is needed for carrying the experiment. High quality spectral data is preferential to high spatial resolution in most cases. Better spatial resolution increases the measurement time mainly because the number of repeated scans has to be increased to simultaneously keep signal-to-noise-ratio at an adequate level.

2.2.2 Spectral resolution

Spectral resolution is the accuracy (measured as wavenumbers) at which the spectral data is acquired. Depending on the type of experiment, spectral resolution is typically 4-16 cm^{-1}. Sharp spectral features are only seen with a high spectral resolution. Therefore, it is advantageous to increase the spectral resolution up to a certain point. On the other hand, required measurement time is directly proportional to the spectral resolution. It is important to remember that as the resolution increases, so does the noise. Thus, number of repeated scans has to be increased if the S/N-ratio is kept unchanged, and therefore duration of the measurement is considerably longer. It has been demonstrated that more spectral features

can be seen with 4 cm^{-1} resolution compared to 8 or 16 cm^{-1}. The effect is particularly evident when 2nd derivative spectra are used in the analysis (Figure 1). Most of the reported FT-IRIS studies have been conducted using either 4 or 8 cm^{-1} resolution. Application of the 2 cm^{-1} resolution is not typically used for two main reasons: 1) protein features generally lack very sharp peaks, i.e., additional information is limited, and 2) measurement time becomes impractical due to poor signal-to-noise-ratio.

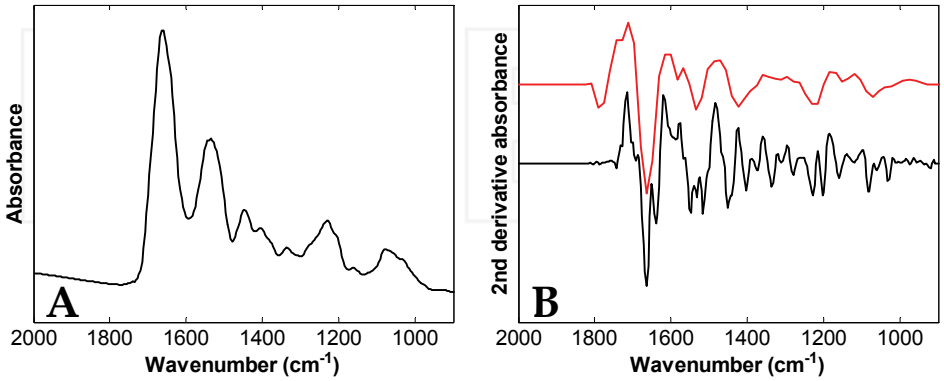


Fig. 1. An IR absorbance spectrum of articular cartilage measured with 4 cm^{-1} spectral resolution (A) and second derivative spectra with 4 cm^{-1} (black) and 16 cm^{-1} (red) spectral resolutions (B)

Advice: Simple univariate analysis can be carried out using spectral resolution of 8-16 cm^{-1}. If second derivative spectra are used for analysis, spectral resolution should be increased to 4-8 cm^{-1} to gain any advantage. From the theoretical point of view, a better spectral resolution reveals more information from the sample, but in practice measurement time is often a limiting factor. Signal-to-noise-ratio gets progressively worse as the spectral resolution increases. Spectral resolution can be considered as the most important parameter when a new FT-IRIS study is planned since it directly affects the subsequent possibilities for data analysis. Thus, data analysis methods (univariate vs. multivariate analysis) required need to be also planned beforehand. If numerous samples are measured, a careful balancing between the optimal spectral resolution, signal-to-noise ratio and measurement time is needed. Pilot studies are often essential to evaluate the spectral resolution needs against the time consumption.

2.2.3 Standardized measurement conditions

IR measurements are significantly affected by the carbon dioxide and water vapour of the atmosphere [24]. Humidity is substantially changed between winter and summer time causing a significant variance to the atmospheric conditions. Altered measurement conditions hinder the data quality when multivariate analyses are carried out. Standardized measurement conditions are, thus, essential for FT-IRIS measurements. Equipment is typically purged with N_2-gas or dried air. For example, water vapour free air can be guided into spectrometer, microscope and sample compartment. Furthermore, thin histological

sections absorb moisture from the environment. Therefore, it is beneficial to store the samples in a desiccator where steady environment can be maintained.
Advice: Each measurement system and sample type has to be separately tested for the measurement stability. Temperature and humidity levels of the laboratory should be kept as constant as possible. Separate purge system can be used to obtain low humidity and carbon dioxide levels in the measurement chamber. It is good to keep in mind that the measurement environment is altered each time the specimen is changed. It is a good idea to measure the duration when steady measurement conditions are reached after the sample compartment has been opened. A regular quality control also ensures that the measurements are consistent with each other regardless of the measurement time.

3. Data analysis – a crucial step to get the answers for your research problem

Data-analysis is the most difficult part of the FT-IRIS experiments. A good quality data set is a prerequisite for successful spectral analysis. Raw spectral data gives only a little information without proper knowledge of spectral pre-processing and analysis. Chemometric methods vary and offer different approaches to gather specific information from the measured data. Analysis has to be designed to match the desired research problems.

3.1 Spectral pre-processing

Spectral pre-processing is an essential procedure prior to the actual spectral analysis. Pre-processing is particularly needed when data from several studies are combined. Data has to be evaluated and filtered so that the non-biological variation between different measurements is minimized. Pre-processing routines include data quality analysis (signal-to-noise-ratio criteria, water vapour limitations, removal of corrupted data), measurement area criteria (definition of the region of interest, removal of unwanted pixels), selection of spectral region (data truncation, masked areas) and baseline corrections. In general, spectral pre-processing should be conducted in a way that do not alter the actual biological information but reduces the variance caused by the measurement itself or by the measurement conditions.
There exists different kind of baseline correction methods. Offset correction is simply conducted by setting, e.g., the minimum value of the spectrum (or alternatively a wavenumber that should not have any absorbance) to zero level. Linear baseline correction is done by fitting a line through two zero-absorbance points. The line is subsequently subtracted from the spectrum. Also polynomial-based fitting is used, but too heavy correction might alter not only the baseline but also the actual spectroscopic data. A more realistic model-based approach has also been used for data pre-processing. In Extended Multiplicative Signal Correction (EMSC), offset error, linear error and second order polynomial error are fitted simultaneously. The model uses a good quality reference spectrum when estimating the baseline errors, which makes EMSC a reliable method for baseline correction [25-28]. Application of second derivative spectra for the analysis eliminates the need for baseline correction as the differentiation removes most significant offset and linear baseline errors [27]. Proper pre-processing is essential, as baseline variations might hide the real biological variation if the baseline errors are not removed.
Advice: Pay attention for the spectral pre-processing before actual data-analysis. Pre-processing steps and demands depend on the data analysis method.

3.2 Univariate methods

Univariate methods are the simplest and most used data analysis techniques in IR spectroscopy. Univariate analyses are usually carried out by using peak height, integrated peak area or peak ratio for quantitative spectroscopic measurements [29]. These methods are hindered by the fact that biological tissues are essentially composed of the same building components regardless of the tissue type. Therefore, univariate methods are not suitable for solving complex research problems. Biological tissues usually lack the distinctive spectral features, which hinders the possibilities of simple univariate analysis. It seems impossible to achieve specificity and sensitivity level of biochemical reference techniques with univariate based parameters in most biological research problems [30]. Univariate analysis is, however, useful if the tissue has major tissue constituents that can be isolated with a single parameter. For example, bone and partially also cartilage, are good examples of tissues where separate components can be measured with reasonable accuracy also with univariate methods [8, 9, 31, 32]. Application of second derivative spectroscopy offers enhanced spectral features that potentially increase the parameter specificity. Second derivative spectroscopy increases significantly the spectral features, but the noise related to measurements is also significantly amplified. This sets higher demands for data quality.

3.3 Multivariate methods

Multivariate methods use more than one variable at a time. Multivariate methods offer means to increase parameter specificity since a larger part of the spectral information is exploited than in univariate analysis. Multivariate methods are suitable both for qualitative and quantitative analysis. The main difference between the univariate and multivariate methods is that the multivariate approach does not require accurate isolation of the differentiating spectral feature before the analysis is conducted. Univariate analysis is dependent on a priori knowledge of the spectral features and therefore the technique has limited efficiency. Multivariate techniques can handle the complete spectral data set with statistical means, and therefore the technique is more potential than user-limited univariate methods. Multivariate approach is often needed when specificity of the IR parameters has to be increased. This is commonly done by multivariate calibration, where a multivariate model is calibrated against some actual reference information. The model can be then used to obtain the same information from the spectroscopic data. Multivariate methods can also be used to separate two or more sample groups by spectral means.

4. How to select appropriate analysis method for different research problems?

A key element for successful use of the FT-IRIS is to be able to define the research question before actual experiments are started. Potential and the limitations of the FT-IRIS technique have to be kept in mind when study protocol is designed. Type of the research question determines the demands for sample harvesting and for type of analysis methods best suited for the particular research problem.

4.1 Quantitative measurement of tissue constituents

Quantitative measurement of different tissue constituents requires a specific parameter for a given compound. This is not a trivial demand to meet and often univariate parameters are not feasible. Multivariate models have a better chance to work since they do not require

fully separated, specific spectral features evident for researcher to be noticed. Univariate parameters are best suited in special situations where spectral differences between tissue components are evident, e.g., bone mineral and matrix content [8-10]. Major tissue constituents can probably be measured using univariate methods with reasonable specificity. However, if the studied compound is present only in small quantities, then only multivariate methods should be considered. Multivariate methods, such as principal component regression (PCR) and partial least squares (PLS) regression, offer more efficient means for quantitative analysis, since they can handle also overlapping spectral data. On the other hand, multivariate analysis needs good understanding of the background of the methods in use. Furthermore, in order to routinely use multivariate techniques to quantificate the composition of different tissue constituents, a comprehensive reference data set needs to be collected. This illustrates the complexity of FT-IRIS spectral analysis.

Example: Univariate vs multivariate analysis of proteoglycan distribution of articular cartilage

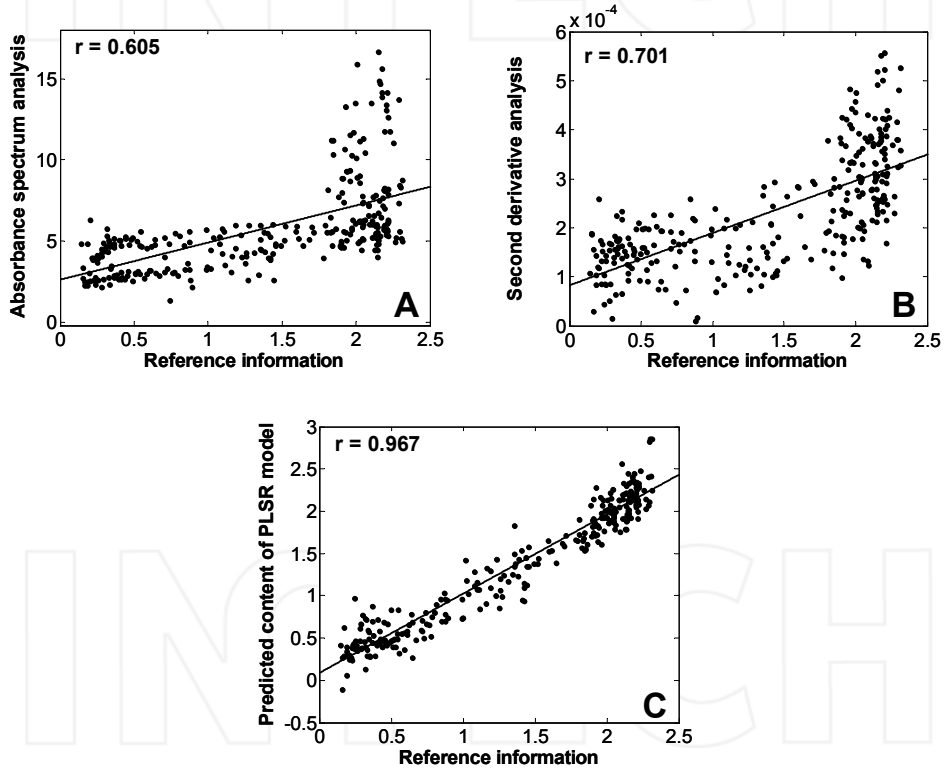


Fig. 2. FT-IRIS analysis of proteglycan content of articular cartilage. Univariate analysis from the absorbance spectrum (A) produces a little worse result than the second derivative analysis (B). However, a multivariate PLS regression model is clearly the most efficient analysis method

As a practical example, let us now consider univariate vs multivariate analysis techniques for determination of proteoglycan distribution in articular cartilage tissue. Univariate based solutions for FT-IRIS analyses of spatial proteglycan content in articular cartilage have been

used in several studies [30-33]. We have demonstrated that specificity for cartilage proteoglycans is significantly increased by taking advantage of the increased spectral separation of the second derivative spectroscopy. Furthermore, we have compared univariate results with the results of the PLS regression model. The PLS regression results were calculated using the whole collected spectral region. The results demonstrated that the PLS regression model is more consistent with the reference technique as compared to univariate methods [34] (Figure 2).

4.2 Qualitative visualization of tissue morphology

Tissue morphology is traditionally investigated through light microscopy of stained thin tissue sections. Morphological features are seen with specific staining patterns of different tissue types. FT-IRIS can produce similar information, but it does not require any staining. Image contrast is created with IR absorption of the tissue. Different tissue types absorb IR energy differently, i.e., their absorption spectra are different. However, it is often difficult to create contrast between different tissue types with univariate techniques. Therefore, if morphological features are one of the main interests of the study, multivariate techniques have to be considered. Application of the neural networks probably gives the most accurate results, but building such a model requires a large data pool and is a very time consuming procedure [3]. A large reference spectra data is collected and the model is trained to recognise spectral features of different tissue types. Nevertheless, neural networks are accurate and fast way to analyze specimens once the neural network is established. Simple multivariate techniques can be used also within a single specimen by using cluster analysis. The aim of cluster analysis methods is to minimize spectral differences within clusters while maximizing differences between clusters. Therefore, tissue types can be classified into their own groups by cluster analysis. Cluster analysis methods, such as K-means clustering or fuzzy c-means clustering, arrange data into desired number of user-determined clusters according to the spectral features [4]. Hierarchical cluster analysis allows unsupervised clustering without pre-determined number of clusters [4].

Clustering methods are often used for isolating the regions of interest from the sample (e.g. cancer area from surrounding tissue). Clustering methods can be used with a limited number of samples but the calculations become very time consuming with large number of spectra or with large number of different samples.

Advice: Morphological information can be gathered with multivariate clustering methods. Data clustering can be done with various multivariate techniques depending on the research application. Clustering methods become essential when large tissue sections are used and only a part of the data is interesting. Proper spectral pre-processing is essential prior to clustering. Any impurities or foreign material, e.g., such as embedding material, are likely to produce a new cluster, which is particularly harmful when fixed number of cluster is used.

4.3 Classification studies

Multivariate techniques are particularly useful when research problem can be simplified into few classes (disease vs. non-disease, tissue types 1,2,3,4 etc). Univariate based parameters have only limited capability for clustering purposes. Multivariate clustering methods are especially useful when different classes are sought (Figure 3). Spectral data can be assigned into subsets according to its spectral features either in unsupervised or supervised manner. In unsupervised techniques, only unlabeled data is used as an input. Clustering is done blind without knowing any additional information of the studied sample.

Data is re-arranged only on the basis of their spectral properties. In supervised clustering, known class information is also included as an input, e.g. disease vs. non-disease. Known information is used to seek spectral features that can be linked to the known information. Once the model is built and verified, the data analysis can be done with great speed. This type of application can be used if the specific feature is looked from the samples.
Advice: Multivariate methods are powerful to find even the smallest chemical differences from the samples. However, analysis methods cannot automatically distinguish artefacts from true differences. Therefore, maintaining stabile measurement conditions and a proper spectral pre-processing is essential for application of multivariate techniques.

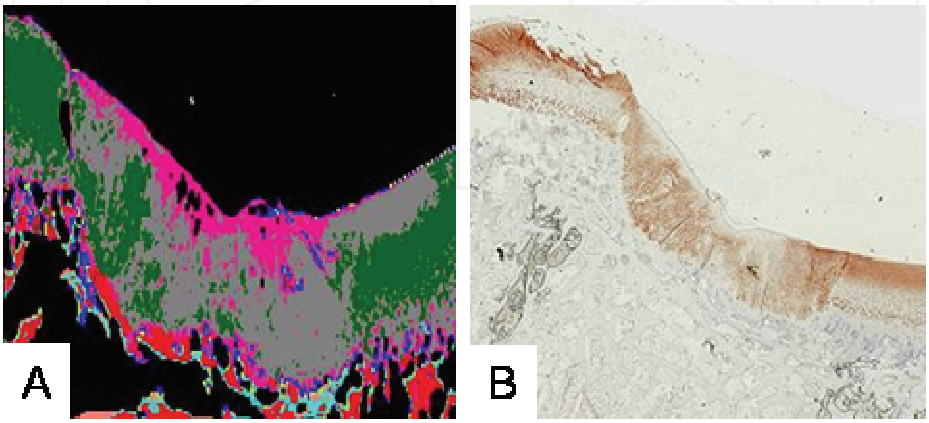


Fig. 3. Pictures of the FT-IRIS k-means clustering map (A) and a histological cartilage section stained with type II collagen antibody (B) that shows a repaired cartilage defect and surrounding intact cartilage. Clustering of the FT-IRIS map is done by mathematical calculation without any user intervention or a priori information of the tissue properties. Grey colour represents repair tissue, pink shows most likely tissue originating from the periosteal flap used in repair surgery and green colour indicates the surrounding intact cartilage

5. Conclusion

Application of FT-IRIS offers new potential for biomedicine. Biologist and medical doctors require training for data mining techniques since the methodology is still relatively new in biomedicine. Multivariate methodology has increasingly been used in spectroscopy, and the research questions are becoming more demanding all the time. Recent progress in biospectroscopy has shown to be fast. From the theoretical point of view, FT-IRIS can be used for the research problems that cannot be answered with traditional imaging techniques. FT-IRIS links the tissue molecular information together with histological imaging. FT-IRIS instruments have rapidly developed and become available for numerous research groups worldwide. Utilization of the modern data mining techniques is likely to further increase the development of the biospectroscopy.

6. References

[1] Lasch P, Naumann D. Spatial resolution in infrared microspectroscopic imaging of tissues. *Biochim Biophys Acta* 2006; 1758: 814-829.

[2] Bhargava R. Towards a practical Fourier transform infrared chemical imaging protocol for cancer histopathology. *Anal Bioanal Chem* 2007; 389: 1155-1169.

[3] Lasch P, Diem M, Hansch W, Naumann D. Artificial neural networks as supervised techniques for FT-IR microspectroscopic imaging. *J Chemom* 2007; 20: 209-220.

[4] Lasch P, Haensch W, Naumann D, Diem M. Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis. *Biochim Biophys Acta* 2004; 1688: 176-186.

[5] Ly E, Piot O, Wolthuis R, Durlach A, Bernard P, Manfait M. Combination of FTIR spectral imaging and chemometrics for tumour detection from paraffin-embedded biopsies. *Analyst* 2008; 133: 197-205.

[6] Sahu RK, Mordechai S. Fourier transform infrared spectroscopy in cancer detection. *Future Oncol* 2005; 1: 635-647.

[7] Zwielly A, Mordechai S, Sinielnikov I, Salman A, Bogomolny E, Argov S. Advanced statistical techniques applied to comprehensive FTIR spectra on human colonic tissues. *Med Phys*; 37: 1047-1055.

[8] Boskey A, Mendelsohn R. Infrared analysis of bone in health and disease. *J Biomed Opt* 2005; 10: 031102.

[9] Boskey A, Pleshko Camacho N. FT-IR imaging of native and tissue-engineered bone and cartilage. *Biomaterials* 2007; 28: 2465-2478.

[10] Boskey AL, DiCarlo E, Paschalis E, West P, Mendelsohn R. Comparison of mineral quality and quantity in iliac crest biopsies from high- and low-turnover osteoporosis: an FT-IR microspectroscopic investigation. *Osteoporos Int* 2005; 16: 2031-2038.

[11] Bi X, Yang X, Bostrom MP, Camacho NP. Fourier transform infrared imaging spectroscopy investigations in the pathogenesis and repair of cartilage. *Biochim Biophys Acta* 2006; 1758: 934-941.

[12] Camacho N, West P, Yang X, Lin J, Bostrom M. An infrared fiber optic probe for detection of degenerative cartilage. *Trans Orthop Res Soc* 2002; 27: 225.

[13] West PA, Bostrom MP, Torzilli PA, Camacho NP. Fourier transform infrared spectral analysis of degenerative cartilage: an infrared fiber optic probe and imaging study. *Appl Spectrosc* 2004; 58: 376-381.

[14] Lasch P, Beekes M, Schmitt J, Naumann D. Detection of preclinical scrapie from serum by infrared spectroscopy and chemometrics. *Anal Bioanal Chem* 2007; 387: 1791-1800.

[15] Lasch P, Schmitt J, Beekes M, Udelhoven T, Eiden M, Fabian H, et al. Antemortem identification of bovine spongiform encephalopathy from serum using infrared spectroscopy. *Anal Chem* 2003; 75: 6673-6678.

[16] Peuchant E, Richard-Harston S, Bourdel-Marchasson I, Dartigues JF, Letenneur L, Barberger-Gateau P, et al. Infrared spectroscopy: a reagent-free method to distinguish Alzheimer's disease patients from normal-aging subjects. *Transl Res* 2008; 152: 103-112.

[17] Colley CS, Kazarian SG, Weinberg PD, Lever MJ. Spectroscopic imaging of arteries and atherosclerotic plaques. *Biopolymers* 2004; 74: 328-335.

[18] Manoharan R, Baraga JJ, Rava RP, Dasari RR, Fitzmaurice M, Feld MS. Biochemical analysis and mapping of atherosclerotic human artery using FT-IR microspectroscopy. *Atherosclerosis* 1993; 103: 181-193.

[19] Király K, Hyttinen MM, Lapvetelainen T, Elo M, Kiviranta I, Dobai J, et al. Specimen preparation and quantification of collagen birefringence in unstained sections of articular cartilage using image analysis and polarizing light microscopy. *Histochem J* 1997; 29: 317-327.

[20] Aparicio S, Doty SB, Camacho NP, Paschalis EP, Spevak L, Mendelsohn R, et al. Optimal methods for processing mineralized tissues for Fourier transform infrared microspectroscopy. *Calcif Tissue Int* 2002; 70: 422-429.

[21] Pleshko NL, Boskey AL, Mendelsohn R. An FT-IR microscopic investigation of the effects of tissue preservation on bone. *Calcif Tissue Int* 1992; 51: 72-77.

[22] Faolain EO, Hunter MB, Byrne JM, Kelehan P, Lambkin HA, Byrne HJ, et al. Raman spectroscopic evaluation of efficacy of current paraffin wax section dewaxing agents. *J Histochem Cytochem* 2005; 53: 121-129.

[23] Rieppo J, Hyttinen MM, Jurvelin JS, Helminen HJ. Reference sample method reduces the error caused by variable cryosection thickness in Fourier transform infrared imaging. *Appl Spectrosc* 2004; 58: 137-140.

[24] Martens H, Bruun SW, Adt I, Sockalingum GD, Kohler A. Pre-processing in biochemometrics: correction for path-length and temperature effects of water in FTIR bio-spectroscopy by EMSC. *Journal of Chemometrics* 2006; 20: 402-417.

[25] Bassan P, Byrne HJ, Bonnier F, Lee J, Dumas P, Gardner P. Resonant Mie scattering in infrared spectroscopy of biological materials--understanding the 'dispersion artefact'. *Analyst* 2009; 134: 1586-1593.

[26] Bassan P, Kohler A, Martens H, Lee J, Byrne HJ, Dumas P, et al. Resonant Mie scattering (RMieS) correction of infrared spectra from highly scattering biological samples. *Analyst*; 135: 268-277.

[27] Kohler A, Bertrand D, Martens H, Hannesson K, Kirschner C, Ofstad R. Multivariate image analysis of a set of FTIR microspectroscopy images of aged bovine muscle tissue combining image and design information. *Anal Bioanal Chem* 2007; 389: 1143-1153.

[28] Kohler A, Kirschner C, Oust A, Martens H. Extended multiplicative signal correction as a tool for separation and characterization of physical and chemical information in Fourier transform infrared microscopy images of cryo-sections of beef loin. *Appl Spectrosc* 2005; 59: 707-716.

[29] Stuart B. Spectral analysis. In: Infrared Spectroscopy: Fundamentals And Applications, Struart B Ed. Kent: John Wiley & Sons 2004.

[30] Rieppo L, Saarakkala S, Narhi T, Holopainen J, Lammi M, Helminen HJ, et al. Quantitative analysis of spatial proteoglycan content in articular cartilage with Fourier transform infrared imaging spectroscopy: Critical evaluation of analysis methods and specificity of the parameters. *Microsc Res Tech* 2009.

[31] Camacho NP, West P, Torzilli PA, Mendelsohn R. FTIR microscopic imaging of collagen and proteoglycan in bovine cartilage. *Biopolymers* 2001; 62: 1-8.

[32] Kim M, Bi X, Horton WE, Spencer RG, Camacho NP. Fourier transform infrared imaging spectroscopic analysis of tissue engineered cartilage: histologic and biochemical correlations. *J Biomed Opt* 2005; 10: 031105.

[33] Bi X, Yang X, Bostrom MP, Bartusik D, Ramaswamy S, Fishbein KW, et al. Fourier transform infrared imaging and MR microscopy studies detect compositional and structural changes in cartilage in a rabbit model of osteoarthritis. *Anal Bioanal Chem* 2007; 387: 1601-1612.

[34] Rieppo L, Jurvelin JS, Rieppo J, Saarakkala S. Partial Least Squares Modeling of FTIR Microspectroscopy for Predicting Proteoglycan Content of Articular Cartilage: Preliminary Results. In: *SPEC: Shedding Light on Disease: Optical Diagnostics for the New Millennium*. Manchester 2010.

2

Analysis of Bioactive Olygosaccharide-Metal Complexes by Modern FTIR Spectroscopy: Copper Complexes

Goran S. Nikolić and Milorad D. Cakić
University of Niš, Department of Organic Chemical Technology, Faculty of Technology, Bulevar oslobodjenja 124, RS-16000 Leskovac Serbia

1. Introduction

Spectroscopy of biomolecules has been founded at the end of last century by the researchers working in the field of optical spectroscopy applied to biosystems. Since this time, the interest of the activity has considerably grown up. Researchers have typically used traditional spectroscopic techniques, such as Raman scattering, IR absorption, UV/Vis absorption, circular dichroism, fluorescence, magnetic resonance, X-rays and neutron scattering. Recently, particular attention has been devoted to the applications of biomolecular spectroscopy in the fields of biomedical imaging, drug characterization for pharmaceutical applications, drug delivery and nanobiotechnology.

Investigations of the bioactive metal complexes are very interesting in medicine and pharmaceutical industry, with the aspects on therapy of different states of anemia or metabolism disorder. On the other hand, polysaccharides and their derivatives, as the most abundant class of biomolecules, are known to have a large variety of biological functions. Through the interaction between these polyfunctional molecules and metal ions in living organisms, the modification of the biological function of both counterparts may be expected. The polysaccharide type compounds as ligands have received considerable interest. Simple sugars and their derivatives with reduced and oxidized groups form metal ion complexes of various composition and stability. One of the known roles of the oligo- or polysaccharide complexes is the transport of metal ions through cell membranes. For example, the commercial copper preparations based on polysaccharide dextran and its derivatives are used for such purpose in both human and veterinary medicine.

In the field of biocoordination chemistry a lot of investigations are based on the synthesis and characterizations of different metal complexes of ligands they present in biological systems, or synthetic ligands, which will serve like the model-molecules for complex biomolecular structures. Bio- or synthetic ligands are mainly natural chemical compounds of macromolecular type. In this group of products of the special importance are chemical compounds of olygosaccharide pullulan, dextran and inulin with cations of the different biometals (Cu, Fe, Co and Zn). It is well known that raw microbiological exopolysaccharides dextran and pullulan, are glucose polymers with the large molar mass from a few millions g/mol, with own toxic and antigen characteristics so that they are not of pharmaceutical

importance. For commercial reasons raw polysaccharides were depolymerized to the products with adequate molar masses, with the aim of getting fractions with narrow molar mass distribution. Synthesis procedures for the complex formation of biometals with poly- or oligosaccharides are described in scientific and patent literature. However, the structure of the bioactive metal complexes with oligosaccharides has not been explained in details yet, despite a number of studies. The work represents further development in research of complex structure and pharmacobiological activity of the complexes. Some new results that are directly related to medical practice and structural physicochemistry of biomolecules based on the oligosaccharide-metal complex are presented in this publication.

Different biometals (Cu, Fe, Co, Zn) complexes with inulin, pullulan and dextran oligosaccharides, as well as reduced or oxidised derivatives, have been analyzed by IR spectroscopy. Spectra-structure correlations of the complexes have been performed by using modern spectroscopic techniques: FTIR microspectroscopy, ATR-IR, LNT-IR and D_2O-FTIR. The techniques are applying in the structure analysis of polysaccharide complexes, as well as for the confirmation of suggested types of complex structure and for the testing of homogeneities of samples. Results of IR microspectroscopic investigation shows that structural form of complexes and metal content considerably depends of constitution and ligands conformation, degree of crystallity, polymerization, polydispersity, and linearity of macromolecules. Also, stability of the synthesized complexes, as well as their pharmacological effect, depends of these parameters. FTIR investigation of the complexes by D_2O isotopic exchange proved to be a very sensitive method for determining OH group coordination and is related to the hydrogen bond strength. Results of our investigations points to the complexes are crystal hydrate molecules. Correlation of physicochemical and spectroscopic investigations of these complexes, and structure of exopolysaccharide chain, are suggesting different model structures of the synthesized complexes.

FTIR spectra and microscopy images were obtained by using an FTIR microspectroscopy system, ATR–FTIR spectrometer Bruker Tensor-27 in conjunction with a FTIR Bruker Hyperion-1000/2000 microscopy attachment equipped with the 4× viewing objective (objective magnification 4×, visible magnification 57×) and 15×IR Schwarzschild objective (objective magnification 15×, visible magnification 215×). The standard detector, a 250 μm liquid nitrogen cooled, mid-band mercury–cadmium–telluride (MCT) detector (ATR objective GMBH, Germany) with preamplifier, with the range of the IR spectrum from 7000 to 400 cm^{-1} was used. The spectra were measured with 2 cm^{-1} resolution and 200 scans co-addition. The spectrometer was linked to a PC equipped with Bruker OPUS software to allow the automated collection of IR spectra. The measurements were conducted in the reflection mode. In the region from 4000–400 cm^{-1} all spectra were Interactive polynomials baseline corrected and area normalized. A Kubelka–Munk arithmetic method was applied to enhance the resolution in this spectral region. Deconvoluted spectra were smoothed by the 40 point Fourier filter method. The IR spectra were imported to GRAMS/AI 7 (Thermo Galactic, USA) for peak area integration.

Thus, various tests can be performed by the Bruker Hyperion microscope, such as transmission, reflection, polarized, and ATR–IR measurements, the linear scan and mapping techniques in terms of software, and optic video technology for true video analysis. In addition, spatial resolution IR spectra and functional group imaging can also be acquired and analyzed. For measuring IR spectra by FTIR microscopy accurately, several primary parameters in the operation need to be selected and set first, which include aperture sizes, number of scans, resolution, velocity of motional mirror, and sampling background.

2. Character and spectroscopy of bioinorganic compound

Metal ions in biological systems are divided into two classes (Nakamoto, 2009). The ions of first class (K, Na, Mg and Ca) are important in maintaining the structure of proteins by neutralizing negative charges of peptide chains and in controlling the function of cell membranes that selectively pass certain molecules. In the second class, ionic forms of Fe, Co, Cu, Zn, Mn, Mo, and so on exist in small to trace quantities, and are often incorporated into proteins (metalloproteins). The latter class is divided into two categories: (A type) transport and storage proteins and (B type) enzymes. Type A includes oxygen transport proteins such as hemoglobin (Fe), myoglobin (Fe), hemerythrin (Fe), and hemocyanin (Cu), electron transfer proteins such as cytochromes (Fe), iron–sulfur proteins (Fe), blue-copper proteins (Cu), and metal storage proteins such as ferritin (Fe) and ceruloplasmin (Cu). Type B includes hydrolases such as carboxypeptidase (Zn) and aminopeptidase (Zn, Mg), oxidoreductases such as oxidase (Fe, Cu, Mo) and nitrogenase (Mo, Fe), and isomerases such as vitamin B12 coenzyme (Co).

During the last decades laboratory (with animals) and clinical researches have shown that many pathologic states of a body are accompanied by statistically reliable disturbances in the metabolism of metals at the molecular and body levels. Any chronic disease, the cause of which has not yet been established, can be due to abnormalities in metal metabolism. The determination of the amount of biometals in the body is suggested as the earliest diagnostic test of diseases (Grigorieva et al., 1983).

The metals participating in metabolism can be divided into the following groups: a) inherent in a living body and involved in the sphere of essential biofunctions (Cu, Fe, Zn, Mn, Mo, Co, Mg, Ca, K, Na); b) introduced, often toxic, whose physiological role has not been fully elucidated and their presence in the body tissue and liquids is due to their abundance in nature and wide application by people (Al, Cr, Cd, Ni, Pb, *etc.*). For the first group of metals both positive and negative balances were detected in different pathologies, and for the second group, as a rule, only the positive balance was observed. One of the reasons for the abnormal accumulation and removal of metals from a human body may be the wide application of drugs in clinics and which, by their chemical nature, are good ligand–complexing agents (up to 80% of all used drugs). Using non-steroidal antiinflammatory compounds (HL) and a copper-containing blood enzyme, ceruloplasmin (CuCPL), the ligands (drugs) were shown to take away competitively the metals from metal-containing and metal-activating enzymes: CuCPL + HL = CuL + CPL. Such an interaction results in a "discomfort" of an enzyme system in the body which is indicative of a side effect of drugs, *i.e.* complexing agents. For some diseases the shifts in metal metabolism are specific: rheumatoid arthritis (−) Fe, Zn; (+) Cu, Al, Mn, Mo, Cr; atherosclerosis (−) Cr, Mn, Zn; cancerogenesis (−) Cu, Fe, Mg; (+) Zn, Mn; diabetis (−) Cu, Mn, Cr; (+) Zn; *etc*. The correction in the concentration of these metals results in a therapeutic effect. The complex compounds of biometals with different types of drugs are the most promising tool for introducing the required metal into the body. It has been established that the application of antiinflammatory agents as complexes with some biometals decreases their toxicity and increases and prolongs their therapeutic effect (chemico-therapeutic synergism); antiulcerogenic, cytotoxic and other helpful properties, unusual to non-complexed agents, appear.

To understand the roles of these metal ions in biological systems, it is first necessary to know the coordination chemistry (structure and bonding) of metal ions in their active sites.

Such information is difficult to obtain since these active sites are buried in a large and complex protein backbone. Although X-ray crystallography would be ideal for this purpose, its application is hampered by the difficulties in growing single crystals of large protein molecules and in analyzing diffraction data with high resolution. As will be discussed later, these difficulties have been overcome in some cases, and knowledge of precise geometries has made great contribution to our understandings of their biological functions in terms of molecular structure. In other cases where X-ray structural information is not available or definitive, a variety of physicochemical techniques have been employed to gain structural and bonding information about the metal and its environment. These include electronic, infrared, resonance Raman, ESR, NMR, ORD, CD, Moossbauer spectroscopy, EXAFS, and electrochemical, thermodynamic, and kinetic measurements.

Infrared spectroscopy has been used extensively for the study of bioinorganic compound. In some cases, however, the vibrations of interest may not be enhanced with sufficient intensity. Then, one must resort to IR spectroscopy, which exhibits all vibrations allowed by IR selection rules. It should be noted, however, that IR measurements in aqueous media are generally limited to the regions where water does not absorb strongly. Furthermore, it is often necessary to use difference techniques to cancel out interfering bands due to the solvent and some solute bands. In the following, we will review typical results to demonstrate the utility of vibrational spectroscopy in deducing structural and bonding information about large and complex bioinorganic molecules. Marked progress has been made in chemistry of the bioinorganic complexes where the active site is modeled by relatively simple coordination compounds. Thus, we compare vibrational spectra of biological molecules and their model systems whenever appropriate or necessary. Since biospectroscopy is one of the most exciting areas of modern research, the volume of literature on biological compounds is increasing explosively. It is clearly not possible to cover all important topics in a limited space. Several excellent monographs (Parker, 1983; Nakamoto & Czernuszewicz, 1993) and review articles cited in each section should be consulted for further information.

Infrared spectroscopy has been used particularly for the study of polysaccharide complexes with metal ions, especially the active sites of the ions in the complexes. FTIR spectroscopy opens up new possibilities for the fine structural analysis of polysaccharides and its derivatives, the establishment of the type of bonding between the elementary links and their rotational isomerism. Weak intermolecular interactions have a significant influence on the specifically valuable properties of biological molecules and polymer compounds. We had to restrict ourselves to a few examples of wide potentialities of the method of FTIR spectroscopy in investigating the relationships between the structure and the properties of extracellular polysaccharides and its complexes with different metal ions.

3. Copper(II) ion and its significance

Copper(II) ion is a biologically active, essential ion, creating ability and positive redox potential allow participation in biological transport reactions. Cu(II) complexes possess a wide range of biological activity and are among the most potent antiviral, antitumor and antiinflammatory agents (Vosburg & Cooper, 1941). On the other hand, condensed triazoles exhibit a range of pharamacological activities such as mitotic (Jackson & Polaya,1951), hypotensive (Walker et al., 1951), CNS stimulant (Lepetil, 1975), antiflammatory (Hardtmann & Kathawala, 1977) and analgesic activites (Kathawala, 1974; Clark et al., 1997).

Copper(II) ion is an essential component of several enzymes such as ceruloplasmin, cytochrome C oxidase, lysil oxidase, superoxid dismutase and tyrosinase that are required to maintain the host homeostasis (Platonova et al., 2004; Gaiduk et al., 2009). At the same time, copper ions can be involved in the reactions producing active radicals, which affect the structure of all types of biomolecules. Therefore, it is not suprising that cells lack free copper ions, while their safe transfer is realized by a special system, metabolic copper system, some genes of which have been recently cloned.

Copper has an important role in the metabolism and transition of iron in the body. Microcytic hypochromic anemia is one of the outcomes of copper deficiency. There is a great number of hypocupremical drugs used commercially nowadays. As active substance, these drugs contain $CuSO_4$ (Expert group, 2002). In the literature is known that metal complex with polysaccharides and their derivatives are of growing importance in medicine and pharmacy. New blood substitutes with hemostimulating and antianemic function, which are complexes of dextran and pullulan with Cu(II) ion differ from the existing analogues in good bio- and hemocompatibility and more pronounced and prolonged action (Klimovich et al., 1998; Gapanovich et al., 1998). These complexes are very stable during prolonged storage and are not toxic. Copper(II) ion is used in the treatment of microcytic hypochromic anemia. It is apsorbed from the lower part gastrointestinal tract. This active pharmaceutical compound has a repetitive dose schedule (0.6-2 mg daily).

4. Bioactive copper-pullulan complex

Pullulan is a linear exopolysaccharide of α-D-glucopyranose that is often described as a α-(1→6) linked polymer of maltotriose subunits. This unique linkage pattern gives pullulan with distinctive physical properties. A number of potential applications have been reported for this biopolymer as a result of its good film-forming properties; pullulan can form thin films which are transparent, oil resistant and impermeable to oxygen. Pullulan may be used as a coating and packaging material, as a sizing agent for paper, as a starch replacer in low-calorie food formulations, in cosmetic emulsions, and in other industrial and medicinal applications (Deshpande et al., 1992). Pullulan is derivatized easily to control its solubility or provide reactive groups. Consequently, pullulan and its derivatives have numerous potential food, pharmaceutical, and industrial applications.

Bernier isolated water-soluble polysaccharides from the cultures of *Aureobasidium pullulans* and reported that α-D-glucopyranose is the major product of acid hydrolysis (Bernier, 1958). Based on the positive optical rotation and IR spectrum of pullulan was concluded that the polymer is a α-glucan in which α-(1→4) linkages predominate (Bender et al., 1959). Subsequent studies using IR, periodate oxidation, and methylation analysis established that pullulan is essentially a linear glucan containing α-(1→4) and α-(1→6) linkages in a ratio of 2:1 (Sowa et al., 1963). Partial acid hydrolysates of pullulan include isomaltose, maltose, panose, and isopanose (Leathers, 2003). The discovery of the enzyme pullulanase provided a critical tool for the analysis of the structure of pullulan (Wallenfels et al., 1961). Pullulanase specifically hydrolyzes the α-(1→6) linkages of pullulan and converts the polymer almost quantitatively to maltotriose (Wallenfels et al., 1965). Based on this result, pullulan is frequently described as a polymer of α-(1→6) linked maltotriose subunits (Fig. 1).

However, pullulan can also be viewed as a polymer of panose or isopanose subunits, which may reflect the biosynthetic origins of the molecule more accurately. Indeed, a number of enzymes that produce panose or isopanose from pullulan have been described since. Catley

(Catley et al., 1970) established that pullulan contains maltotetraose subunits (Fig. 2) in addition to the predominant maltotriose subunits. The frequency of maltotetraose subunits appears to vary on a strain-specific basis, from about 1% to 7% of total residues (Catley et al., 1986). The evidence suggests that maltotetraose subunits are distributed randomly throughout the molecule (Carolan et al., 1983). Unlike the maltotriose subunits in pullulan, maltotetraose residues are substrates for many α-amylases, and it has been proposed that hydrolysis of pullulan at these sites accounts for the decrease in molecular weight commonly observed in late cultures.

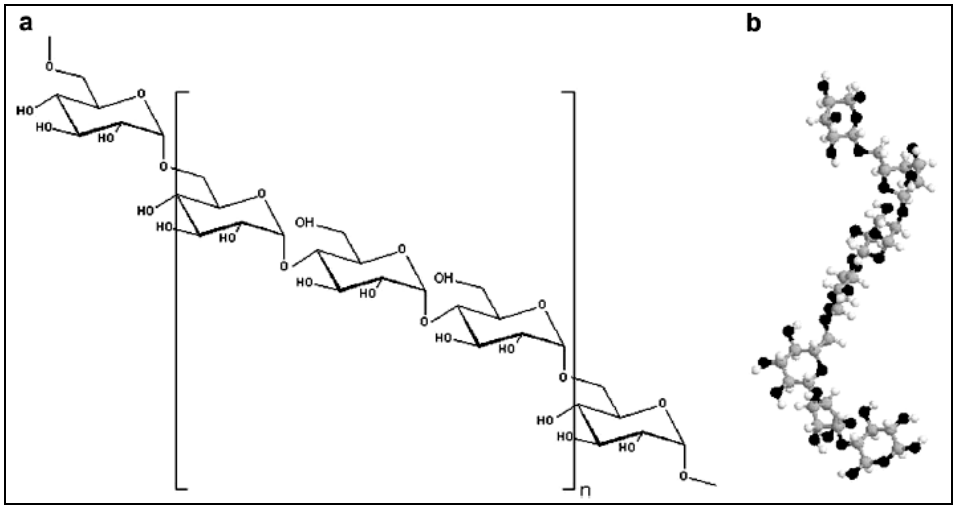


Fig. 1. Molecular structure of a representative portion of pullulan, illustrating the primary structure of repeating linkages: (a) 2D model, (b) 3D model stick and ball

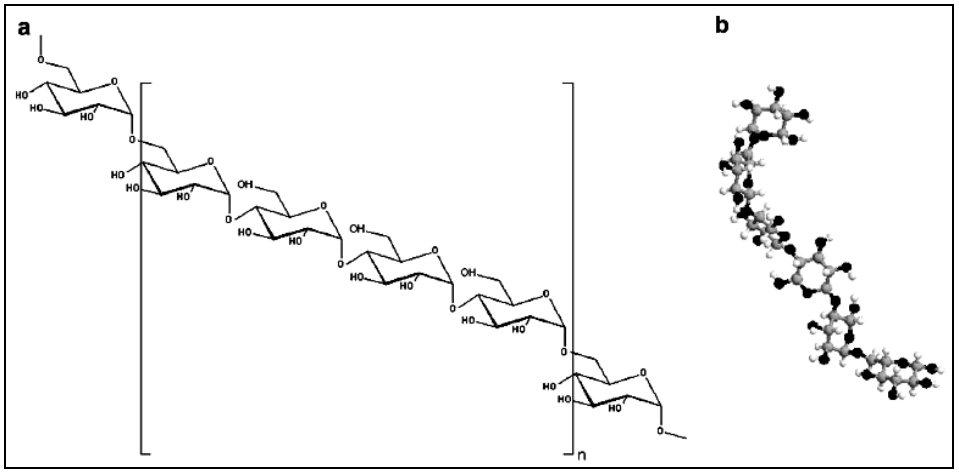


Fig. 2. Molecular structure of the secondary (minor) repeating structure of pullulan, occurring in about 1–7% of total linkage subunits: (a) 2D model, (b) 3D model stick and ball

Many types of carbohydrate derivatives (reduced or oxidized) have been synthesized for biomedical applications. In addition, polysaccharides such as chitin (Tanodekaew et al., 2004), chitosan (Pan et al., 2003; Yin et al., 2003), heparin (Ishihara et al., 2003; Kweon et al., 2003), alginate (Leonard et al., 2003; Perets et al., 2003), inulin (Nikolic & Cakic, 2007), dextran (London, 2004; Serizawa et al., 2003; Lawrence et al., 1997) and pullulan (Ilic et al., 2002; Kim et al., 2003; Nikolic at al., 2002) have been derivatized for biomedical applications. Pullulan is a polysaccharide that has been used in a drug delivery because of its solubility and biocompatibility. In addition, although the polysaccharides have many ionic groups, both anionic and cationic, pullulan is nonionic (Shingel, 2004).
Reduced low-molar pullulan (RLMP), was chosen as a new material for complexing, and the subsequent interactions with Cu(II) ions were investigated. The complexing process begins in a weak alkaline solution (pH > 7), and involves OH groups in C(2) and C(3) or C(6) pullulan monomer units (α-D-glucopyranose). Complexes of Cu(II) ion with reduced low-molar pullulan were synthesized in the water solutions, at the boiling temperature and at different pH values, ranging from 7.5–12. Cu(II) complexes were prepared from sodium salts, and investigated in the solid state. Fourier transform infrared spectroscopic data of synthesized complexes are rare in literature. FTIR spectroscopic characterization is now widely used to study the composition of the complex carbohydrate systems, the molecular interactions, a molecular orientation and conformational transitions of polysaccharides (Zhbankov, 1972; Panov et al., 1976; Panov & Zhbankov, 1988; Shingel, 2002; Zhbankov et al. 1997). The major goal of this section is to use different FTIR spectroscopic techniques (FTIR, LNT-FTIR, ATR-FTIR, and FTIR microspectroscopy) as the main tools to verify the conformation and the structure of this type of ligand around the Cu(II) ions.

Experimental.

Pullulan of average molar mass 2 x 10^5 g mol^{-1} and reduced low-molar pullulan of average molar mass 6000 g mol^{-1} was obtained from PCI ''Zdravlje Actavis Co." (Leskovac, Serbia). $CuCl_2$ x $2H_2O$ was purchased from Merck (Darmstadt, Germany). Cu(II) complex synthesis with RLMP have been described in detail by Nikolic (Nikolic et al., 2008). For FTIR sample preparation the KBr pastille method was used. Fine pulverized, water-free samples (1 mg) were mixed with potassium bromide (150 mg, Merck) stored at 80 ^{0}C for 6 h, and then pressed at 200 MPa to obtain a transparent pellet. The reference measurement was performed with pure KBr. The dryness of the pastille was controlled by the band at ca. 1640 cm^{-1}, which is associated with the deformation vibrations of the O-H bond from water molecules (Nikolic et al., 1996; Bellamy, 1954).
The FTIR spectra as an average of 40 scans were recorded at room (298 K) and liquid-nitrogen (77 K) temperature on a BOMEM MB-100 FTIR spectrometer (Hartmann & Braun, Canada) equipped with a standard DTGS/KBr detector in the range of 4000–400 cm^{-1} with a resolution of 2 cm^{-1} by the Win–Bomem Easy software. The spectrometer was purged with dry N_2. A Specac P/N 21525 variable-temperature cell was used for the LNT measurements. In the region all spectra were baseline-corrected and area-normalized. A Fourier self-deconvolution based on the Griffiths/Pariente method was applied to enhance the resolution in a spectral region of 4000–400 cm^{-1}. A gamma factor of 12 corresponding to a peak width of 24 cm^{-1} was used. Deconvoluted spectra were smoothed by the 30-point Savitzky–Golay filter method.
In addition, FTIR microspectroscopy system, ATR-FTIR spectrometer Bruker Tensor-27 in conjunction with a FTIR Bruker Hyperion-1000/2000 microscopy attachment equipped with

a 15x objective and a 250 μm liquid-nitrogen cooled, narrow-band mercury–cadmium–telluride (MCT) detector (ATR objective GMBH, Germany) with the range of the IR spectrum from 4000 to 400 cm^{-1} was used in this analysis. The spectra were measured with 4 cm^{-1} resolution and 320 scans co-addition. The measurements were conducted in the reflection mode. In the region from 4000–400 cm^{-1} all spectra were Interactive polynomials baseline-corrected and area-normalized. A Kubelka/Munk arithmetic method was applied to enhance the resolution in this spectral region. Deconvoluted spectra were smoothed by the 40-point Fourier filter method.

Results and discussion.

The FTIR spectra of the RLMP and the synthesized Cu(II) complexes (Fig. 3) contain following characteristic bands: ν(O-H) 3400 cm^{-1}, ν(C-H) 2930 cm^{-1}, δ(HOH) 1640 cm^{-1}, δ(C-H) 1450 and 1345 cm^{-1}, δ(O-H) 1420 cm^{-1}, a complex band ν(C-O) and ν(C-C) 1200–1000 cm^{-1}, γ(C-H) 1000–700 cm^{-1}. Between FTIR spectrum of RLMP and FTIR spectra of the synthesized complex on the different pH there is a clear difference in the area of vibrations of all types of OH groups and molecules H_2O (Fig. 3). That is, in the spectrum RLMP has found the wide intensive band on the approx. 3400 cm^{-1} which is the result of valent vibrations OH groups and valent vibration of H_2O constitutional molecules. The band on the 1640 cm^{-1} is the result δ(HOH) (Nikolic et al., 1996; Bellamy, 1954; Nikolic et al., 2007; Nikolic et al., 2008).

The appearance of the spectrum in this region is different, as expected. In the spectrum of the complex which is synthesized on the pH 7.5 (Fig. 3) the centroid of this band is shifted, and decreased temperatures provoke a clear separation of two bands the frequency which is 3378 cm^{-1} and 3246 cm^{-1} (data from LNT-FTIR). By the complex which is synthesized on the pH 8 the frequencies of these bands have been 3453 cm^{-1} and 3333 cm^{-1}, 3458 cm^{-1} and 3348 cm^{-1} on the pH 10 and 3389 cm^{-1} and 3355 cm^{-1} on the pH 12. These bands are sensitive at the decreasing temperatures so those, according to this criterion, need attribute ν(OH) vibrations. These changes in ν(OH) region are results of complexing i.e. the deprotonation of the RLMP ligand OH group, most likely of different surroundings in the first coordination sphere of the Cu(II) ion. Exactly let us say we know that by the complexing Cu(II) ion with dextran (Dex) in the dependence on the pH form different types of the complex (pH 8: $Cu(II)(Dex)_2(H_2O)_2$, pH 10: $Cu(II)(Dex)_2(H_2O)(OH)$, pH 12: $Cu(II)(Dex)_2(OH)_2$) and the spectral picture in this region is very similar (Nikolic et al., 2008). If in the case of the complex with RLMP would form similar complexes, bands in this region would need the attribute valent vibrations of the OH group and coordinate molecules H_2O by the complex which is synthesized on the pH 7.5 with regard to the OH ligand group and the OH group in the first coordination sphere of the Cu(II) ion by the complex on the pH > 10.

In the spectrum presented in Fig. 4a, of the complex which synthesized on the pH 7.5, one from two previous quoted bands would originate from ν(HOH), whose correct position could not be established and this is probably a low-frequent band 3246 cm^{-1} the absence of which in the spectrum of the complex was synthesized on the pH 12. In the area of δ(HOH) vibrations unlike RLMP where spectrum has only one band on the 1645 cm^{-1}, in the spectrum of the complex which synthesized on the pH 7.5 in the area of δ(HOH) vibrations have two bands (1657 and 1642 cm^{-1}) which points to two different types of H_2O molecules (Fig. 4b). The higher frequency band 1657 cm^{-1} with the increasing pH diminishes the intensity. In other words the complex on the pH 10 and pH 12 is absent (Fig. 4b).

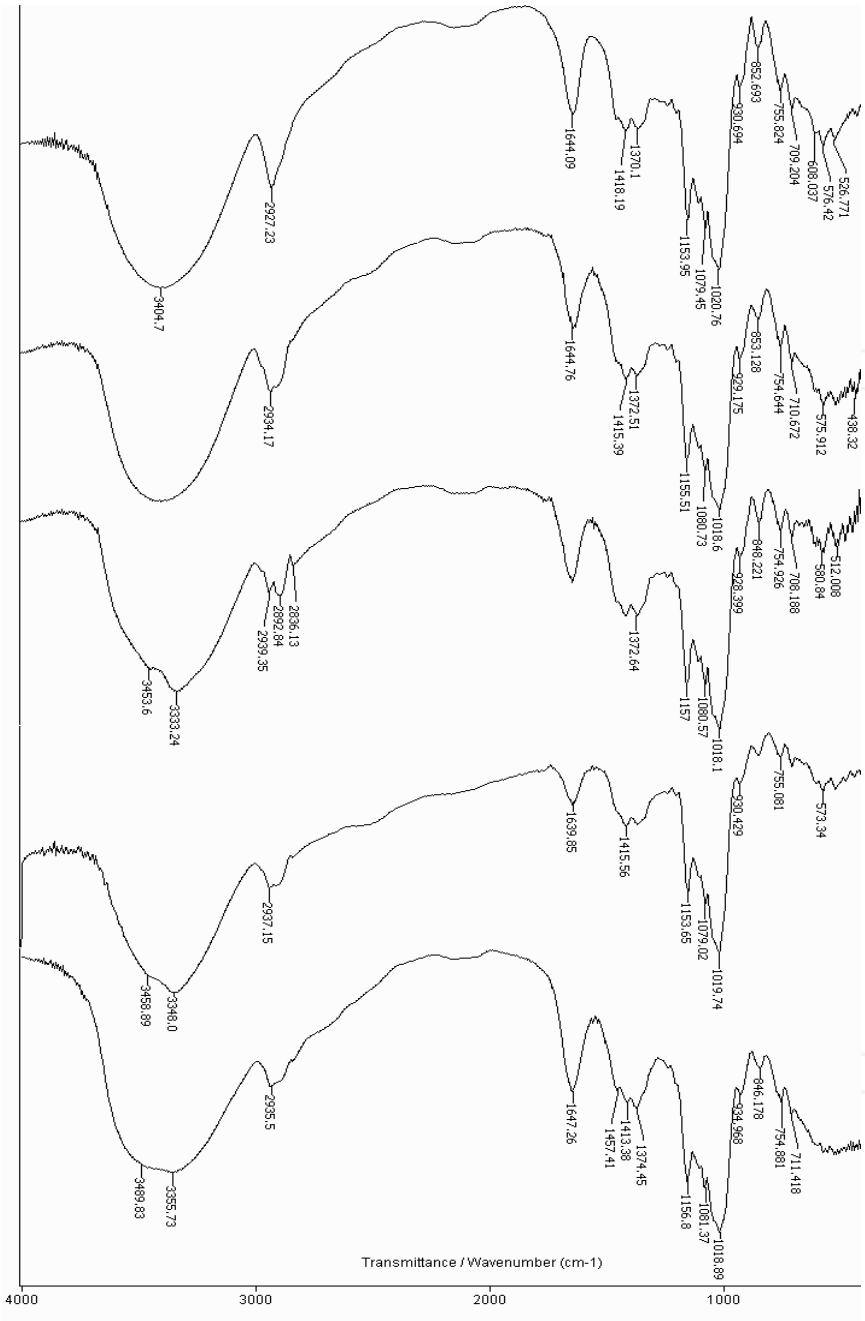


Fig. 3. FTIR spectra of RLMP, Mw = 6000 g mol^{-1} (1) and Cu(II) complexes with RLMP synthesized at boiling point and pH 7.5 (2), 8.0 (3), 10.0 (4) and 12.0 (5)

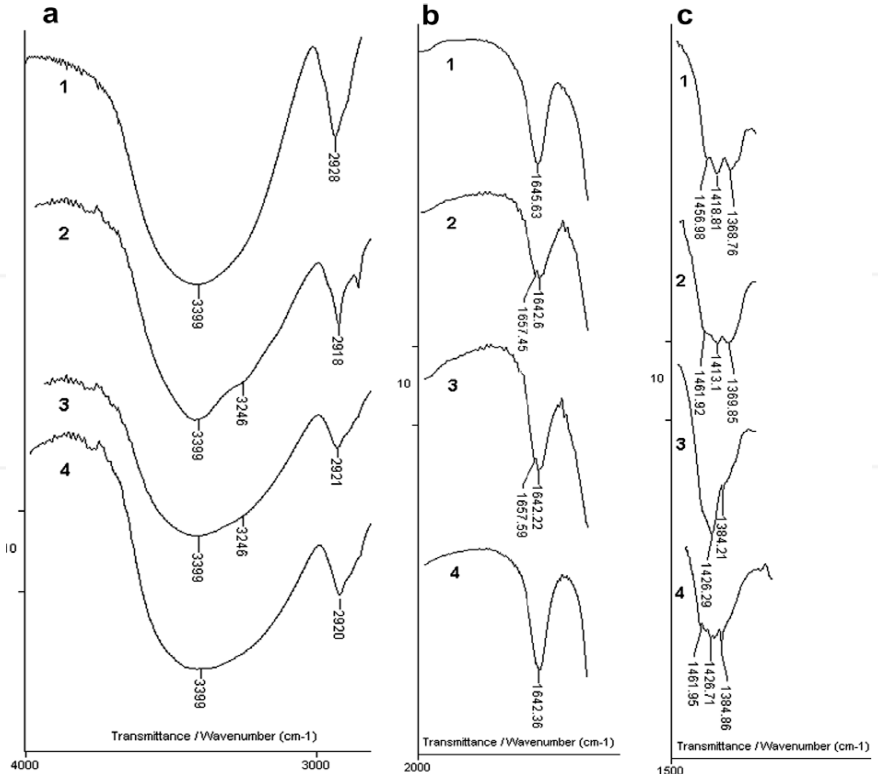


Fig. 4. Stretching (a), bending in plane (b) and bending in plane OH plane (c) region from LNT-FTIR spectra of RLMP (1) and synthesized Cu(II)–RLMP complexes at boiling point and pH 8.0 (2), 10.0 (3), 12.0 (4)

The spectral picture favors the suggested structure in Fig. 5 (type I). The complex which is synthesized on the pH 10 and pH 12 intensity bands from ν(HOH) and δ(HOH) diminishes with increasing pH. Expect the appearance band from δ(OH) from the first coordination sphere of Cu(II) ion whose intensity with increasing of pH growth; this band in the spectrum of the complex which synthesized on the pH 7.5 is absent. In addition to this, the fact is that in the area of δ(HOH) by the complex which is synthesized on the pH 10 and pH 12 only one band exists, which points to one type of H_2O molecule (like by RLMP) and band δ(OH) on the 1384 cm^{-1} (data from LNT-FTIR) (Fig. 4c).

Spectroscopic FTIR study in a particular region of O-H (3400 and 1420 cm^{-1}) and C-H (2900, 1460, and 1350 cm^{-1}) vibrations indicates different binding between the central metal ion and ligand, depending on pH and metal contents. Water protons take part in the formation of relatively weak hydrogen bonds (Nikolic et al., 1996; Bellamy, 1954; Nikolic et al., 2007; Nikolic et al., 2008). In the 1200–1000 cm^{-1} region, the spectra of RLMP and the complex comprise a number of highly fused bands. The enhancement of the resolution by using a Fourier self-deconvolution allows bands to be more accurately detected. The main bands found in the deconvoluted spectra of RLMP and the complex at ca. 1154, 1108 1079, 1042, and 1019 cm^{-1} are due to coupled valent vibrations of the C-O and C-C bonds and

deformational vibrations of the CCH, COH, and HCO bonds. The band at about 1150 cm^{-1} has been assigned to valent vibrations of the C-O-C bond and glycosidic bridge. The broad peak at 1108 cm^{-1} should most likely be ascribed to the vibration of the C-O bond at the C(4) position of the glucopyranose units (Kacurakova et al., 1996). Complex vibrations involving the stretching of the C(6)-O(6) bond with participation of the deformational vibrations of the C(4)-C(5) bond result in the appearance of a band at 1079 cm^{-1} (Sivchik et al., 1979; Nikonenko et al., 2005; Zhbankov et al. 2005; [a]Zhbankov et al., 2003; [b]Zhbankov et al., 2003; Zhbankov et al., 2000). In the spectra of Cu(II) complex with RLMP band at 1079 cm^{-1} is less pronounced than in the spectra of RLMP. In the case of pullulan complexes, part of C(6) atoms participate in the formation of the C(6)-O-Cu(II) linkages; as a result, the band intensity at 1079 cm^{-1} for the Cu(II) complex with RLMP is reduced more than in the case of RLMP. The band at 1079 cm^{-1} in the FTIR spectra of RLMP is attributed to the antisymmetric stretching vibration of C(6)-O-C(1) glycosidic bridge. These findings suggest that the 1079 cm^{-1} band for the Cu(II) complexes with RLMP can be considered as a characteristic for the type of interunit links and for the Ligand–Metal [C(6)-O-Cu(II)] linkage (Shingel, 2002; Zhbankov et al., 2000; [a]Mitic et al., 2008; [b]Mitic et al., 2008).

In the case of the Cu(II)–RLMP complexes O-H groups participate in the formation of the Cu(II)–RLMP linkages. As a result, the band frequency at 3404 cm^{-1} for ν(O-H) vibrations in RLMP is reduced to approximately 3340 cm^{-1} (Fig. 3) in Cu(II)–RLMP. These findings also suggest that the band can be considered as characteristic for the type of Metal–Ligand links. The band at about 1042 and 1019 cm^{-1} found for polysaccharide in the spectra of RLMP and the complex were shown to relate to the crystalline and amorphous phases, respectively (Smits et al., 1998). The changes in intensity of these bands are strongly associated with the alterations in the macromolecular order. These bands in the spectra of RLMP and the complex can be responsible for more and less ordered structures, respectively. The major attention was focused on the bands in the 1160–1010 cm^{-1} region because the absorbance pattern due to ring vibrations in this spectral range is known to be individual for each carbohydrate structure.

Moreover, we attempted to obtain the information about the conformations of these macromolecules in a solvent exhibiting a different influence on the system of intra- and intermolecular interactions. Special interest in the IR range for structural investigation is from 1000 to 700 cm^{-1}. In the spectra of RLMP and the complexes, bands of negligible intensity are found in the region (950, 916, 860, 760 cm^{-1}). According to the normal coordinate treatment on the RLMP model, these bands are interpreted as due to mixed CCH deformation vibrations coupled with CCO, OCO, and COC bending (Buslov et al., 1998; Zhbankov et al., 1997; Zhbankov, 1992; Zhbankov and Avsenev, 1984; Kiselev et al., 1977). Both the number and frequencies of the bands in the IR range depend on the conformation of the D-glucopyranose units. It is well known that the glucopyranose units exist in six different typical conformations (1C, C1, 1B, B1, 3B, and B3) (Panov & Zhbankov, 1976; Komar et al., 1968). The similarities of the γ(C-H) range indicate that there is no difference in the conformation of the glucopyranose unit in the RLMP and complex molecules, and they probably exhibit C1 chair conformation (916 and 850 cm^{-1}).

It appears that the intensity of the 996 cm^{-1} band in the pullulan spectra may indicate the extent of the interchain association. The band at 950 cm^{-1} belongs to the structure-sensitive region, and together with the band at 935 cm^{-1}, characterizes the type of interunit bonds and angles. The band at 935 cm^{-1} was recently used to discover the co-existence of α-(1→6) and

α-(1→4) glycosidic linkages in the pullulan structure (Shingel, 2002; Zhbankov et al., 2000). A decrease of the band at 950 cm^{-1} indirectly confirms an occurrence of the conformational transitions in polysaccharide systems owing to rotational isomerism of pyranose rings about the glycosidic bond. For pullulans the band at 900 cm^{-1} described α-(1→6) linkages. α-(1→4) linkages were observed at 925 cm^{-1} (Shingel, 2002). Ring deformations and scaffold vibrations were observed at 710, 660, 600, 570, and 525 cm^{-1}. In the experiment on the influence of the medium pH on a binding Cu(II) ion with different polysaccharides (Mitic et al., 2007; Norkus et al., 2002; Norkus et al., 2004), there is a possibility of gradual complexing, where their reforming starts at pH 8. Degradation of the Cu(II)–RLMP complex begins at pH values higher than 12. The Cu(II) ions form three different types of complexes with the deprotonated monomeric RLMP unit. Different structural models of the Cu(II)–RLMP complexes of tetragonal distorted O_h coordination in the function of pH synthesis pH 7–8 (type I), pH 8–10 (type II), and pH 10–12 (type III) are given in Fig. 5.

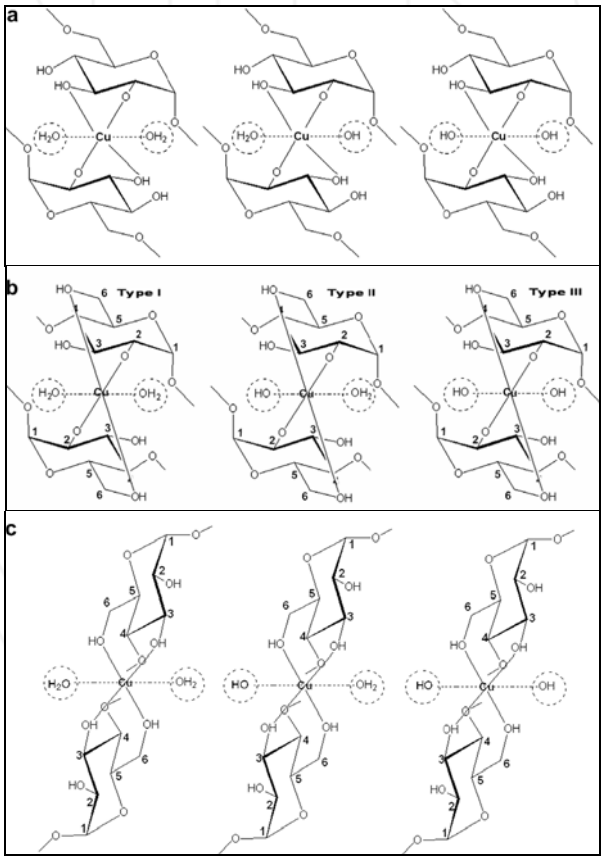


Fig. 5. Structure model of Cu(II)–RLMP complexes, with six O-donor atoms in tetragonal distorted O_h environment of Cu(II) ions, with participation of: (a) C-2 and C-3 ligand OH groups, (b) C-2 and C-6 ligand OH groups, (c) C-3 and C-6 ligand OH groups

The reactivity of the RLMP depends primarily on the reactivity of the secondary, equatorially oriented hydroxyl groups (OH-2, OH-3, OH-4 and OH-6). The contents of the primary O-H groups in RLMP are slightly increased at lower Mw (about 2%). The reactivity of the polysaccharide C-atoms was determined by 13C NMR spectroscopy for pullulan it was C(6)>C(3)>C(2)> C(4) (Mahner et al., 2001). Carbohydrates without anchoring donor groups form a very weak complex with Cu(II) ion in an aqueous solution. The availability of more than one anchoring group can, however, prevent the coordination of the alcoholic O-H groups fulfilling the coordination sphere of the metal ion. The metal interaction with the set of the non-deprotonated OAH groups increases the complex stability. Mainly, the complexes were shown to form, in different protonation states, the deprotonation processes starting from pH 7 (Norkus et al., 2002). After deprotonation of one or more alcoholic O-H groups, the Cu(II) ion complexes having anionic character are usually very stable. In some cases, the formation of various amounts of alkoxo or hydroxo bridged dimeric (or oligomeric) species can be detected. The disugars bound the metal ions less efficiently than the monomeric units, while the trimetric ligands can probably simultaneously use terminal subunits to coordinate the metal ion. In aqueous solution, the RLMP complexes are formed by the displacement of the H_2O molecules from the first coordination sphere of Cu(II) ion by the alcoholic OAH groups. In general, it seems to be true that at least three O-H groups in a favorable steric arrangement are required for the complex formation (Gyurcsik & Nagy, 2000). The general leading rule is that the sugars in pyranose form (in RLMP C1-chair conformation) contain an equatorial–equatorial–equatorial (eq–eq–eq) sequence of three adjacent hydroxyl groups. The possible coordination sites are depicted on the model given in Fig. 5.

The characterization of metal ion coordination equilibrium of polyalcohols and other sugar-type ligands, containing alcoholic and aldehyde (or ketone) oxygen donor atoms, is difficult due to the low stability of the complexes in neutral or acidic aqueous solutions (Gyurcsik & Nagy, 2000; Nagy et al., 2003). The low electron densities on these donor oxygens cause the situation, in spite of their relatively large number in one ligand molecule, that they do not readily substitute the water molecules bonded in the first coordination sphere of the metal ions. With increasing pH, however, the hydrolysis of some metal ions prevents the coordination of the organic ligands. Thus, complex formation can only be expected in strongly alkaline solutions after deprotonation of alcoholic hydroxyl groups. The fact that in solutions of carbohydrates the species are in anomeric and conformational equilibrium and the isomers interact in different ways with metal ions makes the studies even more complicated. Any shift in the above equilibrium due to the metal ion coordination, thereby resulting in the changes in the fraction of the isomers having suitable positioned sequences of alcoholic hydroxyl groups in the total concentration of the ligand, will also influence the complex stability.

The methods, such as FTIR, NMR, ESR, X-ray and UV–Vis made it possible to assign the binding hydroxyl or other groups and also to characterize the metal ion coordination of carbohydrates monitoring the ligand conformation or/and configuration changes forced by the complexation processes. FTIR spectra of the Cu(II)–RLMP complexes were recorded on room (RT) and on low nitrogen temperatures (LNT) in order to detect bands which are sensitive to the reduction temperatures respective bands which have originated from vibrations of all types OH groups and H_2O molecules. In Fig. 6 RT-IR and LNT-IR spectra of the complex synthesized on the pH 7.5 have been presented in the comparison with RLMP.

FTIR spectra show the correlation between O-H stretch frequency and the hydrogen bond strength (3400–3200 cm^{-1}) predict a red shift of the bonded O-H stretching band on cooling (from 3350 to 3246 cm^{-1}) (Fig. 6). It is expected that the non-interacting O-H group (at 3378 cm^{-1}) is much less sensitive to cooling and consequently will show smaller red shifts. The red shift of the band is an indication of the involvement of the appropriate O-H proton in a weak hydrogen bond.

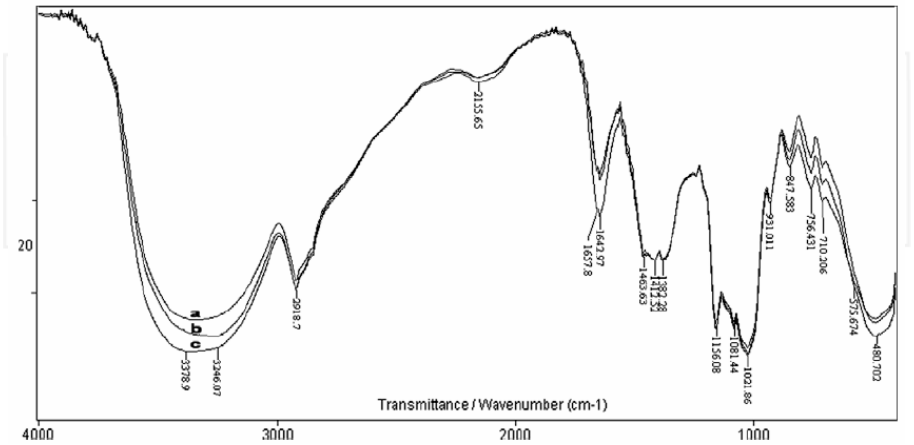


Fig. 6. Stretching and bending region from RT (a) and LNT-FTIR (b and c) spectra of Cu(II)–RLMP complex (type I, pH 7–8) (a – 298 K, b – 173 K, c – 77 K)

In the LNT-FTIR spectra of the complex was synthesized on pH 7.5, two bands (3378 and 3246 cm^{-1}) are found in the region of ν(O-H) vibrations (Fig. 6). In the low-frequency region on LNT-FTIR spectra were presented in Fig. 6, sensitive on the reduction of temperature bands in the γ(OH) bending region from librations of coordinated water molecules on frequencies 847 cm^{-1} and 756 cm^{-1}, show blue shift on cooling. The librations of the O-H group in this region of the complex synthesized on pH 12 were much less sensitive to cooling. The observation allowed one to suggest that the most probable water molecules are coordinated around Cu(II) in the complex type I (Fig. 5). The number and shape of these bands implies that in complexes type III there is the displacement of H_2O molecules by the O-H groups in the first coordination sphere of the Cu(II) ion. The LNT-FTIR results confirm the structural models of complexes presented in Fig. 5. The results obtained from the structural studies of the investigated complexes were based on other spectroscopic techniques (Nikolic et al., 2005; Mitic et al., 2004; Nikolic et al., 2004; Nikolic et al., 2006; Bartkowiak et al., 1998; Cakic et al., 2008).

The changes in number, frequencies, intensity, and width of the FTIR bands in the particular region of ν(O-H) vibrations (3400 cm^{-1}), δ(C-H) vibrations (1500–1300 cm^{-1}) and ν(C-O) vibrations (1200–1000 cm^{-1}) (Fig. 3) were related to changes in the conformation and short-range interactions of the RLMP. Very important changes can be observed in the range of 1500–1300 cm^{-1} by detailed empirical analysis. Otherwise, the FTIR range is specific of bending vibrations of CH_2-OH groups (Fig. 7). Namely, the exchange position and intensity of complex bands can be registered in this range, where C-H and O-H bending vibrations from the CH_2-OH groups take part. The change of intensity on some bands was registered

only in synthesized Cu(II)–RLMP complexes. An approximate effect exists in the stretching of the FTIR range of C-H vibrations (3000–2800 cm^{-1}, Fig. 3). The appearance of bands at about 1460 cm^{-1} and 1370 cm^{-1} from δ(C-H) vibrations and the band at about 1420 cm^{-1} from δ(O-H) vibrations are characteristic for one of more possible positions of the CH_2-OH group, rotating around the C(5)-C(6) bond of the glucopyranose unit. The change of the angle between the methylene CH_2-OH group and the polysaccharide chain axes, consequently decreases the intensity of the appropriate IR bands (ν(C-H) and δ(C-H) vibration). The Cu(II) ions in solution have a possible influence on the rotation of CH_2-OH groups in the complexes (Cakic et al., 2004; Mitic et al., 2007; Nikolic et al., 2006).

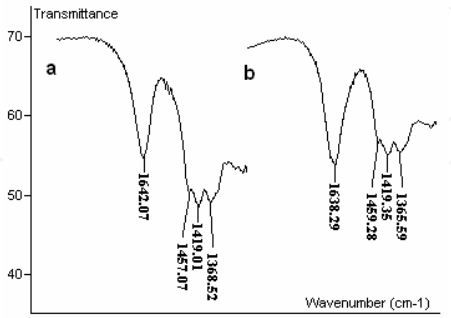


Fig. 7. Region of δ(CH) and δ(OH) CH_2OH grop vibrations in FTIR spectra of: (a) RLMP; (b) and Cu(II)–RLMP complexes

We also applied FTIR spectroscopy to determine spectral manifestation of the changes in the complex structure caused by recrystallization from D_2O and, thereby, complete the structural investigation of this modified polysaccharide. The FTIR spectra of the Cu(II) complexes with RLMP and recrystallized analogs from D_2O were analyzed in order to find the specific spectral peculiarities that allow one to obtain the information about the structure and the conformation of these macromolecules in solvents that exhibit different influences on the system of intra- and intermolecular interactions. No effect of the conformation change was observed for the recrystallized Cu(II)–RLMP complex, especially in the range of 1000–700 cm^{-1}. When the a-D-glucopyranose units with C1 chair conformations are present, the FTIR spectra exhibit one band in the region between 925–885 cm^{-1} and another one around 860–820 cm^{-1}, which are assigned to mixed CCH deformation vibrations (Shingel, 2002; Nikolic et al., 2008). The results allowed one to suggest a predominant crystalline form of the recrystallized Cu(II)–RLMP complexes.

Recently, FTIR spectroscopy was coupled with a microscope and a computer system, capable of microanalysis of minute samples by using a dedicated MCT detector. The resultant FTIR vibrational microspectroscopy can provide molecular information of samples with a high spatial resolution at microscopic level. Samples with microscopic size can be nondestructively analyzed by both vibrational microspectroscopies, particularly in the application of biomedical sciences (Kacurakova et al., 2001; Lin et al., 2007; Chiu et al., 2004; Nikolic et al., 2008). Thus, the use of vibrational microspectroscopy has extensively become a great potential over other spectroscopic techniques for noninvasive investigation of chemical components of ultrastructural samples (carbohydrates, lipids, proteins, nucleotides) (Mousia et al., 2001; Yu et al., 2005).

More recently, FTIR and/or Raman microspectroscopic imaging systems have also been developed for applying to biosciences (Gierlinger & Schwanninger, 2007; Chenery & Bowring, 2003). ATR-FTIR spectra may be simultaneously collected at a time in a stepwise manner from different areas of a sample. The absorbance ATR-FTIR spectra of Cu(II)–RLMP complex which was synthesized at pH 7.5 are shown in Fig. 8. The absorbance of a band corresponding to a specific chemical component may be plotted as a map. ATR-FTIR spectra were presented in Fig. 8(A–C) from different areas of Cu(II)–RLMP complex and show high homogeneity of the sample.

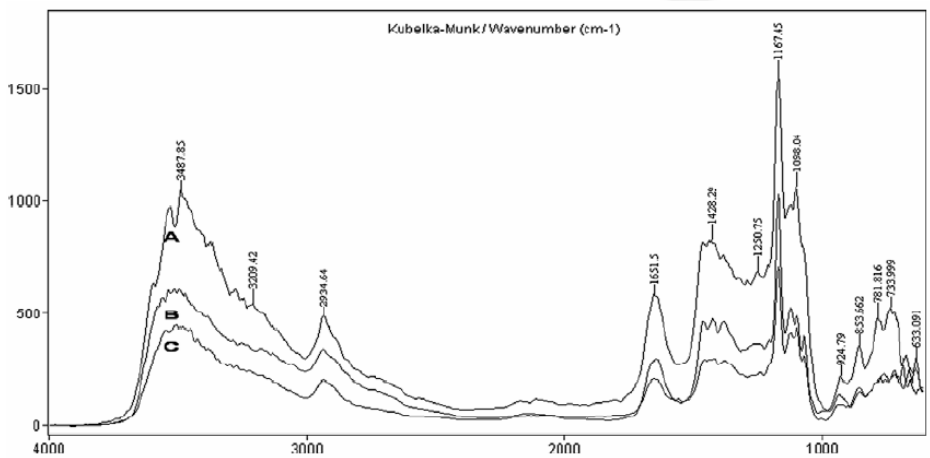


Fig. 8. ATR–FTIR spectra of Cu(II)–RLMP complex synthesized at pH 7.5 (type I) from different areas (A, B and C) of a sample

A new imaging capability has been established not only to image heterogeneous regions of the samples and simultaneously provide spectroscopic and spatial information, but also to show visually the concentrations of components and to highlight their effect from the three dimensional plot. The application of microscopic FTIR imaging system to the ligand RLMP and Cu(II)–RLMP complexes, were synthesized at pH 7.5–12, is shown in Fig. 9. FTIR microscopy images of ligand RLMP, as well as images of the synthesized Cu(II)–RLMP complexes differ which also indicates that the complexation process and the creation of coordination compounds took place. FTIR microscopy images confirmed that the changes in the intensity of the analyzed bands are strongly associated with the alterations in the macromolecular order. These bands in the spectra of the complexes can be responsible for more and less ordered structures, respectively (Fig. 9). The changes in color contour may show the content and distribution of copper, and polysaccharides in Cu(II)–RLMP samples.

Conclusions.

The complexing process begins in a weak alkali solution (pH > 7.5), and involves OH groups in C(2) and C(3) or C(6) pullulan monomer unit (α-D-glucopyranose). A part of FTIR spectra, in the range on 1000–700 cm^{-1} of Cu(II) ion complexes with RLMP, indicates no influence of complexing process on the conformation change of C1 glucopyranose units. The IR band d(HOH) at the frequency of 1640 cm^{-1} indicated the existence of water molecules in a complex structure.

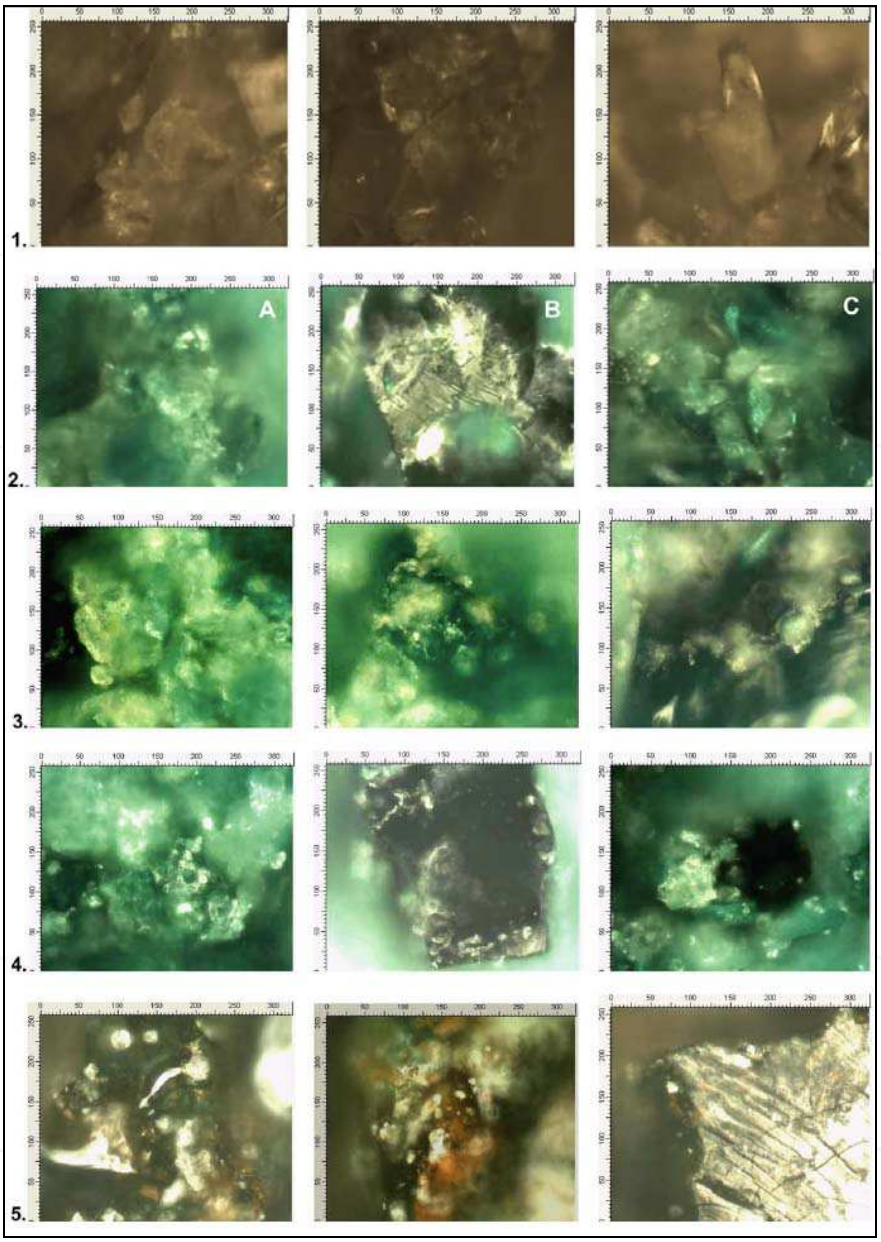


Fig. 9. FTIR microscopy images (250 μm x 300 μm) of RLMP, Mw = 6000 g mol^{-1} (1) and Cu(II) complexes with RLMP synthesized at boiling point and pH 7.5 (2), 8.0 (3), 10.0 (4), 12.0 (5)

From LNT-FTIR it follows that non-interacting OH group is insensitive to temperature variation whereas a bonded OH shows a significant red shift upon cooling. In the low-

frequency region on LNT-FTIR spectra, sensitive on the reduction of temperature bands in the γ(OH) bending region from librations of H_2O molecules on frequencies 847 cm^{-1} and 756 cm^{-1}, show blue shift on cooling.
Cu(II)–RLMP complexes are formed by the displacement of H_2O molecules from the first coordination sphere of Cu(II) ion by the OH groups. Copper(II) ion with RLMP unit (Glc) forms three different types of complex (pH 7–8: $Cu(II)(Glc)_2(H_2O)_2$, pH 8–10: $Cu(II)(Glc)_2(H_2O)(OH)$, pH 10–12: $Cu(II)(Glc)_2(OH)_2$).
The changes of the intensity on some bands were registered in RLMP complexes (in the ranges of a stretching vibration at about 2930 cm^{-1} and a bending vibration at about 1400 cm^{-1}). The bands are characteristic of one of more possible positions of the CH_2-OH group, rotating around C(5)-C(6) bond of the pullulan glucopyranose unit.
FTIR microscopy images shows more and less ordered structures of the Cu(II)–RLMP complexes. ATR-FTIR microspectroscopic data shows homogeneity of the Cu(II)–RLMP samples. The results of the FTIR spectroscopic study by different techniques allowed one to suggest a predominant crystalline form of Cu(II)–RLMP complexes.

5. Bioactive copper-dextran complex

Dextran $H(C_6H_{10}O_5)_xOH$ is a complex, branched glucan composed of chains of varying lengths (from 10 to 150 kDa). It is a polysaccharide similar to amylopectin (Belder, 1985). The straight chain consists of α-(1→6) glycosidic linkages between glucose molecules, while branches begin from α-(1→3) or α-(1→4) linkages (Fig. 10). Dextran is synthesized from sucrose by certain lactic-acid bacteria, the best-known being *Leuconostoc mesenteroides* and *Streptococcus mutans*. Dextran is an oral bacterial product that adheres to the teeth, creating a film called plaque; dental plaque is rich in dextrans. Dextran is also formed by the lactic acid bacterium *Lactobacillus brevis* to create the crystals of tibicos, or water kefir fermented beverage which supposedly has some health benefits. Dextran is freely soluble in water, methyl sulphoxide, formamide, ethylene glycol, glycerol, 4-methylmorpholine-4-oxide, and hexamethylphosphoramide. Some dextran fractions may adopt a certain degree of crystallinity and may only be brought into solution by strong heating.

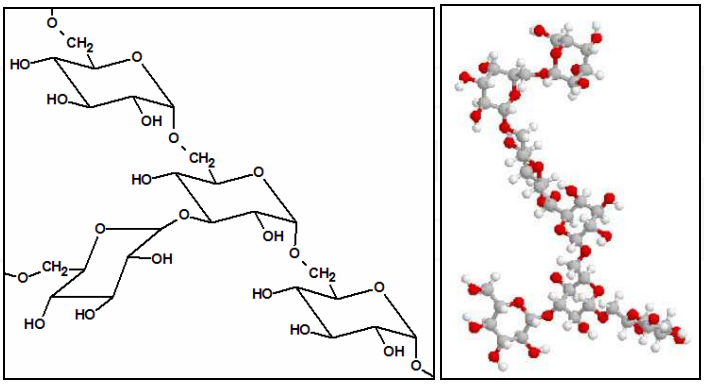


Fig. 10. Molecular structure of a dextran: (a) 2D model, (b) 3D model stick and ball

Dextran and its derivatives have been studied extensively (Zhbankov, 1972; Skornyakov & Komar, 1996). The Fourier-transform infrared spectra of a series of branched dextrans also

were examined. The IR spectra of low molar dextran have been investigated in the range between 4000 and 400 cm^{-1} (Nikolic et al., 1996). The FTIR spectrum of reduced low molar dextran is presented in Fig. 11. Information on the glucopyranosyl units conformation in the polysaccharide can be acquired in the 1000–700 cm^{-1} region in which we expect the deformational γ(CH) vibrations bands. In this particular region, as it can be seen in Fig. 11, at least two weak bands around 911 and another 850 cm^{-1} are observed, which is the proof for C-1 conformation of glucopyranoside units of dextran. In the δ(OH) range of IR spectra one band at about 1645 cm^{-1}, which is sensitive to deuteration, has appeared. The band originates from water molecules.

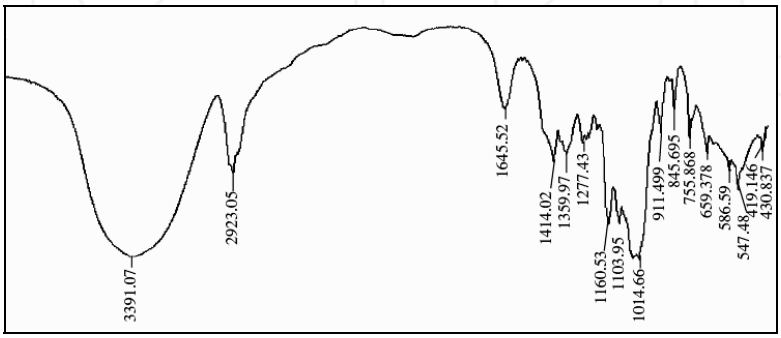


Fig. 11. FTIR spectra of reduced low molar dextran, Mw = 5000 g mol^{-1}

Dextran is used medicinally as an antithrombotic (anti-platelet), to reduce blood viscosity, and as a volume expander in anemia (Lewis et al, 2008). These agents are used commonly by microsurgeons to decrease vascular thrombosis. The antithrombotic effect of dextran is mediated through its binding of erythrocytes, platelets, and vascular endothelium, increasing their electronegativity and thus reducing erythrocyte aggregation and platelet adhesiveness. Dextran also increases blood sugar levels. Biological active polysaccharides dextran have possibility to binding different ions and metals in the solution and making of water-soluble complexes. These complexes have wide application in human medicine and veterine. These compounds have great importance in investigations today.

Copper, essential biometal for living organisms, is a hematopoetical active element of some metaloenzymes regulating the iron absorption in intestines, maintaining, at the same time, the iron in a reduced state and influencing the iron incorporation into hemoglobin (Lewis, 1995). The copper amount necessary is usually supplemented by a normal diet in both humans and animals. It is known that copper deficiency causes a number of pathological states. Complex compounds of Cu(II) ion are important for prevention and treatment of some anemia caused by iron deficiency.

The carbohydrate type compounds as ligands have been of a considerable interest. Simple sugars and their derivatives with reduced and oxidized groups form metal ion complexes of a various composition and stability. In both human and veterinary medicine commercial copper preparations based on polysaccharide dextran and its derivatives are used for such purpose (Ilic et al., 2003). According to literature data, dextran has the ability of complex formation with various biometals (Co, Zn, Ca and Mg) (Mitic et al., 2007; Cakic et al., 2006; Lugovaya et al., 1976; Gyurcsik & Nagy, 2000). Iron complexes with different polysaccharides have special importance, and they have been described in detail (Ilic et al., 2002;

Nikolic & Cakic, 2007; Pekic & Cvetkovic, 1988; Cakic & Nikolic, 2003; Nikolic et al., 2002). The interaction of Cu(II), Ni(II) and Fe(III) ions with dextran may be used for their speciation by ultrafiltration (Solpan & Sahan, 1993). Synthesis procedures for the complex formation of Cu(II) ion with polysaccharides, including dextran, are described in scientific literature (Nikolic et al., 2005; Mitic et al., 2004; Mitic et al., 2007).

In the section, we analyzed the IR spectra of Cu(II) ions complexes with reduced low molar dextran (RLMD). Fourier-transform IR spectra of dextran and its compounds with copper(II) ion, recorded at room temperature, were analyzed in order to obtain the information about the structure and the conformation of these polymer compounds. For IR sample preparation KBr pastille method was used. The IR spectra as an average of 40 scans were recorded at room (298 K) temperature on FTIR spectrometer BOMEM MB-100 (Hartman–Braun) equipped with a standard DTGS/KBr detector, in the range of 4000–400 cm^{-1} with the resolution of 2 cm^{-1}, by Win-Bomem Easy software. FTIR microspectroscopy system, ATR-FTIR spectrometer Bruker Tensor-27 in conjunction with a FTIR Bruker Hyperion-1000/2000 microscopy attachment equipped with a 15x objective and a 250 μm liquid-nitrogen cooled, narrow-band mercury–cadmium–telluride detector (ATR objective GMBH, Germany) was used also in this analysis.

Results and discussion.

The plan for the synthesis of the Cu(II) complex with reduced dextran has required a detailed analysis of the synthesis procedure; both from the aspect of the reaction conditions of the synthesis and from the aspect of obtaining the stable and commercially applicable preparation of the complex. The analysis of the synthesis of similar complexes has pointed to the necessity of defining the physicochemical properties of commercial preparations. By their correlation, the undesired effects can be eliminated and thus a considerably improved pharmacological effect of the complex.

The reactivity of the dextran primarily depends on the reactivity of the secondary, equatorially oriented hydroxyl-groups (OH-2, OH-3 and OH-4). The contents of the primary OH groups in dextran are slightly increased at lower molar masses (about 2%). As it is the case with the other glucans, the reactivity of the OH-2 group to the alkalizing reagents is higher than in the OH-3 and OH-4 groups. This is rationalized in the context of higher acidity of the OH-2 because of the proximity of anomeric centre (Gyurcsik & Nagy, 2000). When the OH-2 and OH-4 ionize, reactivity to OH-3 is reduced; however, the substitution on the OH-2 and OH-4 abolish this action and induce the successive increase in the reactivity of OH-3. On applying these generalizations, it is necessary to exercise some precaution, since the potency of the base may affect the relative and absolute reactivities of the hydroxyl. Through derivatization of the dextran to the reduced form the large number of activation centers with will be at the disposal to the copper ions for the purpose of binding the complex is expected. This creates the possibility of achieving the considerably larger stability of the synthesized complex as well as of their pharmacological effect.

For this reason, the choice and optimization of the low molar dextran in the capacity of the ligand have been made. Considering the importance of physicochemical parameters on the process and the synthesis results, the examination and optimization of ligands in relation to molar mass (*Mw*), as well as the reaction conditions of the synthesis (pH, *T* and *t*) were investigated and optimized (Mitic et al., 2007). The basic characteristics of synthesized Cu(II) complexes with RLMD are given in Table 1.

Sample	pH synth.	Complex color	Cu (%) (by AAS)	Water solubility
1	7.5	Light green	7.23	Soluble
2	7.5	Green	19.85	Very soluble
3	8.0	Green-blue	8.12	Soluble
4	10.0	Blue-green	8.20	Sparingly soluble
5	12.0	Dark blue	6.97	Slightly soluble
6	7.5	Dark green	4.08	Slightly soluble

Table 1. Characteristics of copper-dextran complexes with RLMD (M_w 5000 g/mol) as ligand, synthesized at boiling temperature

On the basis of obtained experimental results (Table 1), a favorable result of Cu(II)-RLMD complexes synthesis is obtained with dextran oligomers *M*w 5000, at boiling temperature and pH 7.5–8.0, within 7 min. Complexes obtained at pH > 8 present unfavorable effects of synthesis. Comparing the obtained complexes of Cu(II) with RLMD, either in solid state or in solution, it is obvious that, depending on pH values, various complex colors are obtained (Table 1). The change of the solutions color during the synthesis may be an indicator whether the syntheses of complexes were successful. The results obtained have shown that, in the range of pH 7.5–12, the color can vary from light green to dark blue. This is confirmed by the green solution color of the most stable complex of Cu(II) with RLMD (procedure 2, Table 1), in comparison with an indigo-blue alkali solution of decomposed Cu(II) at pH>12, where $[Cu(OH)4]^{2-}$ ions dominate.

Water solubility of synthesized complexes of Cu(II) with RLMD is different. The most water soluble complex is obtained at pH 7.5 (Table 1). The solution is permanent and stable after a longer period of time (6 months). The complexes that are synthesized at higher pH are less soluble. The solution of the complexes obtained, following the procedure 5 (Table 1), after resting for long period of time, start layering, precipitate and become opalescent. Medium pH is changed after adding Cu–salts and Cu(II) content in a complex is much influenced by it. Syntheses are performed at the same temperatures and within the same reaction period, but at different pH values (Table 1). The highest Cu(II) content was got at pH 7.5. The possibility of obtaining Cu(II)–RLMD complexes with a higher Cu(II) content has been tested with the increased concentration of Cu–salts. The expected results have not been obtained.

Solution pH probably has the influence on the way of binding of Cu(II) into a complex, i.e. on the type of a bond because, due to the change of pH value, the stability (Nikolic et al., 2006), the color and the solubility of the complex obtained are also changed (Table 1). Thus, by the increase of solution pH values from 7.5 to 12, the percentage of the bounded Cu(II) with RLMD in complex decreases. Some authors (Tolmachev et al., 1975), in the paper of influence of medium pH on binding Cu(II) with dextran, point out the possibility of gradual complexing, i.e. gradual forming of coordination bounds, where their reforming starts at pH 8. Thus, Cu(II) ions form of three different types of complexes with dextran (Norkus et al., 2002). Decomposition of Cu(II)-dextran complex begins at pH values higher than 12.

The results of ATR-FTIR spectroscopic investigations show that spectra of Cu(II)–RLMD complexes and ligand are basically similar (Fig. 12). The similarities of the bending (C–H) range indicate that there is no difference in the conformation of the glucopyranose unit in the dextran and the complex molecule (916 and 850 cm^{-1}), and they probably exhibit C-1 chair conformation (Mitic et al., 2007).

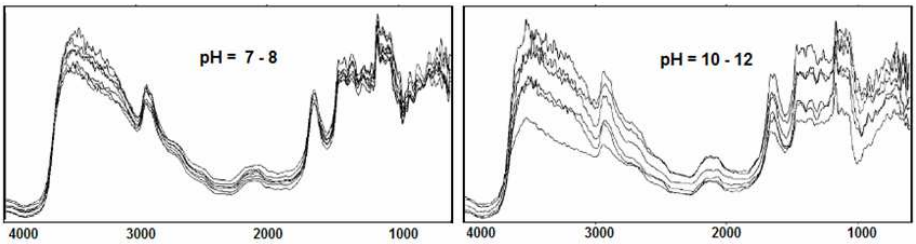
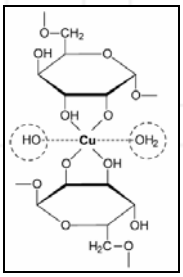
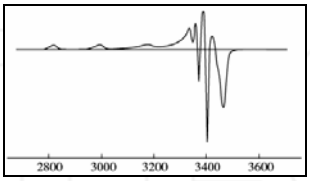


Fig. 12. ATR-FTIR spectra of Cu(II)–RLMD complexes synthesized at pH 7-8 and pH 10-12



Mid FTIR spectra of Cu(II)–RLMD complexes synthesized at different pH (pH 7–8, Fig. 13B, and pH 10–12, Fig. 13C) recorded at 298 K, show that the correlation between the O–H stretch frequency and the hydrogen bond strength. Spectroscopic IR study in particular region of OH (3400, 1420 cm^{-1}) and C–H (2900, 1460, 1350 cm^{-1}) vibrations indicates different binding between central metal ion and ligand, depending on pH and metal contents (Fig. 13). The difference in frequencies, intensity, and shape of these bands in the region 3600–3100 cm^{-1}, implies that in complexes which were synthesized at pH 10–12 there is the displacement of H_2O molecules by O–H groups in the first coordination sphere of the copper(II) ion. Dextran and complexes with Cu(II) ion have one crystallographic type of water molecule (1640 cm^{-1}). Water protons take part in the formation of relatively weak hydrogen bonds (Cakic et al., 2002; Nikolic et al., 2006; Nikolic et al., 2008).



The FTIR investigation corresponds with the results obtained by ESR spectrometry (Mitic et al., 2004), as well as with the results obtained by UV-VIS investigations (Mitic et al., 2007). ESR parameters of the spectra ($A_{||}$ = 187 × 10^{-4} cm^{-1}, $g_{||}$ = 2.23 and $g_{\perp}$ = 2.03), for the complexes synthesized at higher pH values, were close to the values for the frozen Cu(II)–ethylene glycol complex, thus indicating the square-planar coordination of Cu(II) ion with four oxygen atoms. Although the Cu(II) ion content of complexes synthesized at lower pH values was much higher (up to 18.95% for the complex synthesized at pH 7.5) the ESR signal of these complexes was lacking due to strong spin–spin interactions of neighboring Cu(II) ions. ESR spectrum of complex containing 6.97% of copper synthesized at pH 10 is

presented to the left. The ESR spectra indicate the axial symmetry of synthesized complexes and were typical for the Cu(II) ion with one unpaired electron in 3*d* subshell. Asymmetric appearance of the hyperfine spectral lines originates from the unresolved spectral contributions of two natural isotopes, 63Cu and 65Cu. ESR spectral parameters ($A_{||}$ = 187 × 10^{-2} cm^{-1}, $g_{||}$ = 2.23 and $g_{\perp}$ = 2.03) point to the tetragonal coordination of Cu(II) with four oxygen atoms from ligands in the same plane (Mitic et al., 2004; Nikolic et al., 2004).

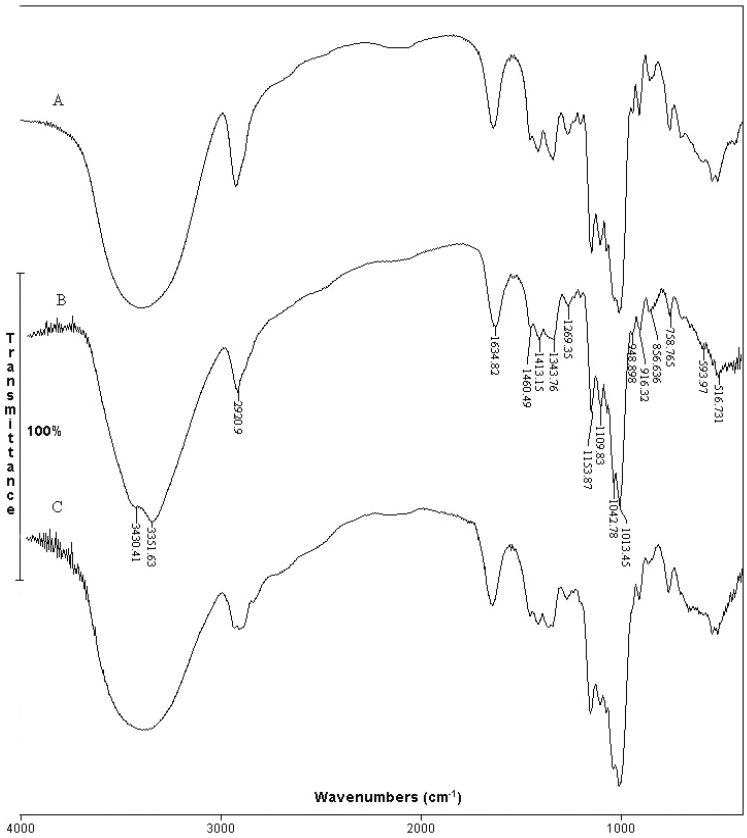


Fig. 13. FTIR spectra of: LM dextran (A); stable Cu(II)–dextran complex, with high metal content (≈ 18%), obtained at pH 7–8 (B); unstable Cu(II)–dextran complex, with low metal content (≈ 5%), obtained at pH 10–12 (C)

In addition, depending on pH values, complexes of Cu(II) with RLMD also behave differently considering wavelength at which they show absorption maximum. This range of wavelengths in the VIS spectra is 650–700 nm (Mitic et al., 2007). Hypsochromic effect of complexes absorption maximums with increase of pH solutions confirms the presence of different types of complexes. Hexaqua copper(II) ion $[Cu(H_2O)_6]^{2+}$ absorb at wavelength 812.7 nm, while synthesized complexes absorb within the ranges of 650–700 nm. With increase in solution pH the absorption maximums change to shorter wavelengths

comparing with $[Cu(H_2O)_6]^{2+}$ ion. Complex, which has been decomposed at pH values over 12, shows absorption maximum at 634 nm. Thereby, these spectrophotometric criteria can be applied for the confirmation of the success of complex synthesis. On the basis of the obtained results by spectroscopic investigations of this complexes, three different types of Cu(II) complexes structure with deprotonized dextran monomer unit (Glc–) are suggested depending on pH synthesis. At pH 7 to 8 $[Cu(Glc)_2(H_2O)_2]$ is formed; at pH 8 to 10 $[Cu(Glc)_2(OH)(H_2O)]^-$ is formed and at pH values over 10 $[Cu(Glc)_2(OH)_2]^{2-}$ is formed.

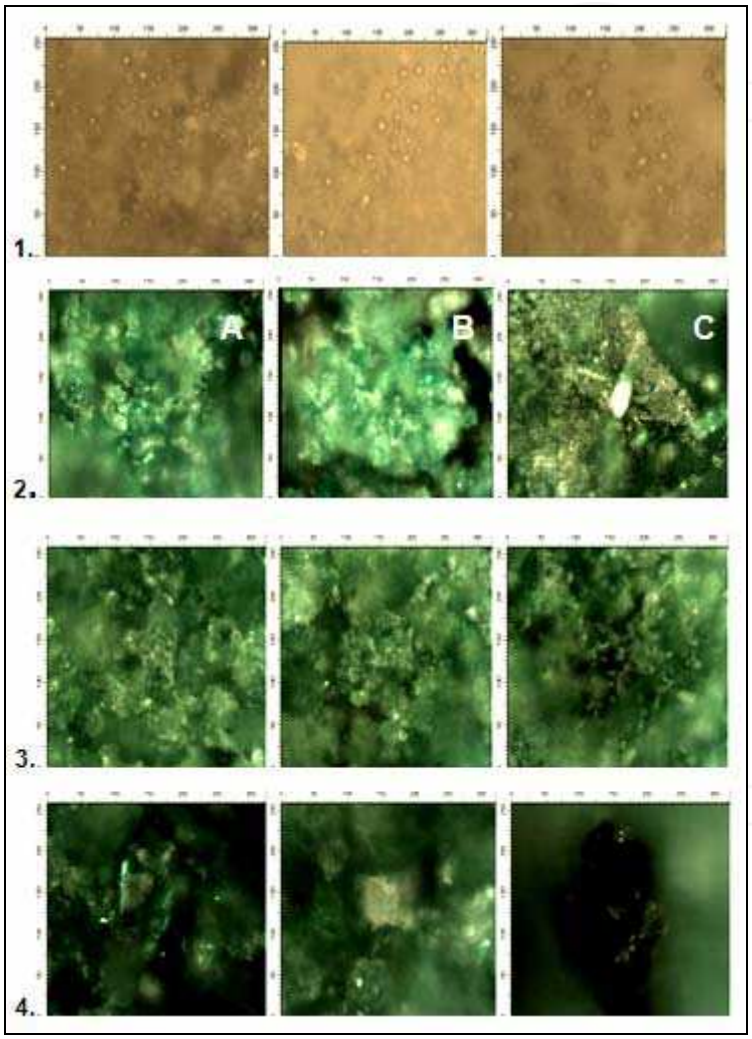


Fig. 14. FTIR microscopy images (250 μm x 300 μm) of RLMD, Mw = 5.000 g mol^{-1} (1) and Cu(II) complexes with RLMD synthesized at boiling point and pH 7-8 (2), 8-10 (3) and 10-12 (4)

The application of microscopic FTIR imaging system to the RLM dextran as ligand and Cu(II)–RLMD complexes, were synthesized at pH 7.5–12, is shown in Fig. 14. FTIR microscopy images of dextran, as well as images of the copper-dextran complexes differ which also indicates that the complexation process and the creation of coordination compounds took place (Mitic et al., 2010). FTIR microscopy images confirmed that the changes in the intensity of the analyzed bands are strongly associated with the alterations in the macromolecular order. These bands in the spectra of the complexes can be responsible for more and less ordered structures, respectively. The changes in color contour may show the content and distribution of copper, and polysaccharides in Cu(II)–RLMD samples.

After physicochemical standardization of the most stable complex obtained according to the procedure 2 (Tab. 1), the preparation for the pharmacological test was provided. The preparation was tested pharmacologically with the aim of determining systemic acute toxicity expressed as a median lethal dose (LD_{50}) and as an equivalent of Cu(II) dose per kg of a mouse body weight (Cakic et al., 2008). None of the applied complex doses at the concentration of 82–169 mg equivalent of Cu(II) per kg of a mouse's body weight was lethal in the tested mice. Therefore, in this case, a median lethal dose could not be determined. The application of higher doses has caused the mortality of one part of experimental animals. Thus, in this range, the preparation toxicity LD_{50} of 1419–1661 was determined, which corresponds to the equivalent of Cu(II) dose of 281–329 mg per kg of the body weight, in the concentration 5–20% of the complex solution. Toxicity investigations of various commercial copper salts show a wide range of values for LD_{50}. The level of the acute toxicity is higher for more soluble than for less soluble Cu(II) salts. The results of our pharmacological investigations point to the lower toxicity of Cu(II) complex with RLMD, what is much better than in the case of commercially applicable inorganic copper salts.

6. References

Bartkowiak, A.; Jezierska, J.; Spychaj, T. (1998). Self-assembled monolayers and polymer brushes in biotechnology. *Polym. Bull*., 41, 199-206.

Bellamy, L.J. (1954). *The infrared Spectra of Complex Molecules*, Methuen, London.

Belder, A.N. (1985). *Dextran*, Pharmacia AB, Uppsala, Sweden.

Bender, H.; Lehmann, J.; Wallenfels, K. (1959). Pullulan, an extracel- lular glucan from Pullularia pullulans. *Biochim. Biophys. Acta*, 36, 309-316.

Bernier, B. (1958). The production of polysaccharides by fungi active in the decomposition of wood and forest litter. *Can. J. Microbiol*. 4(3) 195-204.

Buslov, D.K.; Nikonenko, N.A.; Sushko, N.I.; Zhbankov, R.G. (1998). Analysis of the results of α-D-glucose Fourier transform infrared spectrum deconvolution: comparison with experimental and theoretical data. *Spectrochim. Acta A*, 55(1), 229-238.

Cakic, M.; Nikolic, G.; Ilic, Lj. (2002). FTIR spectra of iron(III) complex with dextran, pullulan and inulin oligomers. *Bull. Chem. Technol. Macedonia*, 21(2), 135-146.

Cakic, M.D.; & Nikolic, G.S. (2003). Iron complexes of pharmaceutical interest, Antianemics. *Chem. Industr.*, 57(10), 437-443.

Cakic, M.D.; Nikolic, G.S.; Mitic, Ž.J.; Ilic, Lj.A. (2004). Conformational analysis of copper(II)-dextran complexes by Fourier-Transform IR spectroscopy. *Proceedings of 4th International Conference of the Chemical Societies of the South–Eastern European Countries*, p. 72, Belgrade, Serbia, 18–21 July 2004.

Cakic, M.D.; Nikolic, G.S.; Mitic, Ž.J.; Ilic, Lj.A. (2006). Synthesis and spectroscopic characterization of zinc(II) complexes with dextran derivatives, *Proceedings of 5th International Conference of the Chemical Societies of the South-East European Countries*, p. 544, Ohrid, Macedonia.

Cakic, M.; Mitic, Ž. Nikolic, G.S.; Ilic, Lj.; Nikolic, G.M. (2008). The investigations of bioactive copper(II) complexes with reduced low-molar dextran. *Spectoscopy Int. J.*, 22(2-3), 177–185.

Carolan, G.; Catley, B.J.; McDougal, F.J. (1983). The location of tetrasaccharide units in pullulan. *Carbohydr. Res.*, 114, 237-243.

Catley, B.J. (1970). Pullulan, a relationship between molecular weight and fine structure. *FEBS Letters,* 10, 190-193.

Catley, B.J.; Ramsay, A.; Servis, C. (1986). Observations on the structure of fungal extracellular-polysaccharide, pullulan. *Carbohydr. Res.*, 153, 79-86.

Chenery, D. & Bowring, H. (2003). Infrared and Raman spectroscopic imaging in bioscience. *Spectrosc. Europe*, 15(4), 8-14.

Chiu, H.C.; Hsiue, T.; Chen, W.Y. (2004). FTIR-ATR measurements of the ionization extent of concentration-induced equilibrium swelling. *Polymer*, 45(5), 1627-1636.

Clark, R.L.; Pesolano, A.A.; Shen, Y.Y. (1997). S. Afr. Patent, 76, 03163, (*Chem. Abstr*. 1977, 87, P189951a).

Deshpande, M.S.; Rale, V.B.; Lynch, J.M. (1992). Aureobasidium pullulans in applied microbiology : a status report. *Enzyme Microb. Technol*., 14, 514-527.

Gaiduk, O.; Pantaler, R.; Grebenyuk, N.; Ostrovskaya, V. (2009). Rapid determination of copper(I, II*)* ions using reagent indicator paper. *J. Analyt. Chem*., 64(2), 201-205.

Gapanovich, V.; Petrov, P.; Melnova, N. (1998). Human and drug. *Abstracts of papers, V Russian National Congr.*, Moscow, 1998, p. 480.

Gierlinger, N. & Schwanninger, M. (2007). The potential of Raman microscopy and Raman imaging in plant research: review. *Spectroscopy Int. J.*, 21, 69-89.

Grigorieva, A.S.; Konakhovich, N.F.; Budarin, L.I. (1983). Possible applications of coordination compounds for correcting biometal metabolism in different pathologies. *Inorganica Chimica Acta*, vol. 79, pp. 297-298.

Gyurcsik, B. & Nagy, L. (2000). Carbohydrates as ligands: Coordination equilibria and structure of the metal complexes. *Coord. Chem. Rev.*, 203, 81–149.

Hardtmann, G.G. & Kathawala, F.G. (1977). US. Patent 053, 800, (*Chem. Abstr*. 88, 2297).

Expert group on vitamins and minerals. (2002). *Draft review of copper*, EVM/99/19, Secretariat, UK Government, London.

Ilic, Lj.; Ristic, S.; Cakic, M.; Nikolic G.; Stankovic, S. (2002). Polynuclear complex Fe(III) with pullulan oligomers, process of its obtaining, and pharmaceutical preparations on the basis of the complex, Pat. WO 02/46241 A2, Int. Cl. C08B 37/00, Int. Pub. Date 13 June, 2002.

Ilic, Lj.A.; Mitic, Ž.J.; Nikolic, G.S.; Cakic, M.D. (2003). Synthesis and pharmacological activity of copper(II) and cobalt(II) complexes with dextran oligomers, *Proceedings of XVII Ulusal Kimya Kongresi*, p. 190, Istanbul, Turkey.

Ishihara, M.; Obara, K.; Ishizuka, T.; Fujita, M.; Sato, M.; Masuoka, K.; Saito, Y.; Yura, H.; Matsui, T.; Hattori, H.; Kikuchi, K.; Kurita, A. (2003). Controlled release of fibroblast growth factors and heparin from photocrosslinked chitosan hydrogels and subsequent effect on in vivo vascularization. *J. Biomed. Mater. Res*., 64A, 551-59.

Jackson, W.D. & Polaya, J.B. (1951). *Aust. J. Soc.*, vol. 13, pp. 149.

Kacurakova, M. & Mathlouthi, M. (1996). FTIR and laser-Raman spectra of oligosaccharides in water: Characterization of the glycosidic bond. *Carbohydr. Res.*, 284, 145-157.

Kacurakova, M. & Wilson, R.H. (2001). Developments in mid-infrared FT-IR spectroscopy of selected carbohydrates. *Carbohydr. Polym.*, 44, 291-303.

Kathawala, F.G. (1974). US. Patent 3, 850, 932 (*Chem. Abstr.* 82, 140175).

Kim, S.; Chae, S.Y.; Na, K. ; Kim, S.W.; Bae, Y.H. (2003). Insulinotropic activity of sulfonylurea/pullulan conjugate in rat islet microcapsule. *Biomaterials*, 24(26), 4843-4851.

Kiselev, V.P.; Komar, V.P.; Skornyakov, I.V.; Firsov, S.P.; Virnik, A.D.; Zhbankov, R.G. (1977). Spectroscopic study of various structural modifications of dextran. *Polym. Sci. USSR*, 19(8), 2140-2146.

Klimovich, O.; Gapanovich, V.; Petrov, P. (1998). Human and drug. *Abstracts of papers, V Russian national Congr.*, Moscow, pp. 493.

Komar, V.P.; Zhbankov, R.G.; Khomyakov, K.P.; Zavada, O.A.; Virnik, A.D.; Rogovin, Z.A. (1968). Dextran and derivatives. *Polym. Sci. USSR*, 10, 1351-1357.

Kweon, D.K.; Song, S.B.; Park, Y.Y. (2003). Preparation of water-soluble chitosan/heparin complex and its application as wound healing accelerator. *Biomaterials*, 24, 1595-1601.

Lawrence, R.; Lange, R.; Wu, C.; Helenek, M. (1997). Iron dextran formulations, US Pat. 5 624 668, Int. Cl A01N059/16, US Cl 424/078, Appl. No. 536 984.

Leathers, TD. (2003). Biotechnological production and applications of pullulan. *Appl. Microbiol. Biotechnol.*, 62(5–6), 468-473.

Leonard, M.; Rastello De Boisseson, M.; Hubert, P.; Dellacherie, E. (2003). Production of microspheres based on hydrophobically associating alginate derivatives by dispersion/gelation in aqueous sodium chloride solutions. *J. Biomed. Mater. Res.*, 68A (2003) 335-342.

Lepetil, G.S.P.A. (1975). Ger. Patent, 2. Vol. 424, 670 (*Chem. Abstr.* 1975, 83, 20628).

Lewis, A.G. (1995). *The Biological Importance of Copper*, Literature Review ICA Project, No. 223, British Columbia, pp. 7–17.

Lewis, S.L.; Heitkemper, M.M.; Dirksen, S.R.; O'Brien, P.G.; Bucher, L. (2008). *Medical Surgical Nursing : Assessment and Management of Clinical Problems*, 7th Edition, Mosby Elsevier.

Li, J.; Pan, J.; Zhang, L.; Guo, X.; Yu, Y. (2003). Culture of primary rat hepatocytes within porous chitosan scaffolds. *J. Biomed. Mater. Res.*, 67(3A), 938-943.

Lin, S.Y.; Li, M.J.; Cheng, W.T. (2007). FT-IR and Raman vibrational microspectroscopies used for spectral biodiagnosis of human tissues. *Spectroscopy*, 21, 1-30.

London, E. (2004). The molecular formula and proposed structure of the iron-dextran complex, imferon. *J. Pharm. Sci.*, 93(7), 1838-1846.

Lugovaya, Z.A.; Tolmachev, V.N.; Naiden, V.I.; Valakhanovich A.I.; Zaretskaya, R.V. (1976). Reaction products of low molecular-weight dextran with copper. *Pharm. Chem. J.*, 10(11), 1536–1538.

Mähner, C.; Lechner, M.D.; Nordmeier, E. (2001). Synthesis and characterization of dextran and pullulan sulphate. *Carbohydr. Res.*, 331, 203-208.

Mitic, Ž.; Cakic, M.D.; Nikolic, R.; Ilic, Lj.; Nikolic, G.S. (2004). Polysaccharide complexes of pharmaceutical interest. Water solutions stability of Cu(II) complexes with reduced low-molecular dextran. *World Polym.*, 7(5), 225–229.

Mitic, Ž.; Nikolic, G.; Cakic, M.; Nikolic, R.; Ilic, Lj. (2007 a). Synthesis and spectroscopic characterization of copper(II) dextran complexes. *Russ. J. Phys. Chem.*, 81(9), 1433–7.

Mitic, Ž.; Cakic, M.D.; Nikolic, G.S.; Nikolic, R.S.; Ilic, Lj.A. (2007 b). The investigation of Co(II) dextran complexes. *Chem. Ind.*, 61(5), 257–262.

Mitic, Ž.; Nikolic, G.S.; Cakic, M.D.; Nikolic, R.; Ilic, Lj. (2007 c). Fourier-transform IR spectroscopy of Cu(II) ion complexes with different polyglucans, *Proceedings of XII European Conference on the Spectroscopy of Biological Molecules (ECSBM)*, p. 319, Bobigny, Paris, France, 1–5 September 2007.

Mitic, Ž.; Nikolic, G.S.; Cakic, M.D.; Premovic, P.I.; Ilic, Lj. (2008 a). FTIR spectroscopic characterization of Cu(II) coordination compounds with exopolysaccharide pullulan and its derivatives. *Proceedings of XXIX European Congress on Molecular Spectroscopy (EUCMOS)*, Opatija, Croatia, 31 Aug–5 Sep 2008, p. 197.

Mitic, Ž.; Nikolic, G.S.; Cakic, M.D. (2008 b). Development of ATR-FTIR microspectroscopy for analysis of bioactive copper(II)-polysaccharide complexes. *Proceedings of Conference SPEC2008, V International Conference of Spectroscopy, Shedding light on disease : Optical diagnosis for the New Millennium*, Sao Paulo, Brazil, 25–29 October 2008, p. 86-88.

Mitic, Ž.; Cakic, M.; Nikolic, G.; Ilic, Lj.; Stankovic, M. (2010). Spectroscopic characterization of bioactive Cu(II) complexes with polysaccharides by FTIR microspectroscopy. *Chem. Ind.*, 64(1), 9–20.

Mousia, Z.; Farhat, I.A.; Pearson, M.; Chesters, M.A.; Mitchell, J.R. (2001). FTIR microspectroscopy study of composition fluctuations in extruded amylopectin–gelatin blends. *Biopolymers (Biospectroscopy)*, 62(4), 208-218.

Nagy, L.; Yamaguchi, T.; Yoshida, K. (2003). Application of EXAFS and XANES methods in coordination chemistry of carbohydrates and their derivatives. *Struct. Chem.*, 14(1), 77-84.

Nakamoto, K. & Czernuszewicz, R.S. (1993). *Infrared Spectroscopy, in J.F. Riordan and B.L. Vallee, eds., Methods in Enzymology*, Vol. 226, Part C, Academic Press, San Diego, CA.

Nakamoto, K. (2009). *Infrared and Raman Spectra of Inorganic and Coordination Compounds. Part B: Applications in Coordination, Organometallic, and Bioinorganic Chemistry*, Sixth Edition, John Wiley & Sons, Inc., Hoboken, New Jersey, pp. 333, ch 3.

Nikolic, G.M.; Ilic, Lj.; Cakic, M.D.; Nikolic, G.S.; Nikolic, R.; Mitic, Ž. (2004). Copper(II) complexes of reduced low-molecular dextran: an ESR study, *Proceedings of 7th International Conference on Fundamental and Applied Aspects of Physical Chemistry*, pp. 775–777, 21–23 September 2004, Belgrade, Serbia.

Nikolic, G.S.; Cakic, M.; Cvetkovic, D.; Ilic, Lj.; Premovic, P.I. (1996). A spectroscopic study of polynuclear complex of iron(III) with inuline oligomers. *Spectrosc. Lett.*, 29(7), 1437-1448.

Nikolic, G.S.; Cakic, M.D.; Ilic, Lj.; Ristic, S.; Cakic, Ž. (2002). Synthesis of some new antianemics I. Iron pullulan complexes of pharmaceutical interest. *Die Pharmazie* 57(3), 155–158.

Nikolic, G.S.; Cakic, M.D.; Mitic, Ž.; Nikolic, R.; Ilic, Lj. (2005). Correlation between UV–VIS spectra and structure of Cu(II) complexes with hydrogenated dextran in alkaline solutions. *Chem. Industr.*, 59(1/2), 9–14.

Nikolic, G.S.; Mitic, Ž.; Cakic, M.D.; Ilic, Lj. (2006 a). Study of copper(II)–dextran complexes stability by conductometry method, *Proceedings of 8th International Conference on Fundamental and Applied Aspects of Physical Chemistry*, pp. 719–721, Belgrade, Serbia.

Nikolic, G.S.; Cakic, M.D.; Mitic, Ž.; Ilic, Lj. (2006 b). Copper(II) complexes of reduced low-molar dextran: a deconvoluted LNT-FTIR study, *Proceedings of 8th International Conference on Fundamental and Applied Aspects of Physical Chemistry*, pp. 722–724, Belgrade, Serbia.

Nikolic, G.S. & Cakic, M.D. (2007). Physical investigation of the colloidal iron-inulin complex. *Colloid. J.*, 69(4), 464–473.

Nikolic, G.S.; Cakic, M.D.; Mitic, Ž.; Ilic, Lj. (2008). Deconvoluted Fourier-transform LNT-IR study of coordination copper(II) ion compounds with dextran derivatives. *Russ. J. Coord. Chem.*, 34(5), 322-328.

Nikonenko, N.A.; Buslov, D.K.; Sushko, N.I.; Zhbankov, R.G. (2005). Spectroscopic manifestation of stretching vibrations of glycosidic linkage in polysaccharides. *J. Mol. Struct.*, 752, 20-24.

Norkus, E.; Vaiciuniene, J.; Vuorinen, T.; Heikkilä, M. (2002). Interaction of Cu(II) with dextran in alkaline solutions. *Carbohydr. Polym.*, 50(2), 159-164.

Norkus, E.; Vaiciuniene, J.; Vuorinen, T.; Macaladay, D. (2004). Equilibria of Cu(II) in alkaline suspensions of cellulose pulp. *Carbohydr. Polym.*, 55(1), 47-55.

Panov, V.P. & Zhbankov, R.G. (1976). *Conformations of Sugars*, Nauka i Tekhnika, Minsk.

Panov, V.P.; Zhbankov, R.G. (1988). *Intra- and intermolecular interactions in carbohydrates*, Nauka i Tekhnika, Minsk.

Parker, F.S. (1983). *Application of Infrared Raman and Resonance Raman Spectroscopy in Biochemistry*, Plenum Press, New York.

Pekic, B. & Cvetkovic, D. (1988). Polynuclear complex Fe(III) with dextran and its derivatives,YU Pat. 42384 B, Int. Cl. C08B37/02, Appl. No. 564.

Perets, A.; Baruch, Y.; Weisbuch, F.; Shoshany, G.; Neufeld, G.; Cohen, S. (2003). Enhancing the vascularization of three-dimensional porous alginate scaffolds by incorporating controlled release basic fibroblast growth factor microspheres. *J. Biomed. Mater. Res.*, 65A, 489-497.

Platonova, N.; Zhiguleva, E.; Tsymbalenko, N.; Mishchenko, B.; Vasin, A.; Zhivulko, T.; Puchkova, L. (2004). Age-related features of ceruloplasmin biosynthesis and distribution in rats. *Russ. J. Develop. Biology*, 35(3), 131-141.

Serizawa, T.; Yamaguchi, M.; Kishida, A.; Akashi, M. (2003). Alternating gene expression in fibroblasts adhering to multilayers of chitosan and dextran sulfate. *J. Biomed. Mater. Res.*, 67A, 1060-1063.

Shingel, K.I. (2002). Determination of structural peculiarities of dexran, pullulan and gamma-irradiated pullulan by Fourier-transform IR spectroscopy. *Carbohydr. Res.*, 337(16), 1445-1451.

Shingel, K.I. (2004). Current knowledge on biosynthesis, biological activity and chemical modification of the exo-polysaccharide pullulan. *Carbohydr. Res.*, 339, 447-460.

Sivchik, V.V.; Zhbankov, R.G.; Astreiko, M.V. (1979). Theoretical Investigation of the Vibrational Spectrum of Dextran. *Acta Polym.*, 30, 689-693.

Skornyakov, I.V. & Komar, V.P. (1996). Infrared spectra of structural modifications of dextran. *J. Appl. Spectrosc.*, Volume 63, No. 3, 309-317.

Smits, A.L.M., Ruhnau, F.C.; Vliegenthart, J.F.; Van Soest, J.J.G. (1998). Ageing of starch-based systems as observed with FTIR and solid-state NMR spectroscopy. *Starch/Staerke*, 50, 478-483.

Solpan, D. & Sahan, M. (1993). Concentration and separation of metallic ions from aqueous solutions by using EC-PEG 4000 alloy membranes. *J. Appl. Polym. Sci*., 48(12), 2091–2093.

Sowa, W.; Blackwood, A.C.; Adams, G.A. (1963). Neutral extracellular glucan of Pullularia pullulans (de Bary) Berkhout. *Can. J. Chem.,* 41(9), 2314-2319.

Tanodekaew, S.; Prasitsilp, M.; Swasdison, S.; Thavornyutikarn, B.; Pothsree, T.; Pateepasen, R. (2004). Preparation of acrylic grafted chitin for wound dressing application. *Biomaterials*, 25(7), 1453-1460.

Tolmachev, V.N.; Lugovaya, A.Z.; Ischenko, I.K.; Valikhonovich, A.I.; Zaboronok, V.U. (1975). *Vysokomol. Soedin., Ser.* A, 17(2), p. 419-425.

Vosburg, W.C. & Cooper, G.R. (1941). Complex Ions. *J. Am. Chem. Soc*., vol. 63, pp. 437-442.

Wallenfels, K.; Bender, H.; Keilich, G.; Bechtler, G. (1961). On pullulan, the glucan of the slime coat of Pullularia pullulans. *Angew. Chem.,* 73, 245-246.

Wallenfels, K.; Keilich, G.; Bechtler, G.; Freudenberger, D. (1965). Untersuchungen an pullulan. Die klarung des strukturproblems mit physikalischen, chemischen und enzymatischen methoden. *Biochemische Zeitschrgt*, 341, 433-450.

Walker, H.A.; Wilson, S.; Atkins, E.C.; Garret, H.E.; Richardson, A.R. (1951). *J. Pharmacol. Exp. Ther.*, vol. 101, No. 4, pp. 368-378.

Yin, Y.; Ye, F.; Cui, J.; Zhang, F.; Li, X.; Yao, K. (2003). Preparation and characterization of macroporous chitosan-gelatin. *J. Biomed. Mater. Res*., 67A, 844-855.

Yu, W. & Xiong, L. (2005). Effect of various morphology and testing conditions on fiber infrared spectrum. *J. Appl. Polym. Sci*., 96(4), 1003-1010.

Zhbankov, R. (1972). *Infrared Spectra and Structure of Carbohydrates*, Nauka i tekhnika, Minsk, pp. 56–95, and pp. 180–227.

Zhbankov, R.G. & Avsenev, N.N. (1984). Study of the ordered conformational states of dextran macromolecules in solutions with additions of alkali halide salts. *Polym. Sci. USSR*, 26(2), 262-271.

Zhbankov, R.G. (1992). Vibrational spectra and structure of mono- and polysaccharides. *J. Mol. Struct*., 275, 65-84.

Zhbankov, R.G.; Andrianova, V.M.; Marchewka, M.K. (1997). Fourier Transform IR and Raman spectroscopy and structure of carbohydrates. *J. Mol. Struct*., 436, 637-654.

Zhbankov, R.G.; Firsov, S.P.; Korolik, E.V.; Petrov, P.T.; Lapkovski, M.P.; Tsarenkov, V.M.; Marchewka, M.K.; Ratajczak, H. (2000). Vibrational spectra and the structure of medical biopolymers. *J. Mol. Struct*., 555, 85-96.

Zhbankov, R.G.; Firsov, S.P.; Kolosova, T.E.; Prihodchenko, L.K.; Grinshpan, D.D.; Baran, J.; Marchewka, M.K.; Ratajczak, H. (2003). Fourier-transform Raman spectroscopy of biological molecules: Weak intermolecular interactions. *J. Mol. Struct*., 656, 275-286.

Zhbankov, R.G.; Zhbankova, M.R.; Baran, J.; Marchewka, M.; Ratajczak, H. (2005). Anomalously high intensity of bands in the FTR-spectra of individual kinds of polymers. *J. Mol. Struct*., 744/747, 585-587.

3

Precision Quantitative Proteomics with Fourier-Transform Mass Spectrometry

Qingbo Li
University of Illinois at Chicago
U.S.A.

1. Introduction

The Fourier-transform mass spectrometry (FTMS) instrument offers a mass resolution higher than most of other mass spectrometers. This high resolution is in part due to the better stability offered by a superconductor magnet in FTMS than a radio frequency voltage utilized in many other mass spectrometers. The extremely high resolution of FTMS has very important application in biomedical proteomics research. The high resolution not only dramatically improves the reliability of protein identification but also the accuracy of protein quantitation.

In this chapter, we present several examples of proteomics study that takes advantage of the high resolution offered by FTMS. Particularly, we describe examples of proteome dynamic study with isotopomer analysis, and precise peptide and protein label-free quantitation with rigorous statistical assessments.

In protein dynamic studies, FTMS readily resolves all of the isotopomer peaks for peptides. The well-resolved isotopomer peaks allow a direct integration of the intensities of different isotopologue envelopes without the need of a deconvolution algorithm. With the high-resolution data, we were able to show that protein turnover measurement revealed more subtle changes in the dynamics of a proteome.

The high resolution of FTMS helps to reduce the interference of contaminant peaks (**Fig. 1**). The ability to resolve the targeted peptide isotopomer peaks from interfering ones greatly facilitates the implementation of a label-free quantitative proteomics method that relies on peptide cross reference between liquid chromatography runs and the integration of extracted ion chromatographic intensities (Lipton et al., 2002). With both high confidence in peptide identification and quantitation, we showed that the major source of variability lies more in sample preparation than in liquid chromatography/mass spectrometry analysis. Such a result has a direct consequence in the statistical approaches utilized to assign significance in label-free quantitative proteomics.

Two sections are presented in this chapter. The first one briefly discusses the general aspect of a protein turnover analysis followed by examples of proteome dynamics study in acid stressed and iron limited mycobacterial cells. The second one describes the utilization of high-resolution FTMS data for label-free quantitation of proteins with a rigorous statistical assessment of significance in differential protein abundances.

The spectra shown in the panels a to e of Fig. 1 are for a tryptic peptide from a superoxide dismutase in *M. smegmatis* (MSMEG_6427; Mn) with a sequence of AFWNVVNWDDVQNR.

The panels a and b are the ion-trap mass spectrometry (ITMS) MS^2-scan spectra for a +2 precursor ion of the peptide in the nanoLC/LTQ-FT and the nanoLC/LTQ respectively. Shown atop of panel a is the peptide sequence labeled with the detected y-series ions in the MS^2-scan spectra. For clarity, only y-series ion fragments are shown with labels. The panels c to e show the MS-scan spectra obtained with a FTMS scan (c), an ITMS zoom MS-scan (d), and an ITMS full MS-scan (e), respectively. Only an m/z range of 880 to 885 is shown for the +2 peptide charge state. The three short arrows in the panels c, d, and e indicate the first three isotopic peaks of the +2 peptide charge state. The asterisks in panel c indicate the isotopic peaks probably not related to the peptide. The theoretical molecular weight of the peptide is shown atop of panel c. The long dashed arrows point to the respective MS-scans from which a precursor ion is selected for the MS^2 scans. Adapted from (Li, 2010a).

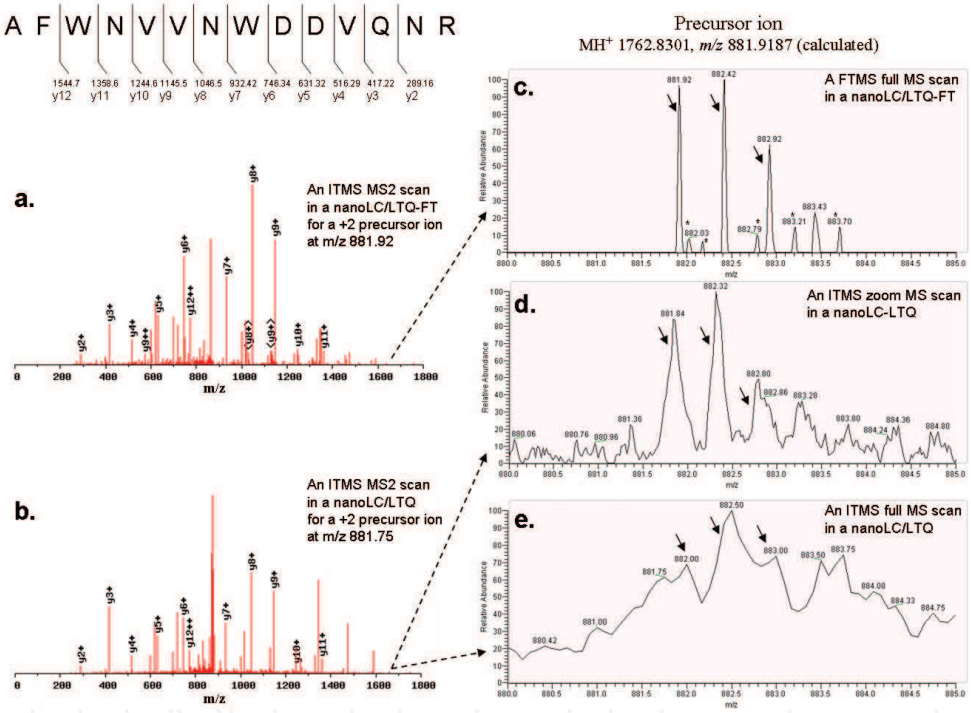


Fig. 1. A resolution and a signal-to-noise ratio in MS-scan and MS^2-scan spectra are compared between a nanoLC/LTQ-FT and a nanoLC/LTQ mass spectrometry systems

2. Protein turnover analysis with FTMS

2.1 Protein turnover

Protein turnover is a fundamental cellular process in all cell types having important implications in many aspects of biological science (Larrabee et al., 1980; Wilkinson, 2005). Advancement of high resolution proteomic technologies has provided the possibility to study protein turnover for multiple proteins simultaneously in complex cellular protein extracts (Beynon, 2005; Cargile et al., 2004; Pratt et al., 2002; Rao et al., 2008a; Rao et al.,

2008b; Vogt et al., 2005). We showed that a combination of protein abundance and turnover data provides a highly interesting insight into the dynamic process of and interconnection among protein synthesis, degradation, and secretion (Rao et al., 2008a).

Protein turnover shares an equally important role with gene transcription and protein translation. Synthesis of new proteins and degradation of old ones form a dynamic process in an organism. Turnover does not only help to clear old proteins but also aid in a fast adaptation to a new condition or environment by adjusting the rate of protein synthesis and degradation (Goldberg & Dice, 1974). Apart from this, turnover also brings new proteins into action with reduced strain on the resources of an organism because preexisting cellular materials are reused. Some early global protein turnover studies identified different *E. coli* proteins that might have different turnover rates. One of those earlier studies showed that a dynamic state for individual proteins existed in non-growing as well as growing cells (Larrabee et al., 1980). Studies of turnover offer a dynamic view of the abundances of proteins. When being applied to a larger scale of the proteome, protein turnover analysis allows one to study the dynamic nature of the entire proteome (Li, 2010a).

Mass spectrometry continues to serve as a major approach for a protein turnover study due to its wide availability and flexibility to analyze both single-cell cultures and multi-cellular organisms. Except for the required administration of stable isotope-labeled metabolites, amino acids, or water to the study subjects, mass spectrometry-based approaches do not require any genetic manipulation of the study subjects. The avoidance of genetic manipulation of the study subjects helps to minimize any unwanted perturbation to a biological system and delivers the most physiologically relevant results. A range of methodologies has been established to measure protein turnover based on stable isotope labeling and mass spectrometry (Doherty & Beynon, 2006).

2.2 High-resolution mass spectrometry for protein turnover analysis

The advent of highly automated high-resolution mass spectrometry technology promises to bring about in-depth insight into the dynamic nature of a proteome at the global level. The work done by Pratt et al. (Pratt et al., 2002) demonstrated the determination of protein degradation rate constants in a steady state population of yeast grown in a chemostat. The authors used isotope labeling along with 2D gel analysis to study protein turnover and advocated that protein turnover is 'a missing dimension in proteomics.' Another study done by Cargile et al. (Cargile et al., 2004) labeled *E. coli* cells with ^{13}C to study the relative synthesis over degradation ratio (S/D). These earlier works demonstrated the global analysis of protein turnover with individual protein identifications but their data were not correlated with abundance values.

With one-dimensional SDS/PAGE fractionation and subsequent nanoLC/LTQ-FTMS analysis, Rao et al. determined the global turnover profiles of *Mycobacterium smegmatis*, a non-pathogenic surrogate of *Mycobacterium tuberculosis*, under acid-shock and iron-limitation conditions (Rao et al., 2008b). A dynamic range of 3-orders of magnitude was demonstrated for relative turnover measurements. The study provided direct evidence that relative turnover in growing mycobacterium cells, with or without stress, was highly heterogeneous. The results obtained in that study addressed the long-standing question whether a 'dynamic state' exists in growing bacteria (Borek et al., 1958), and illustrated the benefits and needs to study protein turnover at the global level with the most advanced mass spectrometry technology.

In the work by Rao et al. (Rao et al., 2008b), the cells were grown with two different methodologies for both different types of stresses. For the pH stress the cells were initially grown in [^{14}N]-containing media at pH 7.0. Once the cells are in the initial log phase, the cells were divided into two flasks and the media was doped with 50% [^{15}N] and the pH was reduced to 5.0 in one of the flasks. The cells were harvested after one doubling and analyzed for protein turnover using LC/LTQ-FTMS. For low iron analysis, the cells were first grown in [^{15}N]-containing media until mid-log phase. The cells were then collected by centrifugation and the media was then exchanged with [^{14}N]-containing media. The cells were then allowed to grow to one doubling and harvested to be analyzed by LC/LTQ-FTMS.

In either the complete isotope swapping (iron-limitation experiments) or the partial isotope labeling conditions, the isotopologue envelopes and the individual isotopomer peaks of a peptide are clearly resolved. The complete resolution of the isotopomer peaks and the isotopologue envelops for the old proteins and the de novo synthesized proteins facilitates simple calculation of the abundances of the new and old fractions of a peptide (**Fig. 2**).

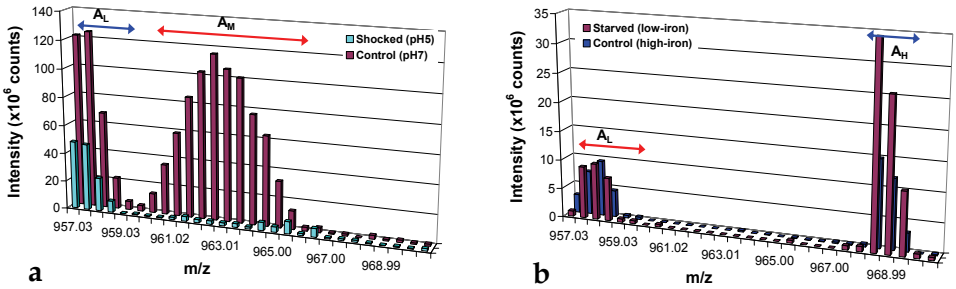


Fig. 2. Calculation of isotopologue intensities for a representative tryptic peptide ANLLGLSAPEMTTLVGGLR (MH_2^{2+}, 23 N atoms) of protein KatG (MSMEG6346) in shocked (pH5) and control (pH7) cultures (panel a), and in starved (low-iron) and control (high-iron) cultures (panel b). A_L, A_H, and A_M represent the isotopologue intensities with light label (99.6At%[^{14}N]), heavy label (99At%[^{15}N]) and medium label (50At%[^{15}N]) respectively. Red arrows and text labels indicate *de novo* synthesized proteins. Blue arrows and text labels indicate old proteins. Adapted from (Rao et al., 2008b) with permission

Turnover analysis of *M. smegmatis* under both stressful conditions revealed two different patterns (Rao et al., 2008b). In the low pH condition, many proteins had increased turnover at pH 5.0 as compared to pH 7.0. It was an obvious reaction since the bacteria has to readjust its proteome in order to counter the stress posed by increased proton concentrations. The correlation coefficient for the low pH shock cells was small which indicated that the proteins in the cells exposed to pH 5.0 underwent extensive readjustment in different directions. In the low iron stress the correlation coefficient being high suggested that either there was not much rearrangement of turnover values or all the proteins had changes in a similar direction. KatG and Tpx, which are important for protection of mycobacterial cells against oxidative stress, had low protein turnover values in both low iron as well as low pH conditions. A study on *M. tuberculosis* Tpx suggested that it might be an important protein

against oxidative stress because Tpx mutants were unable to survive in the macrophages in an infected mouse model. However, it would be interesting to analyze how the low turnover of Tpx correlates with the survival of mycobacteria in the cell.

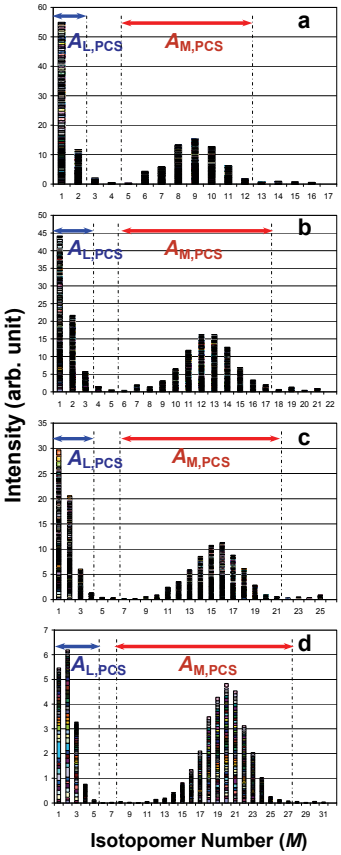


Fig. 3. Average isotopomer profiles and selected isotopomer ranges to calculate the abundances of peptides in *M. tuberculosis*. $A_{L,PCS}$ and $A_{M,PCS}$ are peptide abundance for old and new proteins respectively. Each panel was the stacked column graph of the normalized isotopomer profiles of the detected peptide charge states (PCSs) having the same number of N atoms (n). Profiles are shown for n equal to 11 (a), 16 (b), 20 (c), and 26 (d) respectively. The blue and red arrows indicate the M ranges for calculating $A_{L,PCS}$ and $A_{M,PCS}$ respectively

An open question is how the protein turnover values correlate with protein abundances. To investigate the correlation between protein abundance and protein turnover values in *M. tuberculosis*, Rao et al. analyzed *M. tuberculosis* cells in an iron replete and iron depleted condition using the high resolution LC/LTQ-FTMS instrument (Rao et al., 2008a). The approach employed many large-scale quantitative proteomics techniques to make it readily accessible for protein turnover studies at the global level. The concomitant measurement of protein turnover and abundance was previously shown by Gerner et al. with 2D gel

electrophoresis for separation and fluorography and autoradiography for quantitation of individual gel spots without any protein quantification (Gerner et al., 2002). With the advanced nanoLC/LTQ-FTMS system and a label-free quantitation approach, Rao et al. demonstrated that protein abundance and turnover could both be measured with a dynamic range of at least 3-orders of magnitude in an automated fashion (Rao et al., 2008a).

The study compared the sensitivity of relative turnover and relative abundance measurements to detect a dynamic response of *M. tuberculosis* when the cell culture was shifted from an iron-starved to iron-sufficient condition. An unlabeled iron-depleted *M. tuberculosis* culture was grown to late-log phase and diluted with a fresh iron-replete medium that was labeled with [^{15}N]-labeled nitrogen source (Rao et al., 2008a). The incorporation of the [^{15}N]-labeled nitrogen source into the newly synthesized proteins resulted in a complete separation between the old and the newly synthesized peptide isotopologue profiles (**Fig. 3**). Similar to that shown in Fig. 2 for *M. smegmatis*, the complete separation of the isotopologue profiles and the isotopomers allows the quantitation of the abundances of old proteins, newly synthesized proteins, and the total proteins.

In this work, we are able to obtain both the protein abundance and turnover values to more comprehensively assess the dynamic response of the H37Rv cells when they were shifted from iron-starved stationary-phase to fresh low- and high-iron media. This is achieved by applying both the isotope chasing and a label-free quantitation method (Rao et al., 2008a).

The results indicated that a relative turnover measurement was much more sensitive to monitor the dynamic response of the *M. tuberculosis* cells. Meanwhile, a combination of turnover and abundance measurements provided insight into the correlation of protein synthesis, degradation, and secretion. A further principal component analysis of the *M. tuberculosis* proteome dynamics reveals that protein relative turnover properties are orthogonal to protein relative abundance properties (Rao & Li, 2009a). Thus, a study of protein turnover at the global level would likely bring forward new findings that can be missed with a protein abundance analysis alone.

The data obtained from the comparison of protein abundance and protein turnover values showed that protein turnover is a much more sensitive measurement to discern the changes in the proteome than abundance measurements. Upon the transfer of late-log phase cells from a low iron to a high iron media, protein abundance measurements showed that out of the 104 proteins that we identified, only 5 proteins were upregulated and 16 proteins were downregulated in the HI media. Relative abundance of KatG was upregulated in cells grown in the high iron media.

Protein turnover analysis of the proteins compared between cells grown in a low iron and a high iron media showed that more proteins had increased synthetic activity in the high iron grown cells. The S/D had increased for 24 proteins in the cells grown in the low iron media. Eight proteins had decreased turnover. However, for cells grown in the high iron media, 56 proteins had increased S/D and 5 proteins had decreased S/D. A comparison of protein abundance measurements to protein turnover measurements clearly suggests that protein turnover does give more information to uncover the dynamic response of the proteome (**Fig. 4**).

In addition to providing the information about synthesis and degradation of proteins, the protein turnover analysis can also provide information regarding whether a protein has been secreted when the protein turnover values are analyzed together with the protein abundance measurements. In our study of proteome dynamics, we found that some proteins had low changes in relative abundances even though their synthesis had increased

significantly. As stated before the relative abundance of a protein in the cell can be affected not only by synthesis or degradation but also by a secretion process. In our turnover analysis, we found discrepancies between protein abundance values of certain proteins and their turnover values. A previous proteomic study of *M. tuberculosis* culture filtrates showed many of those proteins to be secreted into the culture filtrate. Proteins such as FbpC2, KatG, and the mammalian cell entrance protein Rv0172 were also predicted to be secreted (Malen et al., 2007).

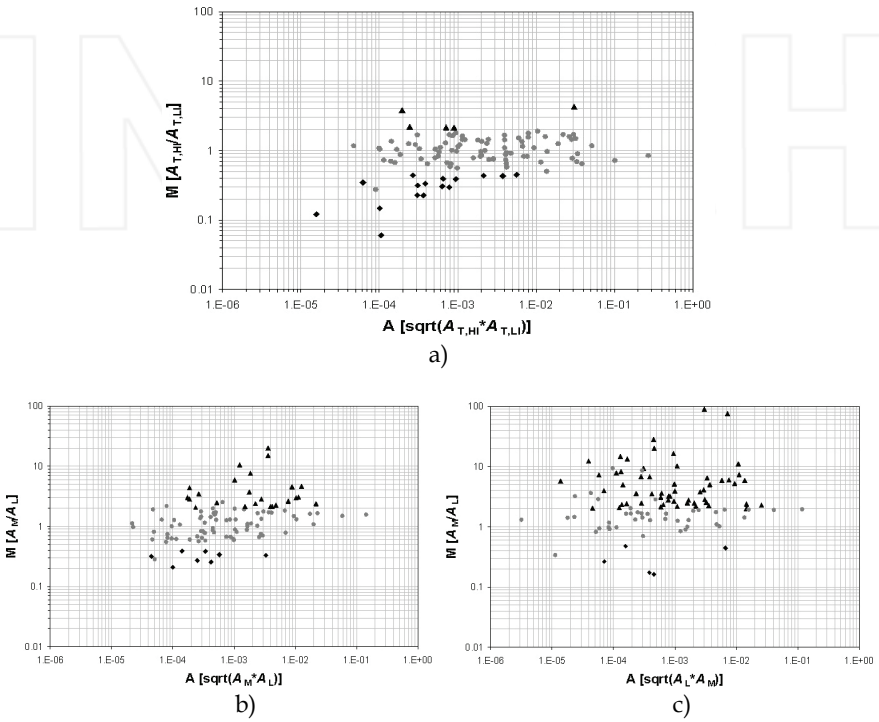


Fig. 4. M-A plots representing the total protein abundances in high iron (HI) versus low-iron (LI) cells (panel a), the newly synthesized protein abundances versus the old protein abundances in LI cells (panel b) and in HI cells (panel c) respectively. The proteins with 2-fold significant (p <.05) change in relative abundance are marked with black triangles and diamonds. The M–axis represents the relative abundance values and the A-axis represents the average abundance values. Adapted from (Rao et al., 2008a) with permission

These results support that protein turnover in combination with abundance analysis could predict the secretion of proteins and reveal the interconnected roles of protein synthesis, degradation, and secretion in determining the protein abundances in cells. These analyses illustrate that protein turnover can divulge information that classical proteomics does not provide. The integration of data from transcriptome studies, abundance measurements and turnover analyses will likely provide a more complete picture of the dynamics associated with the proteome. To some extent, it will probably reconcile the discordances between transcriptome and proteome analyses.

2.3 Implication of proteome turnover studies in mycobacteria

With the advent of high-precision and automated mass spectrometry instrumentation to support large-scale proteomic studies, protein turnover analysis at the global level potentially has an increasing importance for biomedical research. One example is the application of protein turnover analysis to study *M. tuberculosis*, especially at its non-replicating or dormant state. Whereas over a hundred research articles have been published on mycobacterial proteomes, only a few dealt with non-replicating *M. tuberculosis* (Cho et al., 2006; Rosenkrands et al., 2002). Information about protein turnover in non-replicating and dormant *M. tuberculosis* is scarce in the literature. With the potential importance of proteome dynamics in bacterial cell sporulation or dormancy (Bernlohr, 1967; Bernlohr, 1972; Mandelstam, 1958; Spudich & Kornberg, 1968), a study of mycobacterial protein turnover at the global level (Rao et al., 2008a; Rao et al., 2008b) will likely help to advance our understanding of the molecular basis of *M. tuberculosis* persistence (Rao & Li, 2009b).

Over the last century, a variety of control and eradication measures have been implemented against tuberculosis such as vaccination, aggressive chemotherapy, and public health surveillance. But tuberculosis still remains a major global health problem to continue to cause nearly 2 million deaths and 9 million new infection cases per year. *Ca.* one third of the world population is infected with *M. tuberculosis*. A majority of the infected individuals remain asymptomatic whereas they carry a lifetime risk to develop an active disease i.e., the latent tuberculosis infection. The state of latency represents the greatest obstacle to eradicate the tuberculosis disease.

The metabolic requirement of *M. tuberculosis* in latency is unclear and difficult to study because the bacilli presumably remain dormant in a granuloma (Pagan-Ramos et al., 2006). The anti-tuberculosis drugs used today have their maximum effect against the growing but not the dormant bacilli. The long therapeutic regime required to treat latent tuberculosis infection is probably explained by the lack of a direct target that is specific to dormant *M. tuberculosis*. There are new drug treatment regimes and several new anti-tuberculosis drugs in a development pipeline that aim to shorten the treatment period and to overcome multi-drug resistant strains (Murphy & Brown, 2008). Most of these new drugs are still based on existing classes of antimicrobial compounds to target the conventional pathways and molecular machinery that are critical for the growth of *M. tuberculosis*. These drugs could still be countered by drug resistant strains that emerge from the non-adherence of a prolonged regime against latent tuberculosis infection. Thus, the need to discover novel drug targets, especially those against dormant bacteria, is urgent.

A 'simple but nonetheless vexing problem' in target discovery against non-replicating *M. tuberculosis* is that many methods rely on a growth-inhibition measurement to assess the effect of drug treatment (Murphy & Brown, 2008). Rao et al. showed that a protein-turnover measurement was much more sensitive than a protein relative abundance measurement alone to uncover protein synthesis activities in *M. tuberculosis* (Rao et al., 2008a); those data suggest that protein dynamics analysis with turnover and abundance measurements could potentially add a valuable alternative to the drug target discovery problem for non-replicating *M. tuberculosis*. The sensitive protein turnover analysis could also be useful to detect and validate drug treatment effects at an early phase, during which the most relevant drug effect can be isolated from other non-specific cell stress responses.

3. Label-free quantitation of proteins

There is an increased use of LC/MS instrumentation for proteomics studies at a large scale. The depth of a proteomic analysis, i.e. the number of protein species that can be precisely identified and characterized in an experiment, depends on the precision and sensitivity of a mass spectrometry instrument. The impact of the precision and sensitivity of an LC/MS instrument on a proteomic study is manifested in many areas of proteomic studies. For example, in the analysis of intracellular bacterial proteomes, FTMS-based approach clearly identified more proteins from scarce and complex intracellular bacterial samples compared to many other proteomic methods.

The precise retrieval of biological information from a large LC/MS dataset critically depends on algorithms for data interpretation, which remains a current bottleneck in the rapid advance of proteomics technology (Mortensen et al., 2010). The quantitation of differentially regulated proteins represents a major type of proteomics application in biological studies. Protein quantitation with LC/MS data includes three conceptually different methods i.e., spectral counting, differential stable isotope labelling, and label-free LC/MS measurements by using extracted ion chromatographic intensities (Mueller et al., 2008). Due to the increased time and complexity of sample preparation in stable isotope labelling, cost of labelling reagents, and requirement of higher starting sample amount, however, researchers are increasingly using label-free proteomics for faster and simpler protein quantitation (Zhu et al., 2010).

Most of proteomics studies infer proteins with ≥2 identified peptides as reliable protein identifications and usually disregard proteins with a single-peptide hit as unreliable for quantitation. This "two-peptide" rule was recently challenged with the evidence that it reduced protein identifications more in a target database than in a decoy database and thus increased false discovery rates in protein identification (Gupta & Pevzner, 2009). Indeed, it was shown that proteins with a single-peptide hit could represent 30% of the proteins identified with ≥2 MS^2 spectrum matches at p <.01 (Li & Roxas, 2009). Because those single-peptide proteins had ≥2 MS^2 spectrum matches (p <.01) in multiple LC/MS analyses under the same condition, they had an adequate level of statistical confidence to be included for quantitation.

But the inclusion of single-peptide proteins in a differential quantitative proteomics analysis raises two issues. The first is that a conventional statistical test such as a t-test can not be applied toward these single-peptide proteins when the t-test relies on multiple quantified peptides as replicates to calculate the t-statistic for the protein relative abundance (Li & Roxas, 2009). The second is that many single-peptide proteins are at a lower abundance and thus noisy. More stringent thresholds are needed to control the false discovery rate when these single-peptide proteins are included for the selection of differentially regulated proteins.

Pavelka et al. applied a power law global error model (PLGEM) and the signal-to-noise ratio (STN) statistic (Pavelka et al., 2004) to select differentially regulated proteins based on a spectral counting quantitation method (Pavelka et al., 2008). The PLGEM-STN statistic utilized a re-sampling approach to estimate the null distribution from replicates of a sample. After the error model was calculated from a pool of re-sampling statistics that constituted the null distribution, a set of STN thresholds was applied at a specified confidence level toward samples with any level of replicates. The PLGEM-STN method is attractive in that it could be applied toward samples with no replicates if several replicates for one sample are

provided to estimate the null distribution. It is also applicable to proteins with any number of identified peptides. The PLGEM-STN method, however, has not been demonstrated for label-free quantitation with extracted ion chromatographic intensities.

In the work described in the following, the PLGEM-STN statistic was applied toward a LC/MS dataset obtained with a high-resolution mass spectrometer (Roxas & Li, 2009). The peptide and protein abundances were quantified with a label-free approach based on extracted ion chromatographic intensities. The false discovery rate was estimated at different confidence levels of the PLGEM-STN statistics.

The PLGEM-STN statistic alone did not provide a desired level of false discovery rate control. Insufficient stringency in false discovery rate control was similar to the situation when a t-test statistic was used alone (Li & Roxas, 2009). With the combination of a t-test and the rule of minimum number of permuted significant pairings (MPSP), however, the false discovery rate was significantly reduced in that study.

The combination of MPSP and PLGEM-STN was further tested to control the false discovery rate. PLGEM-STN does not require that a protein have to have at least two detected peptides for an assessment of statistical significance. Thus, the combination of MPSP and PLGEM-STN has the potential to extend the selection of differentially regulated proteins to those with lower fold-changes and to those with single-peptide hits. Similarly, a fold-change threshold can be applied toward proteins with any number of detected peptides. With a control, the statistical significance of a differentially regulated protein can also be assessed based on a fold-change threshold. Therefore, the combination of MPSP and fold-change thresholds was also tested and compared with the PLGEM-STN-MPSP approach.

It is important to use the high-resolution FTMS instrument to acquire data for the label-free quantitation so that proteins with a single peptide hit can be quantified based on peptide cross reference and extracted ion chromatographic intensity. Proteins with fewer than three peptide hits are typically difficult to quantify by the spectral count method.

3.1 Purpose of the study

There are two purposes to investigate the combination of the PLGEM-STN statistic or a fold-change threshold with the MPSP method to identify differentially regulated proteins. One was to extend the selection of differentially regulated proteins to those that had single-peptide hits. The other was to select differentially regulated proteins at smaller fold-changes and at a false discovery rate ≤.05. The approaches to achieve this two-fold purpose were investigated under a scenario where the number of sample replicates was small. When the number of sample replicates is small, other typical statistics such as a t-test might not perform well to provide the necessary specificity in the label-free quantitation of differentially regulated proteins. Therefore, it was found necessary to insert an additional measure, such as the MPSP rule to compensate for the lack of sample replicates (Li & Roxas, 2009). Even when more sample replicates are available, the additional use of MPSP might still further increase the specificity although this possibility will need to be tested.

3.2 Results

With a null distribution built from the labelled control sample to establish thresholds, different approaches were experimented with to select differentially regulated proteins by using the combination of MPSP, PLGEM-STN, and fold-change methods. Differentially regulated proteins were selected from the unlabeled sample pair S_P and R_P.

The following two subsections of Results are described. Subsection 3.2.1 analyzes the source of variability in the peptide and protein quantitation processes. Subsection 3.2.2 performs multi-step extended selection of differentially regulated proteins.

3.2.1 Source of variability

An observed differential abundance of a PCS or protein between samples arose not only from the difference in biological samples but also from measurement noise that included the variability from multiple steps that involve LC/MS injection replicates, sample preparation replicates, biological replicates, or the data processing method. The multi-step experimental procedures are summarized in **Fig. 5**.

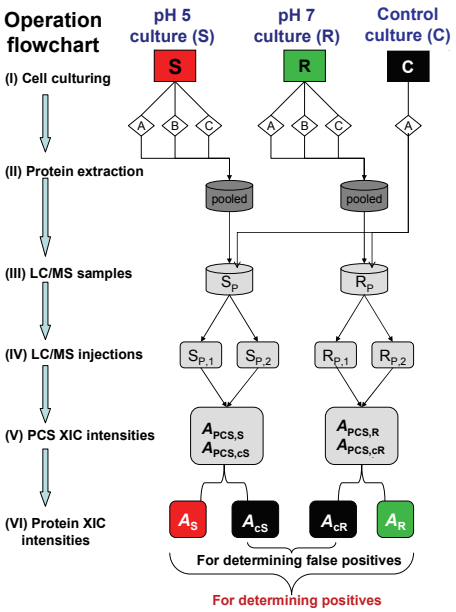


Fig. 5. Experimental outline of the label-free protein quantitation approach to assess the acid stress response between the unlabeled stressed culture (S) and the unlabeled reference culture (R) with the [^{15}N]-labeled culture as control (C)

The biological sample model used in the study was the proteome response of an acid stressed *M. smegmatis* culture (S) in reference to a neutral pH culture (R) (Roxas & Li, 2009) (Fig. 5). Both S and R cultures were unlabeled. The proteins from a [^{15}N]-labelled control culture (C) were used as an internal standard to mix with the proteins from the unlabeled cultures. Because the proteins from the control culture were analyzed repeatedly with two other unlabeled samples, the repeated analyses of the labelled control provided replicates to construct a null distribution.

In the null distribution, there were no true differentially regulated proteins. The null distribution was thus useful to model the noise in the experiment. The error model was derived from the null distribution that consists of at least two replicates of sample preparation. The error model was preferred not to derive from the unlabeled protein

samples S_P and R_P, because each of these two unlabeled samples had only LC/MS run replicates but no sample preparation replicates. The use of only LC/MS replicates to model the noise is likely to underestimate the noise level in the experiment.

The experimental procedures were divided into six stages (I-VI). Briefly, equal amounts of protein extract from the S culture triplicates were pooled. Equal amounts of protein extract from the R culture triplicates were also pooled. Into these two pooled unlabeled protein samples, an equal amount of protein extract from the C culture was added. This resulted in the two pooled samples i.e., S_P and R_P. The proteins differentially expressed between the S and R cultures were determined based on comparison of the abundances of the unlabeled proteins i.e., A_S and A_R, between samples S_P and R_P. For the purpose of false discovery rate assessment, the abundances of the [^{15}N]-labeled proteins i.e., A_{cS} and A_{cR}, were quantified and compared between S_P and R_P in the same way as between A_S and A_R. The proteins found differentially expressed between A_S and A_R were considered positives, because they reflected the difference between the S and R cultures. The proteins found differentially expressed between A_{cS} and A_{cR} in the labeled form were false positives, because difference was not expected from the identical C sample that was run concurrently with two unlabeled samples in separate runs.

To assist in the assessment of the source of variability in the label-free quantitation of the LC/MS data, another three samples were used in addition to S_P and R_P. The three additional samples were the biological replicates of the S culture sample, namely S_A, S_B, and S_C. S_P was generated by pooling S_A, S_B, and S_C.

The 3rd of the five fractions of an SDS/PAGE gel lane was processed for LC/MS analysis for the protein samples S_A, S_B, S_C, S_P, and R_P with duplicate injections for each sample (Li & Roxas, 2009). The five samples with two LC/MS injections per sample resulted in 10 LC/MS runs. These 10 LC/MS runs of the 3rd fraction allowed the quantitation of 349 proteins for the 3rd fraction (Li & Roxas, 2009). Because a protein was quantified in both the unlabeled form (for culture S or R) and the labelled form (for culture C), there were 20 quantitation categories for each protein (**Table 1**). Thus, these 349 proteins and the 20 quantitation categories formed a 349 x 20 matrix. The 349 x 20 matrix was examined by a clustering analysis (Eisen et al., 1998). The clustering analysis provides an overview of the correlation among the protein samples and LC/MS injections, thus reveals the major source of variability.

From the dendrogram of the 20 quantitation categories shown in **Fig. 6**, it could be seen that the distance between each pair of duplicate LC/MS injections was the shortest compared to those between any other sample pairings. The closest distance of the duplicate LC/MS injections for a sample indicated that the variability between LC/MS injections was the smallest, which also indicated that the label-free data analysis methodology (Li & Roxas, 2009) did not introduce a more significant variability.

	Unlabeled protein samples from culture S or R					[^{15}N]-labeled protein samples from control culture C				
	S_P	R_P	S_A	S_B	S_C	cS_P	cR_P	cS_A	cS_B	cS_C
Quantitation category	$S_{P,1}$ $S_{P,2}$	$R_{P,1}$ $R_{P,2}$	$S_{A,1}$ $S_{A,2}$	$S_{B,1}$ $S_{B,2}$	$S_{C,1}$ $S_{C,2}$	$cS_{P,1}$ $cS_{P,2}$	$cR_{P,1}$ $cR_{P,2}$	$cS_{A,1}$ $cS_{A,2}$	$cS_{B,1}$ $cS_{B,2}$	$cS_{C,1}$ $cS_{C,2}$

Table 1. Twenty quantitation categories arising from the duplicate LC/MS analyses of the 3rd gel fraction for samples S_A, S_B, S_C, S_P, and R_P along with the labeled control in them

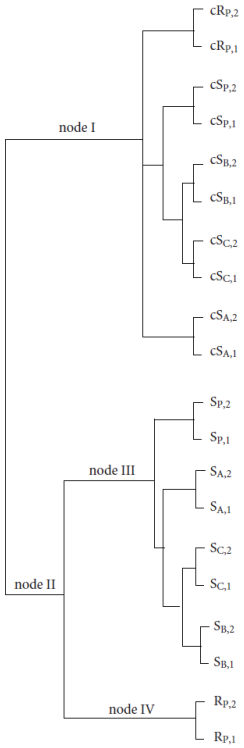


Fig. 6. Clustering of the 20 quantitation categories based on the 349 proteins quantified from the 3rd gel fraction for the five protein samples S_P, R_P, S_A, S_B, and S_C (Li & Roxas, 2009)

In Fig. 6, it was also apparent that the unlabeled and labelled quantitation categories were separated into two distinct branches represented by nodes I and II, respectively. The separation of the unlabeled and labelled quantitation categories into the two distinct clusters indicated that the difference between cultures C and S or C and R was larger than the difference between S and R. From the tree branch under node II, it could be seen that the distance between the unlabeled protein samples S_P and R_P was larger than the distance among the S culture replicates i.e., S_A, S_B, S_C. The result indicated that the difference between cultures S and R exceeded the difference among the S culture replicates, suggesting that the variability in biological sample replicates was less than the actual difference between the biological samples treated with different conditions.

Therefore, the clustering result in Fig. 6 indicated that the variability increased in the order of LC/MS injections < sample preparation replicates (under node I) ~ biological replicates (under node III) < biological samples (between nodes III and IV). Because these differences were evaluated based on the proteomic quantitation data, a variability observed among biological replicates also included the variability introduced during sample preparation for LC/MS analysis. The similarity between the variability observed among the sample preparation replicates and the variability observed among the biological replicates suggested that the variability among biological replicates was not larger than the variability among sample preparation replicates.

3.2.2 Extended selection of differentially regulated proteins

This subsection describes the multiple steps leading to the extended selection of differentially regulated proteins from all quantified proteins including those with only a single-peptide hit. The proteins with a single-peptide hit represent 1/3 of the identified proteins (Li & Roxas, 2009). Therefore, it is desirable to have a procedure to select regulated proteins from all of the proteins including those with a single-peptide hit to maximize the potential of the global protein expression profiling.

Establishing a null distribution

Based on the evaluation with the clustering analysis (Fig. 6), the variability among sample preparation replicates appeared to be comparable with that among biological replicates. Samples S_P and R_P represented the average of triplicate biological replicates for cultures S and R respectively, because each of them was the pooled sample of three biological replicates. The pooling process further reduced the biological variability between S_P and R_P. Therefore, the [^{15}N]-labelled control sample replicates (Table 1) were adequate to represent a null distribution in which there was no differentially regulated protein.

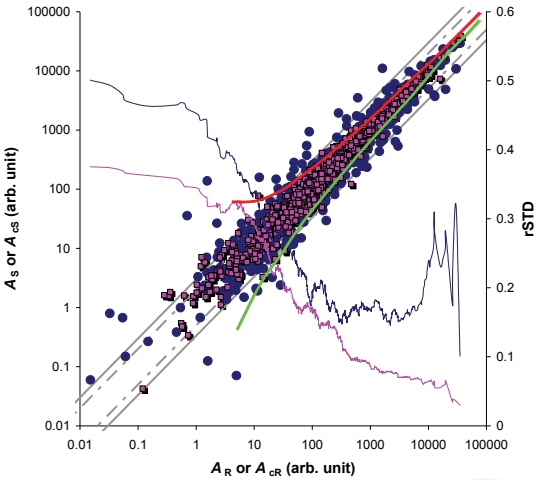


Fig. 7. A_{PRO} scatter plots, local variability, and thresholds for selecting differentially regulated proteins. The blue dots represent the A_{PRO} scatter plot of A_S vs. A_R corresponding to the unlabeled proteins in sample S_P vs R_P. A_S is the average of $A_{S,1}$ and $A_{S,2}$. A_R is the average of $A_{R,1}$ and $A_{R,2}$. The red dots represent the A_{PRO} scatter plot of A_{cS} vs. A_{cR} corresponding to the labeled proteins in control sample replicate cS_P vs cR_P. A_{cS} is the average of $A_{cS,1}$ and $A_{cS,2}$. A_{cR} is the average of $A_{cR,1}$ and $A_{cR,2}$. $A_{S,1}$, $A_{S,2}$, $A_{R,1}$, $A_{R,2}$, $A_{cS,1}$, $A_{cS,2}$, $A_{cR,1}$, and $A_{cR,2}$ were the A_{PRO} values for the eight quantitation categories defined in Table 1. To evaluate the local noise of A_{PRO} measurement, the relative standard deviation (rSTD) for each protein was calculated from its four unlabeled A_{PRO} values $A_{S,1}$, $A_{S,2}$, $A_{R,1}$, and $A_{R,2}$ (the blue trace) or its four labeled A_{PRO} values $A_{cS,1}$, $A_{cS,2}$, $A_{cR,1}$, and $A_{cR,2}$ (the pink trace). The rSTD- A_{PRO} traces were smoothed with a 100-point moving box. The grey straight lines indicated a 3-fold (solid line) and a 2-fold (dashed line) change threshold. The solid red and green curves represent the fold-change thresholds established with the PLGEM-STN statistics based on the local variance in the null distribution (the pink-dot scatter plot)

The null distribution afforded an estimation of measurement noise. The determined measurement noise was then used to estimate the false discovery rate for the selected differentially regulated proteins between samples S_P and R_P. The null distribution provided a reference for setting thresholds to maximize the selection of differentially regulated proteins (positives) while minimizing false positives. In **Fig. 7**, such a null distribution was illustrated with the scatter plot represented by the pink dots.

To investigate the relationship between measurement variability and protein abundance A_{PRO}, relative standard deviation (rSTD) was plotted against the mean A_{PRO} value for each protein in the unlabeled protein samples (blue trace) or the labelled control protein samples (pink trace) (Fig. 7). The rSTD-A_{PRO} trace in pink reflected the local noise of the null distribution. The local noise of the null distribution was mainly due to the variability that was introduced during sample preparation (Fig. 6). The rSTD-A_{PRO} trace in pink clearly suggested that the A_{PRO} measurement noise had a reciprocal dependence on the A_{PRO} amplitude. The rSTD-A_{PRO} trace in blue reflected both sample preparation variability and biological sample difference between cultures S and R. Thus, the blue trace had higher rSTD values than the pink trace throughout the A_{PRO} range.

Modelling local noise in the null distribution

Because of the reciprocal dependence of A_{PRO} rSTD on the A_{PRO} value, a universal 3-fold-change cut-off missed some positives at higher A_{PRO} values where a <3-fold change was already significantly different from the local noise. Missed positives at higher A_{PRO} values could be observed in Fig. 7 by examining the spread of the two scatter plots in the high A_{PRO} ranges. At A_{PRO} >1000, the rSTD was a few times smaller than that at A_{PRO} of ~100. From the figure, it could be seen that it was possible to detect a < 2-fold change for the proteins with A_{PRO} >1000. To the contrary, at A_{PRO} <10, a 3-fold change threshold was not sufficient to eliminate many false positives. Therefore, a criterion adaptive to the dependence of A_{PRO} noise on A_{PRO} values would uncover more differentially regulated proteins. This extended selection of differentially regulated proteins could be achieved by penalizing proteins with higher A_{PRO} values less than proteins with lower A_{PRO} values. Such an adaptive criterion, however, requires a systematic modelling of the noise to establish the thresholds according to local variability (Pavelka et al., 2004).

In this study, the PLGEM-STN statistic was experimented with for the selection of differentially regulated proteins quantified with label-free proteomics based on protein extracted ion chromatographic intensities. There were two reasons for the choice of the PLGEM-STN method.

First, the PLGEM-STN method allowed statistical analyses of the proteins quantified with a single PCS because the PLGEM-STN statistic did not rely on multiple PCSs of a protein like a t-test (Li & Roxas, 2009). Because single-peptide proteins constituted a third of the quantified proteins, being able to quantify these single-peptide proteins was important to maximize the potential value of the data. Second, the PLGEM-STN method took into account the dependence of A_{PRO} noise on A_{PRO} levels. A threshold adjustable to the local dependence of A_{PRO} noise on A_{PRO} levels allowed the selection of differentially regulated proteins with a smaller fold-change threshold at a higher A_{PRO} level.

Therefore, the PLGEM-STN method potentially could select more differentially regulated proteins by applying a smaller fold-change threshold in the higher A_{PRO} range where the variability was smaller. This possibility was tested as described below.

Selecting differentially regulated proteins with PLGEM-STN

PLGEM-STN confidence level	FP, P, and FDR	PLGEM-STN					PLGEM-STN-MPSP
		Permuted sample pairings				Average	
		I	II	III	IV		
0.01	FP (cS_P/cR_P)	31	68	22	46	42	13
	P (S_P/R_P)	141	155	134	148	145	101
	FDR	0.22	0.44	0.16	0.31	0.29	0.13
0.002	FP (cS_P/cR_P)	6	15	3	9	8	2
	P (S_P/R_P)	47	50	46	51	49	44
	FDR	0.13	0.30	0.07	0.18	0.16	0.05

Table 2. Numbers of differentially regulated proteins selected with PLGEM-STN alone or in combination with MPSP

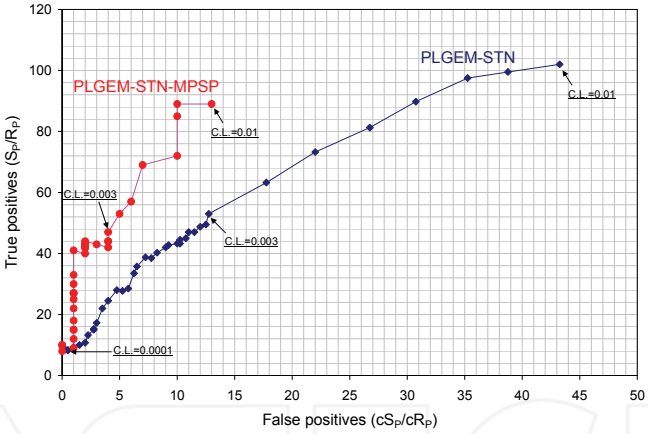


Fig. 8. Receiver operating characteristic analysis of the PLGEM-STN approach with (red curve) or without (blue curve) the combination with MPSP. *Positives* are the differentially regulated proteins selected from the comparison of protein abundances between samples S_P and R_P. *False positives* are the differentially regulated proteins selected from the comparison of proteins abundances between samples cS_P and cR_P. *True positives* are estimated by subtracting false positives from the positives. For each approach, i.e. PLGEM-STN-MPSP or PLGEM-STN, 37 data points at different confidence levels (C.L.) are plotted in this figure, starting from C.L.=0.0001 up to C.L.=0.01. The increment is 0.001 between C.L. of 0.0001 and 0.003 (30 data points). Between C.L. of 0.003 and 0.01, the increment is 0.01 (7 data points)

Table 2 shows the result of the PLGEM-STN analysis for the unlabeled samples S_P and R_P and the labelled sample replicates cS_P and cR_P. cS_P and cR_P were the labelled control samples analyzed concurrently with S_P and R_P, respectively. The differentially regulated proteins

found between S_P and R_P were positives (P), and those found between cS_P and cR_P were false positives (FP). Because each protein sample was analyzed with duplicate LC/MS injections, permutation of the four LC/MS injections for a sample pair resulted in four permuted sample pairings (Li & Roxas, 2009). These four permuted sample pairings were numbered as I to IV in Table 2. In each column for a permuted sample pairing in Table 3, the numbers of false positives and positives and the false discovery rate (FDR) were listed. The false positives were determined as the differentially regulated proteins for the sample pair cS_P/cR_P. The positives were determined as the differentially expressed proteins for the sample pair S_P/R_P.

In Table 2, the positives and false positives were selected with the PLGEM-STN method at the confidence level of 0.01 and 0.002, respectively. The results indicate that the numbers of positives or false positives were not the same among the four permuted sample pairings. To estimate an average false discovery rate, the numbers of positives and false positives were respectively averaged among the four permuted sample pairings. The false discovery rate was then calculated as the ratio of the average number of false positives divided by the average number of positives. The false discovery rate was determined at two different PLGEM-STN confidence levels (Table 2). With a receiver operating characteristic analysis, the PLGEM-STN approach is examined over a broader confidence level range (**Fig. 8**) and will be compared with another approach that is to be described below.

Incorporating the MPSP rule

Initially, the PLGEM-STN approach was carried out by comparing the duplicate LC/MS injections from the two samples R and S without permutation pairings. But the false discovery rate stayed high unless the sensitivity was severely compromised to reduce the false discovery rate. For example, at a confidence level of 0.0001, only 16 differentially regulated proteins were selected at 6% false discovery rate (data not shown). With all of the permutation pairs and a combination of PLGEM-STN and MPSP, 44 differentially regulated proteins were selected at a false discovery rate of 5% (Table 2). Therefore, a high sensitivity is achieved to uncover differentially regulated proteins by utilizing all possible permutation pairs with a combination of PLGEM-STN and MPSP.

Because of the variable numbers of positives and false positives among the four permuted sample pairings, it was necessary to determine a consensus list of differentially regulated proteins from the four permuted sample pairings. Previously, the rule of MPSP was applied to determine the consensus list of differentially regulated proteins from four permuted sample pairings (Li & Roxas, 2009). The MPSP rule required that only those proteins that were found differentially regulated in a certain number of permuted sample pairings were counted as positives (for S_P/R_P) or false positives (for cS_P/cR_P). When a sample pair such as S_P/R_P had no sample replicates but had duplicate LC/MS injections, MPSP was found to be optimum at four (Li & Roxas, 2009). Setting MPSP at four meant that a differentially regulated protein had to be found differentially regulated in all of the four permuted sample pairings.

Selecting differentially regulated proteins with a PLGEM-STN-MPSP approach

The application of the MPSP rule towards the PLGEM-STN results decreased both false positives and positives (Table 2). But the false discovery rate was also decreased relative to that when only the PLGEM-STN statistic was applied. From Table 2, it could be seen that the number of true positives, which was estimated from the difference between the numbers of

positives and false positives, remained about the same. Therefore, the combination of the MPSP rule with the PLEGM-STN method reduced the false discovery rate by 2-3 times without compromising the sensitivity.

As summarized in Fig. 8, the receiver operating characteristic analysis clearly shows that the PLGEM-STN-MPSP approach significantly reduces false positives to improve the specificity without significantly affecting the sensitivity. Compared to the use of the PLGEM-STN statistic alone, the combination of PLGEM-STN and MPSP performs better in controlling false discovery rates without compromising the sensitivity to select differentially regulated proteins.

Selecting differentially regulated proteins with a fold-change-MPSP approach

Fold change	FP, P, and FDR	Fold-change					Fold-change-MPSP
		Permuted sample pairings				Average	
		I	II	III	IV		
2	FP (cS_P/cR_P)	68	77	118	45	77	22
	P (S_P/R_P)	171	154	186	147	165	104
	FDR	0.40	0.50	0.63	0.31	0.47	0.21
3	FP (cS_P/cR_P)	30	33	47	20	33	9
	P (S_P/R_P)	66	70	85	60	70	42
	FDR	0.45	0.47	0.55	0.33	0.47	0.21
4	FP (cS_P/cR_P)	17	24	32	10	21	1
	P (S_P/R_P)	42	50	53	35	45	26
	FDR	0.40	0.48	0.60	0.29	0.47	0.04

Table 3. Number of differentially regulated proteins selected with a fold-change threshold alone or in combination with MPSP

The use of MPSP with fold-change criteria was also examined (**Table 3**). With fold-change criteria alone, the false discovery rate did not drop below 46% at 2- to 4-fold changes. With the combination of MPSP and the fold-change criteria, the false discovery rate was reduced from 46% to 21% at 2- and 3-fold changes. At a 4-fold change, the false discovery rate was reduced to 4%. Compared to the combination of PLGEM-STN and MPSP, however, the combination of fold-change and MPSP reduced more true positives at the similar false discovery rate of 4-5%. Therefore, the application of MPSP with the fold-change criteria reduced sensitivity. The reduced sensitivity was due to the increase in the fold-change threshold.

With the 4-fold-change-MPSP and the PLGEM-STN-MPSP approaches, 26 and 44 proteins were respectively selected as differentially regulated at a false discovery rate of 4% or 5% (Tables 2 and 3). Among these 26 and 44 proteins, there were 55 unique proteins (Li, 2010b). These 55 unique proteins included all of the 20 high-confidence differentially regulated proteins identified previously with an empirical fold-change and abundance level cut-off approach (Roxas & Li, 2009).

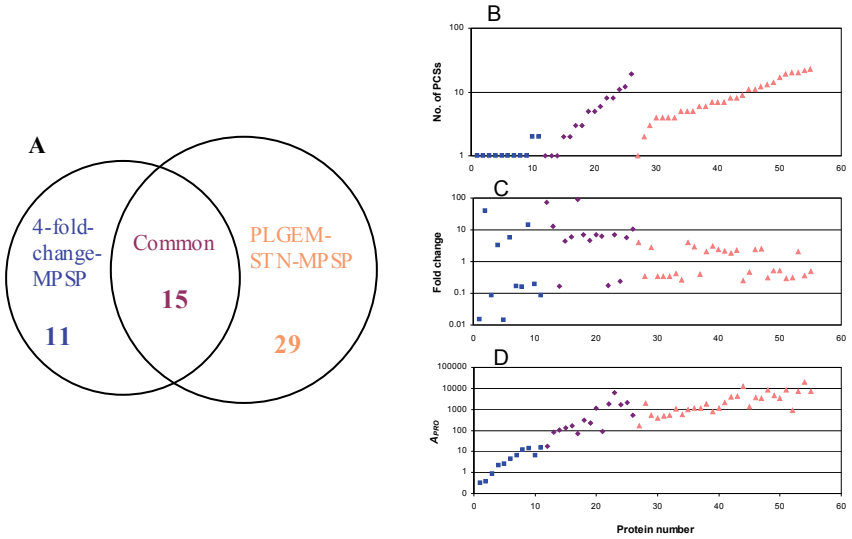


Fig. 9. Comparison of the 26 and 44 differentially regulated proteins respectively selected by the 4-fold-change-MPSP and PLGEM-STN-MPSP approaches at a 5% false discovery rate. (A) Overlap of the two sets of differentially regulated proteins. Panels B-D show the distributions of (B) the number of detected PCSs, (C) the fold changes, and (D) the abundances of the quantified proteins. The blue square, the purple diamond, and the tan triangle markers represent the differentially regulated proteins selected by 4-fold-change-MPSP only, by both, and by PLGEM-STN-MPSP respectively. The protein number was from 1 to 55 on the x-axis representing the 55 unique proteins ranked according to their A_{PRO} in each of the three groups (blue, purple, or tan)

Comparing the PLGEM-STN-MPSP and fold-change-MPSP approaches

Only 15 proteins were common between the two sets of differentially regulated proteins selected with the 4-fold-change-MPSP and the PLGEM-STN-MPSP approaches (**Fig. 9A**). The 4-fold-change-MPSP approach selected more single-PCS proteins than the PLGEM-STN-MPSP approach (Fig. 9B). The PLGEM-STN-MPSP approach selected proteins with a fold-change as low as 1.8-fold (Fig. 9C). However, these differentially regulated proteins selected with PLGEM-STN-MPSP had a protein abundance higher than most of the differentially regulated proteins selected with the 4-fold-change-MPSP approach (Fig. 9D). Thus, the two approaches complement each other and could be used simultaneously.

3.3 Discussions

3.3.1 Motivation of the extensive label-free quantitative proteomics analysis

Despite the relative complexity in label-free proteomics data analysis and the demand of more stringently controlled LC/MS experimental conditions, there are strong motivations stemming from biological and experimental perspectives to use the label-free approach, as discussed below.

As shown in Fig. 6, the unlabeled and labelled quantitation categories are separated into two distinct clusters. One includes the quantitation categories from the labelled control culture C

(under node I). The other includes the quantitation categories from the two unlabeled cultures S and R (under node II). Thus, there was a larger difference between the labelled (C) and either of the two unlabeled samples (S or R) than between the two unlabeled cultures (S and R). The number of differentially regulated proteins between the labelled culture and either of the unlabeled culture was about three times as many as that between the two unlabeled cultures. Compared to the difference between the two unlabeled cultures, the difference between the labelled culture and either of the unlabeled cultures was larger. This larger difference was probably because the labelled culture was cultured in a synthetic minimal medium while the two unlabeled cultures were grown in a commercial 7H9 broth that was richer in ingredients. Another factor was that the acidic growth condition was a relatively mild stress so that not many proteins were differentially regulated.

The apparent difference in proteome profile for cells cultured in different media is actually a strong motivation for this study. In microbiological works, it is not always convenient to make a [^{15}N]-labelled medium with complex ingredients required to cultivate bacteria under more physiologically relevant conditions. Even some of the stable-isotope-labelled media are technically feasible to make, they often bear a costly price tag. For microbiological works, one might not want to be restricted by the type of medium that can be used because of the stable isotope labelling limitation. For example, some mycobacteria are difficult to cultivate on simple synthetic media and prefer complex media. Thus, unlabeled media are always convenient choices if the down-stream proteomic analysis is established to proceed with the quantitation.

For such reasons, the focus of this study was on the comparison of protein expression profiles between the two unlabeled cultures S and R. The labelled control culture C was used as an internal standard to estimate false discovery rates.

3.3.2 The use of a [^{15}N]-labeled internal standard for null distribution construction

The label-free quantitation scheme presented in this study incorporated a labelled internal control to provide replicates for noise modelling without a requirement of other unlabeled sample replicates. The inclusion of a labelled internal control facilitates the estimation and control of false discovery rates.

Internal standards are commonly used to improve reliability of quantitative proteomics such as to aid in removing outlier data and to detect fluctuation in instrument performance (Mirzaei et al., 2009). Compared to other synthetic peptide internal standards(Mirzaei et al., 2009; Winter et al., 2010), the [^{15}N]-labelled control culture C provides more comprehensive peptide internal standards. For most of the peptides, the extracted ion chromatographic intensities can be matched among the three protein samples originated from the two unlabeled (S and R) cultures and the labelled (C) culture. The C protein sample was mixed and run together with either S or R protein sample, so that the reliability of the internal standards was improved.

To construct the null distribution for the error model in PLGEM-STN, it would be ideal to have the labelled internal standard identical to an unlabeled sample in protein composition. As mentioned above, however, that requirement could restrict the culturing conditions available for biological experiments. Thus, it is acceptable and sometimes necessary to use a labelled protein mixture sample as internal standard, even though the internal standard sample might be somewhat different from the unlabeled samples in protein abundance profiles.

Nevertheless, the null distribution is only utilized to establish the relation between the signal-to-noise ratio and the peptide abundance in the PLGEM-STN method. There is no requirement of direct one-to-one comparison between the labelled and unlabeled version of a protein during this process. Therefore, the difference in proteome composition between the labelled internal standard sample C and the two unlabeled samples S and R is not expected to affect the modelling parameters derived from the null distribution constructed from the labelled C sample.
One could choose to run multiple replicates of an unlabeled sample and use the replicates to construct the null distribution. That approach would require more LC/MS runs as discussed previously (Li & Roxas, 2009).

3.3.3 Label-free data analyses and selection of differentially regulated proteins

The LC/MS data used in this work was acquired with a high-resolution mass spectrometer that resolved peptide peaks from a complex sample mixture to allow the determination of the extracted ion chromatographic intensities of peptides and proteins. Repeated LC/MS injections showed the highest reproducibility among several other types of replicates, indicating that the major variability of the label-free quantitation did not lie within the LC/MS separation and the data analysis method. Rather, sample preparation replicates represented a major source of the variability. With a labelled control sample to run concurrently with each of the unlabeled samples, replicates for the labelled control sample were obtained. The control sample replicates provided data to model the noise in the label-free quantitation with extracted ion chromatographic intensities.
We performed a two-step normalization procedure in which the information about the abundance of a peptide or protein in a sample was preserved (Li, 2010b). The preservation of the information about the abundance of a peptide or protein in the samples is critical for performing the PLGEM-STN analysis. In addition, because protein extracted ion chromatographic intensity was represented by the sum of the PCS extracted ion chromatographic intensities belonging to that protein, the summation weighed the low-intensity PCSs less than the high-intensity PCSs. Such a summation of PCS extracted ion chromatographic intensities probably suppressed noise from lower-intensity PCSs. When a protein abundance ratio is calculated as the average of PCS abundance ratios without weighing, the noise from a lower-intensity PCS would be amplified. We have avoided this potential issue by summing the PCS intensities to represent protein abundances before calculating protein abundance ratios.
Single-peptide proteins made up about 35% of the quantified proteins (Li, 2010b). Selection of differentially regulated proteins from these single-peptide proteins required a significance assessment method that did not rely on multiple-peptide detection to calculate a statistic about the confidence of a protein differential abundance. The use of a statistic that does not rely on the detection of multiple peptides is especially useful when the sample replicates were too low to use a typical statistical test such as a t-test. PLGEM-STN was a method that fits this criterion.
However, PLGEM-STN alone was not strict enough to control the false discovery rate without further diminishing the number of positives (Fig. 8). The lack of stringency by using the PLGEM-STN method alone was similar to that by using the t-test alone (Li & Roxas, 2009). In that prior study, the lack of specificity with a t-test alone was overcome by introducing the rule MPSP. The MPSP rule simply requires that a protein be selected as differentially regulated only when it was repeatedly found so in certain number of

permuted sample pairings. The MPSP rule was introduced to deal with datasets with small replicates where other more sophisticated statistical tests could not be applied (Li & Roxas, 2009). Although the MPSP rule was originally used in combination with a t-test statistic and a fold-change threshold, this study shows that it can be used in combination with other types of statistical tests such as the PLGEM-STN method (Fig. 8).

The combination of the MPSP rule allowed the selection of differentially regulated proteins at a false discovery rate <5%, which would have been impossible for a fold-change method (Table 3). The MPSP rule significantly reduced false positives while keeping the number of true positives relatively constant, thus effectively improving the statistical confidence of the selected differentially regulated proteins by lowering the false discovery rate (Table 3). The results suggest that MPSP is a rule that can be used in combination with different types of statistics to select differentially regulated proteins.

The label-free quantitation simplified cell culturing and sample preparation. Another useful aspect of the label-free quantitation is that peptide cross-reference could be used to increase the number of proteins quantified in all of the samples run under the same condition (Andreev et al., 2007). Lipton et al. introduced the concept of accurate mass and elution time peptide tag for global protein quantitation using high resolution mass spectrometry (Lipton et al., 2002). One advantage of this method over using the spectral counting method is that the large number of identifications that occur in a LC/MS injection can be used as the basis for improved quantitation of another LC/MS injection (Andreev et al., 2007; Fang et al., 2006; Strittmatter et al., 2003). The accurate mass and elution time peptide tag approach uses the extracted ion chromatographic intensities as the quantitative measurement of peptides and proteins. The linear response of peptide extracted ion chromatographic intensities to protein quantities was demonstrated (Hochleitner et al., 2005; Lundrigan et al., 1997; Wang et al., 2006). This method was thus used to improve the comparability of proteins quantified between samples, among LC/MS injections, and for different isotopic forms of a protein (Rao et al., 2008a). The quantitation of 349 proteins from a single gel fraction for several samples clearly demonstrated the power of the peptide cross-reference feature in extracted ion chromatographic intensity-based label-free quantitative proteomics (Li & Roxas, 2009).

One drawback of extracted ion chromatographic intensity-based label-free quantitative proteomics is that the success of an analysis critically depends upon the reproducibility of LC/MS runs that have to be maintained across multiple samples. The reproducibility of LC/MS runs across multiple samples is a prerequisite to reliable peptide cross reference (Andreev et al., 2007). With the advancement in LC/MS instrumentation and the availability of improved LC/MS chromatogram alignment methods (Fischer et al., 2006; Podwojski et al., 2009), the reproducibility of LC/MS runs is unlikely to remain an obstacle for the increasing use of label-free quantitative proteomics.

3.4 Summary

A label-free quantitative proteomics scheme was demonstrated to select differentially regulated proteins with single-peptide hits and <2-fold changes at a 5% false discovery rate. The scheme incorporated a labeled internal control into multiple unlabeled samples to facilitate error modeling when there were no replicates for the unlabeled samples. The error modeling allowed the use of the PLGEM-STN statistic to facilitate the selection of differentially regulated proteins with single-peptide hits. The PLGEM-STN statistic also facilitated the selection of differentially regulated proteins at different fold-change

thresholds according to the local abundance level of the proteins. While the PLGEM-STN statistic uncovered more differentially regulated proteins at higher abundance with smaller fold-changes, the PLGEM error modeling of local variance versus abundance over-penalized the proteins with lower abundance. With a constant fold-change threshold, however, differentially regulated proteins with higher abundance were overlooked. Thus, the results from this study showed that the PLGEM-STN and a constant fold-change threshold were complementary to each other and could be used simultaneously. But, neither the PLGEM-STN nor the 4-fold-change criterion alone was stringent enough for selecting differentially regulated proteins at a 5% false discovery rate.

MPSP was introduced and shown to be a rule that could decrease false discovery rates when being used in combination with the PLGEM-STN statistic or the 4-fold-change threshold. The MPSP rule played a critical role in extending the selection of differentially regulated proteins to those with single-peptide hit or with a lower fold-change in label-free proteomics when sample replicates were limited. Although the approaches were demonstrated for a representative replicate-limited scenario, they potentially can also be applicable to a situation where more sample replicates are available.

4. Conclusion

This chapter presents several examples of proteomic studies utilizing the LTQ-FTMS instrument. The instrument typically provides a high resolution (> 500,000), a large mass range (one order magnitude in a single scan), and high mass accuracy (< 2 ppm) in many experiments.

The protein turnover studies benefit greatly from the high resolution in a large mass range, which allows the resolution of isotopomers and isotopologue profiles that could possibly be generated by almost any degree and type of stable isotope labeling. The well resolved full-range spectra simplify the data processing steps and improve data quality. Because a spectral count method reports the MS^2 events mostly on monoisotopic peaks, it is usually not suitable for protein turnover study. Therefore, a high-resolution mass spectrometer such as an FTMS instrument is highly desired for protein turnover studies.

The high mass accuracy is critical in the implementation of a label-free proteomics approach. The label-free quantitative proteomics relies on a cross reference of peptides between runs and an integration of extracted ion chromatographic intensities. The peptide cross reference allows the quantitation of a peptide in a run in which the peptide is not identified but is identified in another run. It is based on the accurate mass and reproducible liquid chromatographic elution time of the peptide in multiple runs. The cross reference allows more peptides to be quantified. It also improves the comparability of samples because the same peptides, thus the same proteins, are quantified in multiple samples.

The label-free quantitation is useful not only for protein differential expression studies, but also for proteome dynamics studies where the abundances of newly synthesized proteins are to be quantified separate from those of the total proteins and the old proteins.

Thus, the label-free quantitation approach is universally applicable to different proteomic studies. The assessment of statistical significance in such a quantitation approach is challenging especially for the proteins with few peptides identified and when the replicates are limited. A combination of the MPSP rule with a global error modeling method and a fold-change threshold proves to be effective in controlling the false discovery rates in protein differential analysis.

5. References

Andreev, V. P.; Li, L.; Cao, L., et al. (2007). A new algorithm using cross-assignment for label-free quantitation with LC-LTQ-FT MS. *J Proteome Res,* 6, 6, 2186-94.

Bernlohr, R. W. (1967). Changes in amino acid permeation during sporulation. *J Bacteriol,* 93, 3, 1031-44.

Bernlohr, R. W. (1972). 18 Oxygen probes of protein turnover, amino acid transport, and protein synthesis in Bacillus licheniformis. *J Biol Chem,* 247, 15, 4893-9.

Beynon, R. J. (2005). The dynamics of the proteome: strategies for measuring protein turnover on a proteome-wide scale. *Brief Funct Genomic Proteomic,* 3, 4, 382-90.

Borek, E.; Ponticorvo, L. & Rittenberg, D. (1958). Protein Turnover in Micro-Organisms. *Proc Natl Acad Sci U S A,* 44, 5, 369-74.

Cargile, B. J.; Bundy, J. L.; Grunden, A. M. & Stephenson, J. L., Jr. (2004). Synthesis/degradation ratio mass spectrometry for measuring relative dynamic protein turnover. *Anal Chem,* 76, 1, 86-97.

Cho, S. H.; Goodlett, D. & Franzblau, S. (2006). ICAT-based comparative proteomic analysis of non-replicating persistent Mycobacterium tuberculosis. *Tuberculosis (Edinb),* 86, 6, 445-60.

Doherty, M. K. & Beynon, R. J. (2006). Protein turnover on the scale of the proteome. *Expert Rev Proteomics,* 3, 1, 97-110.

Eisen, M. B.; Spellman, P. T.; Brown, P. O. & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A,* 95, 25, 14863-8.

Fang, R.; Elias, D. A.; Monroe, M. E., et al. (2006). Differential label-free quantitative proteomic analysis of Shewanella oneidensis cultured under aerobic and suboxic conditions by accurate mass and time tag approach. *Mol Cell Proteomics,* 5, 4, 714-25.

Fischer, B.; Grossmann, J.; Roth, V., et al. (2006). Semi-supervised LC/MS alignment for differential proteomics. *Bioinformatics,* 22, 14, e132-40.

Gerner, C.; Vejda, S.; Gelbmann, D., et al. (2002). Concomitant determination of absolute values of cellular protein amounts, synthesis rates, and turnover rates by quantitative proteome profiling. *Mol Cell Proteomics,* 1, 7, 528-37.

Goldberg, A. L. & Dice, J. F. (1974). Intracellular protein degradation in mammalian and bacterial cells. *Annu Rev Biochem,* 43, 0, 835-69.

Gupta, N. & Pevzner, P. A. (2009). False discovery rates of protein identifications: a strike against the two-peptide rule. *J Proteome Res,* 8, 9, 4173-81.

Hochleitner, E. O.; Kastner, B.; Frohlich, T., et al. (2005). Protein stoichiometry of a multiprotein complex, the human spliceosomal U1 small nuclear ribonucleoprotein: absolute quantification using isotope-coded tags and mass spectrometry. *J Biol Chem,* 280, 4, 2536-42.

Larrabee, K. L.; Phillips, J. O.; Williams, G. J. & Larrabee, A. R. (1980). The relative rates of protein synthesis and degradation in a growing culture of Escherichia coli. *J Biol Chem,* 255, 9, 4125-30.

Li, Q. (2010a). Advances in protein turnover analysis at the global level and biological insights. *Mass Spectrom Rev,* 29, 5, 717-36.

Li, Q. (2010b). Assigning significance in label-free quantitative proteomics to include single-peptide-hit proteins with low replicates. *Int J Proteomics,* 2010, 731582, 15 pages.

Li, Q. & Roxas, B. A. (2009). An assessment of false discovery rates and statistical significance in label-free quantitative proteomics with combined filters. *BMC Bioinformatics,* 10, 43.

Lipton, M. S.; Pasa-Tolic, L.; Anderson, G. A., et al. (2002). Global analysis of the Deinococcus radiodurans proteome by using accurate mass tags. *Proc Natl Acad Sci U S A,* 99, 17, 11049-54.

Lundrigan, M. D.; Arceneaux, J. E.; Zhu, W. & Byers, B. R. (1997). Enhanced hydrogen peroxide sensitivity and altered stress protein expression in iron-starved Mycobacterium smegmatis. *Biometals,* 10, 3, 215-25.

Malen, H.; Berven, F. S.; Fladmark, K. E. & Wiker, H. G. (2007). Comprehensive analysis of exported proteins from Mycobacterium tuberculosis H37Rv. *Proteomics,* 7, 10, 1702-18.

Mandelstam, J. (1958). Turnover of protein in growing and non-growing populations of Escherichia coli. *Biochem J,* 69, 1, 110-9.

Mirzaei, H.; Brusniak, M. Y.; Mueller, L. N., et al. (2009). Halogenated peptides as internal standards (H-PINS): introduction of an MS-based internal standard set for liquid chromatography-mass spectrometry. *Mol Cell Proteomics,* 8, 8, 1934-46.

Mortensen, P.; Gouw, J. W.; Olsen, J. V., et al. (2010). MSQuant, an open source platform for mass spectrometry-based quantitative proteomics. *J Proteome Res,* 9, 1, 393-403.

Mueller, L. N.; Brusniak, M. Y.; Mani, D. R. & Aebersold, R. (2008). An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *J Proteome Res,* 7, 1, 51-61.

Murphy, D. J. & Brown, J. R. (2008). Novel drug target strategies against Mycobacterium tuberculosis. *Curr Opin Microbiol,* 11, 5, 422-7.

Pagan-Ramos, E.; Master, S. S.; Pritchett, C. L., et al. (2006). Molecular and physiological effects of mycobacterial oxyR inactivation. *J Bacteriol,* 188, 7, 2674-80.

Pavelka, N.; Fournier, M. L.; Swanson, S. K., et al. (2008). Statistical similarities between transcriptomics and quantitative shotgun proteomics data. *Mol Cell Proteomics,* 7, 4, 631-44.

Pavelka, N.; Pelizzola, M.; Vizzardelli, C., et al. (2004). A power law global error model for the identification of differentially expressed genes in microarray data. *BMC Bioinformatics,* 5, 203.

Podwojski, K.; Fritsch, A.; Chamrad, D. C., et al. (2009). Retention time alignment algorithms for LC/MS data must consider non-linear shifts. *Bioinformatics,* 25, 6, 758-64.

Pratt, J. M.; Petty, J.; Riba-Garcia, I., et al. (2002). Dynamics of protein turnover, a missing dimension in proteomics. *Mol Cell Proteomics,* 1, 8, 579-91.

Rao, P. K. & Li, Q. (2009a). Principal Component Analysis of Proteome Dynamics in Iron-Starved Mycobacterium Tuberculosis. *J Proteomics Bioinform,* 2, 19-31.

Rao, P. K. & Li, Q. (2009b). Protein turnover in mycobacterial proteomics. *Molecules*, 14, 9, 3237-58.

Rao, P. K.; Rodriguez, G. M.; Smith, I. & Li, Q. (2008a). Protein dynamics in iron-starved Mycobacterium tuberculosis revealed by turnover and abundance measurement using hybrid-linear ion trap-Fourier transform mass spectrometry. *Anal Chem,* 80, 18, 6860-9.

Rao, P. K.; Roxas, B. A. & Li, Q. (2008b). Determination of global protein turnover in stressed mycobacterium cells using hybrid-linear ion trap-fourier transform mass spectrometry. *Anal Chem,* 80, 2, 396-406.

Rosenkrands, I.; Slayden, R. A.; Crawford, J., et al. (2002). Hypoxic response of Mycobacterium tuberculosis studied by metabolic labeling and proteome analysis of cellular and extracellular proteins. *J Bacteriol,* 184, 13, 3485-91.

Roxas, B. A. & Li, Q. (2009). Acid stress response of a mycobacterial proteome: insight from a gene ontology analysis. *Int J Clin Exp Med,* 2, 309-328.

Spudich, J. A. & Kornberg, A. (1968). Biochemical studies of bacterial sporulation and germination. VII. Protein turnover during sporulation of Bacillus subtilis. *J Biol Chem,* 243, 17, 4600-5.

Strittmatter, E. F.; Ferguson, P. L.; Tang, K. & Smith, R. D. (2003). Proteome analyses using accurate mass and elution time peptide tags with capillary LC time-of-flight mass spectrometry. *J Am Soc Mass Spectrom,* 14, 9, 980-91.

Vogt, J. A.; Hunzinger, C.; Schroer, K., et al. (2005). Determination of fractional synthesis rates of mouse hepatic proteins via metabolic 13C-labeling, MALDI-TOF MS and analysis of relative isotopologue abundances using average masses. *Anal Chem,* 77, 7, 2034-42.

Wang, G.; Wu, W. W.; Zeng, W.; Chou, C. L. & Shen, R. F. (2006). Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: Reproducibility, linearity, and application with complex proteomes. *J Proteome Res,* 5, 5, 1214-23.

Wilkinson, K. D. (2005). The discovery of ubiquitin-dependent proteolysis. *Proc Natl Acad Sci U S A,* 102, 43, 15280-2.

Winter, D.; Seidler, J.; Kugelstadt, D., et al. (2010). Minimally permutated peptide analogs as internal standards for relative and absolute quantification of peptides and proteins. *Proteomics,* 10, 7, 1510-4.

Zhu, W.; Smith, J. W. & Huang, C. M. (2010). Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol,* 2010, 840518, 6 pages.

4

Application of Fourier Transform Mid-Infrared Spectroscopy (FTIR) for Research into Biomass Feed-Stocks

Gordon G. Allison
Institute of Biological, Environmental and Rural sciences, Aberystwyth University
United Kingdom

1. Introduction

Non-food biomass crops e.g. switchgrass (*Panicum virgatum L.*), *Miscanthus* x *giganteus*, and short-rotation coppice poplar (*Poplus* spp.) and willow (*Salix* spp.) offer a sustainable source of energy and platform chemicals (Sims et al., 2006). The majority of the energy stored in the crop biomass is in the cell wall which constitutes the largest fraction of lignocellulosic biomass. The three polymers that constitute the bulk of plant cell wall (cellulose, hemicellulose and lignin) rank amongst the most abundant biopolymers on the planet and their proportional concentrations range generally between 40 – 50%, 10 – 40% and 5 – 30% of biomass by weight respectively (McKendry, 2002). The absolute and relative concentrations of the components of the cell wall have a great influence on biomass quality i.e. its suitability for conversion to heat, power and chemical products. However, because biomass can be utilised by a number of conversion routes with differing feedstock demands measures of feed-stock quality are often quite specific to how the material is to be used. For example, biomass can be processed thermochemically. These routes include combustion or co- combustion with coal to generate heat and electricity (Allison et al., 2010). Alternatively, biomass can be converted by fast pyrolysis to bio-char, which is receiving much attention as a soil improver and a means of sequestering carbon from the atmosphere into the soil (Laird, 2008; Woolf et al., 2010), and bio-oil, a liquid fuel (Bridgwater, 2003; Mohan et al., 2006). Biomass can also be gasified to produce a combustible gas which has application for the generation of heat and power, and for the chemical synthesis of liquid transport fuels and industrial chemicals (Ptasinski et al., 2007). These thermochemical processes demand feed-stocks with low moisture content and high energy density, often equating with high levels of the poly aromatic polymer, lignin.

In contrast, non-thermochemical processes e.g. the production of bioethanol and industrial platform chemicals by biological conversion processes are often inhibited by high levels of lignin. High concentrations of lignin in the feedstock necessitate harsh chemical and heat pre-treatments of the biomass prior to enzymic saccharification. This increases energy inputs and often damages the polysachharide components of the cell wall giving rise to inhibitory products (Carroll and Somerville, 2009; Chang, 2007; Fahmi et al., 2007; Fahmi et al., 2008; Grabber, 2005). There is therefore considerable pressure to optimise feedstock composition and at present the most feasible way to achieve this at a commercial scale is by breeding

improvement (Clifton-Brown et al., 2008), although agronomic practice may also influence composition (Hodgson et al., 2010). Crop breeding places huge demands on the analyst in terms of large sample numbers and there is a demand for methods that can cope with high rates of sample throughput whilst being of low unit cost.

Biomass is traditionally assessed in terms of cell wall fibre content measured by direct, or as they are often called, gravimetric methods. The different types of fibre are isolated by successively harsh chemical treatments which selectively and sequentially remove the different classes of structural carbohydrates until only lignin remains. The proportion of each fibre fraction in the biomass is quantified by sample weight change. Measurements of this type are not entirely quantitative as treatments are not wholly selective, and it is often impossible to directly compare cell wall parameters measured by different methods (Hatfield et al., 1999). In addition, these procedures are time consuming, costly and of low through put (Giger-Reverdin, 1995). The acetyl bromide method for the quantification of lignin is an indirect method that has gained in popularity over recent years. The method was first published by Johnson et al. in 1961 (Johnson et al., 1961) and a modified version of the method has been used to analyse lignin in a wide range of species (Fukushima and Dehority, 2000) from relatively small samples of tissue. Lignin is dissolved from purified cell wall material by extraction into a solution of acetyl bromide in concentrated acetic acid at 50˚C, reacted with hydroxylamine and quantified by absorbance change at 280 nm. Quantification requires reference to a standard curve that is set up using known amounts of standard lignin extracted from similar plant material with acetyl bromide (Fukushima and Dehority, 2000) or acidic dioxane (Fukushima and Hatfield, 2001). The concentrations of lignin detected in samples by this method are comparable with those obtained using the widely accepted gravimetric Klason method (Hatfield and Fukushima, 2005) and the method has the advantage that it is more readily worked into a high-throughput scheme (Foster et al., 2010). Adapting direct and indirect methods for high rates of sample throughput is often complex and most likely requires investment in expensive robotised laboratory equipment that may be beyond the resources of many groups (Foster et al., 2010). Consequently many researchers have turned to less expensive technology for new high-throughput methods.

Infrared (IR) spectrometry offers researchers and breeders an alternative approach that is robust and rapid. The majority of the carbon based molecules in animals and plants are highly active in the IR. At its simplest mid-IR spectroscopy is a useful analytical approach in its own right that provides structural information on samples of pure compounds. Generally, the approach taken is to correlate IR spectra with analytical data obtained from the same samples using multivariate regression methods. This approach although seemingly complex overcomes possible nonlinear relationships between absorption and concentration that are encountered in the IR. Such deviations from the Beer-Lambert law may be caused by spectral shifts due to hydrogen bonding, co-variance with other components in the sample and poor design in older instruments (Hsu, 1997). The regression models produced can be then applied to predict the concentrations of the cell wall components in new samples. Properly executed, this approach has been shown to be rapid and robust and therefore reduces the requirement for standard chemical analysis.

Two types of IR spectroscopy have found application for the measurement of chemical composition in biomass: mid-infrared spectroscopy and near infrared reflectance spectroscopy (NIRS). NIRS has a longer history of being used to predict chemical composition in bulk plant samples as until recently the method was more applicable to bulk

plant samples. Mid-IR spectroscopy in contrast to NIRS informs on fundamental molecular vibrations, rather than on harmonic and overtone absorptions, and for this reason gives a better insight into the molecular bonds present in the sample. Over recent years the technique has been revolutionised by the development of Fourier transform instruments (FTIR) and by improvements in methods for sample presentation and many groups are choosing analysis by FTIR in preference to NIRS. Several reports have shown that the combination of FTIR and multivariate analysis is highly effective. For example, spectra of wood can be used to discriminate between tree species (Huang et al., 2008). Recently, we reported on using FTIR to predict the concentration of lignin and hydroxycinnamic acids (Allison et al., 2009), nitrogen and alkali index (Allison et al., 2009) in samples of grasses by partial least squares regression. In addition to the analysis of bulk samples, FTIR has long had application in the field of microscopy. For example, FTIR microscopy has been used to identify and characterise cell wall mutants and transgenic plants altered in cell wall biosynthetic genes (Chen et al., 1998; McCann et al., 2007; Mouille et al., 2003; Stewart et al., 1997).

Although it is not intended to discuss FTIR microscopy in depth in this chapter because it is not a method of choice for high-throughput screening it deserves some brief discussion and comparison with allied methods. For the study of biological samples FTIR microscopy is often used as a transmission technique where the IR beam passes through the sample. Special objectives are available (see Section 4) which work through contact with the sample but this results in damage to most biological specimens. FTIR microscopy is best suited to the high resolution mapping of chemical composition at the level of the cell. In this type of study high rates of sample through-put and the prediction of composition in bulk samples are not usually experimental objectives. Earlier FTIR microscope instruments took a long time to acquire spectra because the beam passing through the sample filled only part of the detector. New developments in FTIR microscope detector design have decreased spectral acquisition time substantially and many groups have taken advantage of focal plane array (FPA) detectors. These were originally developed for military use and they allow the simultaneous acquisition of many spectra. Over the last decade, FPA detectors have grown from arrays of 8x8 elements to 16 x 16 and now 64 x 64 or even 128 x 128 element detectors are available for groups with sufficient funds. FPA-FTIR microscopes allow very rapid chemical mapping over large areas but the technique is not without fundamental problems. FTIR microscopy is incompatible with water, which prevents the analysis of living samples, and the analyst must be judicious in the use of chemical fixatives to preserve the structure of sectioned materials as the absorption peaks from these substances may mask important features in the spectral data. Furthermore, spatial resolution in the IR is limited to 2 – 5 μm by wavelength and whilst this does not prevent resolution of most plant cells, which range generally between 10 – 50 μm in diameter, cellular components and structures may be substantially smaller than this size and therefore below the limit of resolution. Raman microscopy, which will not be discussed in any depth here, offers a solution to many of these problems. For many years Raman microscopy was considered an expensive and unusual technique despite offering a spatial resolution of better than 1 μm (Schmidt et al., 2010). In addition, Raman microscopy is not affected by water thus making possible the analysis of living samples and samples where cellular structure has been preserved cryogenically. The relatively low signal strength of Raman emission were problematic and often led to poor signal and lengthy analysis times. This matter has been largely overcome by improvements in instrument design and new sensitive Raman techniques e.g. surface enhanced Raman (Knauer et al., 2010) may result in Raman soon becoming the best way to

study and understand the structural organisation of the cell wall at the cellular level. The resolution and flexibility of Raman imaging has been demonstrated recently by several groups studying the ultra-structure and composition of cell wall in tree species and corn stover (Agarwal, 2006; Gierlinger and Schwanninger, 2006; Sun et al., 2010).
In this review we discuss the principles underlying IR spectroscopy, the developments in instrumentation and sample presentation methods that have led to Fourier transform instruments becoming much more compatible with high through-put analysis, and we present data showing how FTIR is being used to assist breeding improvement work in Aberystwyth.

2. The interaction of matter and IR energy

To understand IR spectroscopy it is necessary to have some insight into how energy interacts with the molecules present in living cells. The bonds in molecules are not rigid nor of fixed length and both, bond length and their angle to each other vary about a mean position and vibrate with specific frequencies. Covalent molecules absorb energy in the IR when the frequency of the energy correlates exactly with the vibrational frequency of a chemical bond within the molecule. This absorbance of energy by the molecule causes an increase in the amplitude, but not the frequency of the bond's vibration about its mean centre. By definition, absorption of IR energy can only occur when it causes a change to the bond's electric dipole moment, i.e. the bond's polarity, or the separation of positive and negative electrical charges comprising the covalent bond. Simple homonuclear molecules e.g. N_2 and O_2 cannot absorb in the IR because there is no dipole moment to change. In contrast, it is common for many bonds in covalent heteronuclear molecules to exhibit some degree of polarity e.g. H_2O and CO_2 and most organic compounds absorb strongly in the IR with characteristic absorption frequencies.
The visible, IR and ultraviolet parts of the electromagnetic (EM) spectrum are defined by frequency (υ; the number of wave-cycles passing through a point in one second) measured in Hertz (Hz) and wave length (λ; the length of one complete wave-cycle). The mid-IR region stretches between 3×10^{-4} and 3×10^{-3} cm and is of greatest practical use to the organic chemist because this range corresponds to the fundamental frequencies of molecular vibrations. Frequency and wavelength are related to each other by equation 1, where c is the velocity of light in a vacuum, and has the value 2.997×10^8 m s^{-1}:

$$c = \lambda \upsilon \tag{1}$$

Experiments in the early part of the 20th Century demonstrated that when interacting with matter EM radiation is best described as discrete packets of energy called quanta. The energy (E) of a quantum is directly proportional to its frequency and is calculated by the Bohr equation (equation 2) where h is the Planck constant ($h = 6.626 \times 10^{-34}$ Joules-second). Substitution of equation 1 into equation 2 shows that energy decreases with increasing wave length.

$$E = h\upsilon \tag{2}$$

In IR spectroscopy wavelength is often expressed as its reciprocal, wave number (cm^{-1}), which equates to the number of waves per centimetre. In the mid-IR wave numbers span from 4000 – 400 cm^{-1}; this is a more convenient notation and has the advantage of being proportional with energy.

The infrared spectrum of a sample of a pure compound shows the position of absorptions within the IR spectrum and therefore indicates the functional groups in the molecule. The spectrum can be presented in terms of transmission, where the y axis ranges from 0 – 100% transmission over a range of wave numbers, or in terms of absorption, in which percentage transmission (%T) is converted to absorption (A) by the simple relationship shown in equation 4.

$$A = \log_{10} (100/\%T) \tag{3}$$

For qualitative analysis absorption has the advantage of being proportional to the concentration of the analyte, which can be calculated by equation 4, the Beer-Lambert Law in which c is the molar concentration of analyte, ε the molar extinction coefficient (L mol-1 cm-1) and *l t*he light path through the sample (cm).

$$A = \varepsilon cl \tag{4}$$

Even simple compounds have relatively complex IR spectra showing many absorption bands. This is because there may be many bonds within an individual molecule that are able to absorb in the mid-IR. Bonds within a molecule can vibrate in two ways corresponding to the movements of atoms sharing a bond. Atoms can move relative to each other causing the bond to vary in length, this causes bond stretching, or one atom can move out of plane relative to the other, causing bond bending. For reasons not explained here the maximum number of absorptions in the mid-IR that might be expected from a pure compound comprised of *N* atoms can be calculated by equation 5 for linear molecules such as CO_2 and by equation 6 for non-linear molecules such as H_2O.

$$\text{Linear} = 3N - 5 \tag{5}$$

$$\text{Non-linear} = 3N - 6 \tag{6}$$

Not all possible vibrational modes are active in the IR, for example CO_2 would be expected to have 3 *N* – 5 absorption bands in its spectrum, in reality it has two as one of the possible stretching modes is symmetrical and not IR active, furthermore the two bending modes are degenerate and show as one combined band. Other reasons why fewer than the theoretical number of IR bands are seen in the spectrum include the absorption not being in the 4000–400 cm^{-1} range; an absorption being too weak to be observed and absorptions being too close to each other to be resolved on the instrument. The vibration frequency of the absorption in wave numbers ($\vec{v}$) is proportional with the force constant or bond stiffness (*k*, dyne cm^{-1}) and the masses of the atoms (in grams) sharing the bond (equation 7).

$$\overline{v} = \frac{1}{2\pi c}\sqrt{\frac{k(m1+m2)}{(m1m2)}} \tag{7}$$

3. The Fourier transform infrared spectrometer

Modern mid-IR spectrophotometers are almost exclusively Fourier transform instruments as this design offers greater sensitivity and considerably faster scan speeds compared with the older and now largely obsolete dispersive instruments. The main difference between the

two designs is the manner in which the IR beam is produced which is passed through the sample. Dispersive instruments are based on a monochromator; this device splits a polychromatic beam produced by the source into a beam of very narrow wave number. As the instrument scans over a wave number range the monochromator shifts the wave number of the beam through the scan range. This of course means that the time required for each scan is dependant on how quickly the monochromator can be made to work and at any one point in the scan the beam transmitted through the sample is only a small portion of the sources output. In contrast, a beam covering the entire wave number range of the source is passed through the sample in a Fourier transform instrument. Therefore scan time and beam intensity are considerably faster and brighter respectively. The heart of a Fourier transform instrument is an interferometer. Most commonly this is a Michelson interferometer (Figure 1). In this device IR light from a Nernst or Globar source is split at a beam splitter into two beams of equal intensity, one beam is reflected onto a stationary mirror whist the other is reflected onto a moving mirror. The two beams are recombined to form the transmitted beam at the beam splitter, and this beam is at 90˚to the input beam. The moving mirror produces a varying optical path difference between the two beams resulting in constructive and destructive interference when they are combined. A helium neon laser beam is included in the FTIR spectrophotometer to provide a reference beam of known wave number that can be used to measure precisely the displacement of the moving mirror. This laser is not shown in Figure 1 for the purpose of simplicity. Two types of detector are commonly used in the mid-IR: The deuterium triglycine sulphate (DTGS) detector works at room temperature and has the advantage of great stability and ease of use. For more demanding work increased sensitivity of the mercury cadmium telluride (MCT) may be necessary but this type of detector requires cooling to near liquid nitrogen temperatures in order to work. The resulting interferogram (Figure 2) contains the source's frequency information modulated in a time domain as a function of the moving mirror's displacement.

The entire spectrum is therefore measured simultaneously in the interferogram and FTIR instruments offer considerable speed benefits compared to dispersive instruments which mechanically scan from one wave number to another. An interferogram can be obtained in only a few seconds compared to many minutes so allowing many interferograms to be collected in a comparatively short time. These are averaged allowing great improved signal to noise ratios and increased sensitivity. The absorbance spectrum of the sample is produced from the interferogram by Fourier-transformation; the process requires that the analyst also records an interferogram where no sample is present in the beam path. The ratio of the resulting absorbance spectra corresponds to the absorbance spectrum of the sample alone. Key developments in technology were necessary for Fourier transform based instruments to become practical. Firstly, the development of Fast Fourier transformation (FFT) by Cooley and Tukey (Cooley and Tukey, 1965) offered an algorithm for Fourier transformation that was several orders of magnitude simpler that previous calculation routes whilst only slightly less accurate. Without FFT the lengthy computation would have required unfeasibly powerful and expensive computers. Even the much shorter computation underlying FFT however required the technology to wait for development of affordable desk-top computers to supply the necessary computing power before FTIR based instruments could become practical. In addition, not only does the computer handle the FFT allowing generation of spectra it also allows the analyst the ability to manipulate spectra for integration, base-line correction and averaging, and to analyse the spectra using multivariate approaches such as principal components analysis (PCA) and multivariate regression.

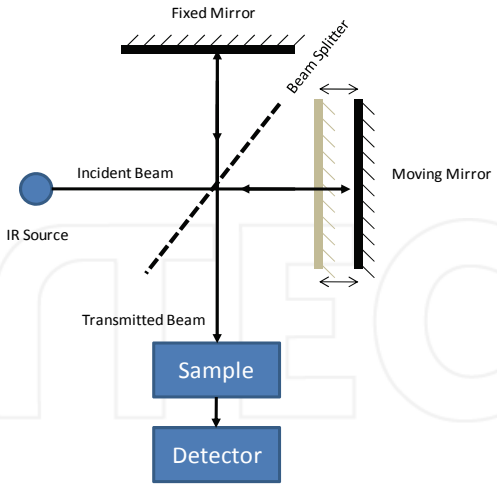


Fig. 1. Schematic of a Michelson interferometer

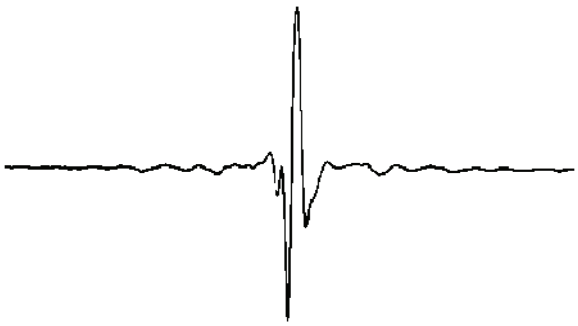


Fig. 2. Typical interferogram (courtesy of Analytical Spectroscopy by R. P. W. Scott, Essential Infromation for the Analytical chemist http://www.analyticalspectroscopy.net/)

4. The development of effective sample analysis techniques

Samples of plant material are generally prepared by oven or freeze-drying to eliminate water which has high absorption in the IR; the dry samples are then finely ground. Generally, the finer that samples can be ground the better but often the analyst finds that for fibrous samples there is trade off between particle size and sample through-put and it is often necessary to determine empirically the maximum size of particle that permits robust analysis. Furthermore, the milled sample must be truly representative of the whole sample and chemical differences between particles of different sizes must be accounted for. It is a good idea to have access to a number of mills differing in design as each offers a combination of through-put and effectiveness that is highly sample specific. The preparation of samples by grinding is often the rate limiting step in IR analysis.

Traditionally, analysis by IR spectroscopy was highly time-consuming. In this approach small amounts of ground sample are ground with a larger quantity of an IR transparent salt such as KBr in a pestle and mortar. The mixture is then pressed hydraulically into transparent discs with between 5 – 10 tonnes of pressure (Kacurakova et al., 2000; Xu et al., 2007). At all stages it is essential that the amount of water in the disk is kept to a minimum and spectra must be taken quickly as the discs are hydroscopic and soon become opaque due to the absorption of water from the atmosphere. This process is time-consuming but allows spectra of a high quality to be obtained. It's use has been reported by several groups engaged on the structural study of various cell wall components including pectin and hemicellulose; (Kacurakova et al., 2000; Sun and Tomkinson, 2002; Xu et al., 2006; Xu et al., 2007), cellulose (Liu et al., 2006), lignin (Gosselink et al., 2004; Sun et al., 2002; Sun et al., 2002; Xu et al., 2007), pyrolysis char (Hu et al., 2008) and lignin (Faix, 1991; Jung and Himmelsbach, 1989; Monties, 1989; Nakanishi and Kawakami, 1991; Stewart et al., 1997). Alternatively, samples can be ground with an oily mulling agent (typically Nujol) and smeared onto salt disks prior to spectral acquisition.

In recent years a much more rapid method of analysing powdered samples has become widely available and this has made the use of FTIR for larger sample numbers more practicable (Allison et al., 2009; Gosselink et al., 2004). The method relies on crushing a powdered sample against a flattened face of a trapezoid crystal of diamond, germanium or zinc selenide mounted in a rugged plate. The accessory is most often placed in the spectrophotometers sample bay and the IR beam is passed through the crystal by a series of mirrors. Because the IR beam hits the walls of the crystal at a very shallow angle it is reflected within the crystal and exits the far side where it passes to the detector. The technique is known as attenuated total reflectance (ATR). Most ATR accessories are of the single bounce design but for samples of only low absorption it is possible to purchase multi-bounce ATR accessories. Spectra are obtained from the sample only when they are placed in direct contact with the crystal and it is important that the particle size of the sample is sufficiently small and uniform to allow spectral acquisition to be reproducible. Because the IR beam only penetrates a few microns from the crystal surface it is possible to directly analyse aqueous samples. All methods require for spectra to be collected in a two step process; first, a blank spectrum is collected with no sample in place, next a sample spectrum is collected and the spectrum of the sample is obtained by calculating the ratio of the two measurements. Using ATR an operator using a manual spectrophotometer can comfortably collect spectra from approximately 200 samples comfortably in one day.

5. Spectral interpretation and multivariate analysis

The handling and interpretation of the IR spectral data is as much a part of compositional analysis as sample preparation and spectral acquisition and is discussed in this chapter in order for the reader to have a clear insight into the entire process. From the information provided in Section 2 it would appear relatively straight forward for the analyst to assign absorption peaks in the spectrum of a pure sample to particular chemical moieties or molecular structures. In practice, IR spectra are more complex than might be expected. Often additional absorption bands are present in the spectrum due to overtones or combinations of the fundamental vibrations. In addition, the spectra of larger molecules have an additional level complexity caused by the coupling of vibrations over part or the entire molecule. When two oscillating bonds share a common atom, the vibrations of the two bonds are coupled and as one bond contracts, the other bond contracts or expands, as in

asymmetrical and symmetrical stretching. When coupling occurs, the spectral position of bands due to particular bonds become shifted rather than superimposed as might be expected. These shifted vibrations are known as skeletal vibrations and give information on the entire molecule rather than on specific functional groups within the molecule. However, despite these complexities spectral interpretation can generally be achieved without resorting to first principles by remembering the principles underlying equation 7, and by reference to either a text book on the subject e.g. Stuart (2005) or to one of the collections of IR spectral fingerprints e.g. Movasaghi et al. (2008).

Spectra from samples of complex mixtures such as plant cell wall preparations or ground samples of dried lignocellulosic crops, present an additional level of difficulty to interpret as it is next to impossible to deconvolute the superimposed spectra of the individual components in the sample. Prior knowledge of the likely composition of the sample often allows the identification of spectral features known to be associated with likely components e.g. in plant material the presence of bands at 1590 cm^{-1} and 1610 cm^{-1} are correlated with lignin (Monties, 1989), and if no knowledge of sample composition exists IR spectra provides clues as to the kind of compounds which may be present. The analyst can thus make an informed decision on the best methods to use for initial chemical analysis of the sample.

Whilst useful for qualitative analysis, these approaches are not sufficiently robust to allow the prediction of composition in spectra from large numbers of samples as they are too easily influenced by co-variance between chemical components in the samples and by random differences in the spectral data. The goal of using IR spectra to predict chemical composition in complex samples has been made possible by the adoption of multivariate approaches to data analysis and the availability of powerful and affordable desk top computers to handle the data processing and file storage. The simplest and possibly the most commonly used multivariate approach is PCA. This was first described in theoretical terms by Karl Pearson in 1901 (Pearson, 1901) and later made into a practical reality by Hotelling in 1933 (Hotelling, 1933). PCA allows the variance in the spectra data at hundreds of wave numbers to be condensed into a much smaller number of new principal components which explain most of the variance in the original data. Because there are fewer variables the data can be more easily explored graphically or statistically for correlation with experimental, environmental and sample effects. PCA works best with highly correlated data and spectral data is highly correlated. Any single data point in a spectrum is highly influenced by the value of its immediate neighbours. PCA is of great value for the biologist and analytical chemist hoping to understand how differences between samples in geographic location, growth year, species or some experimental treatment such as the addition of fertiliser, have an effect on chemical composition. This is illustrated by Figure 3A which shows a plot of the first and second principal components obtained from 194 mid-IR spectra of reed canary grass and switchgrass samples collected at the end of two consecutive growth years. The two components plotted in this figure represent the majority of the variance in the data set. The data points are coded according to species and it is quite apparent that spectral differences exist between the two grass species and these differences are accounted for primarily by the variance explained by principal component 2. The separation of the data points is however only partial showing that chemical differences between samples from the two species occur in only a proportion of cases. The analyst would usually wish to explore this further and establish whether this was the result of processing and sample preparation e.g. differences in sample moisture content, or whether

these differences reflected or factors influencing the growth and development of the plants during their growth in the field. Re-plotting these data but coding the data points by growth year would result in an identical distribution of data points but show no separation of the data according to growth year by these components (data not shown).

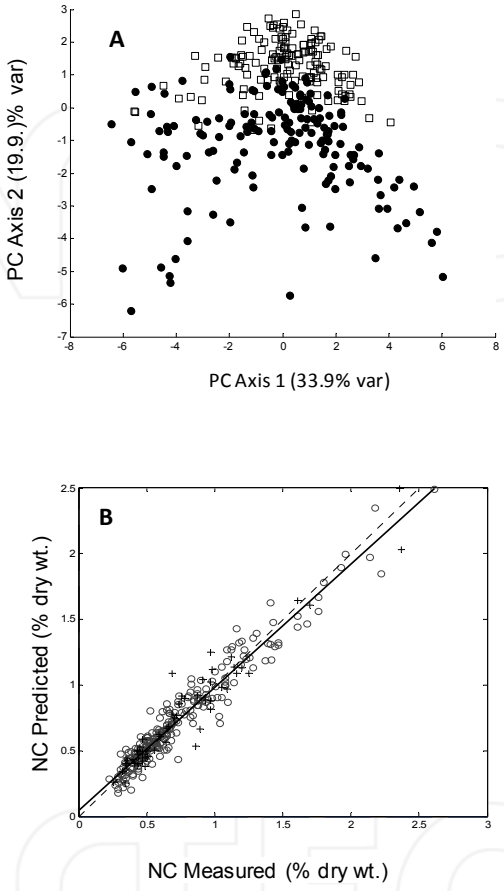


Fig. 3. (A) PCA of 194 mid-infrared attenuated total reflectance absorbance spectrums of samples of reed canary over the spectral range of 500- 4000 cm^{-1} (□ = switchgrass, ● = reed canary grass). (B) Scatter plots showing measured vs. PLS model predicted values for nitrogen content in the training (●) and test (+) spectral data. (Data from Allison et al., 2009)

PCA is obviously a highly useful tool for evaluating differences and similarities between samples. In addition, the analyst may identify the wave numbers where these differences occur. This information is contained in the PCA loadings, a matrix of data produced during the PCA process that shows the influence of each of the original variables i.e. absorbance's at groups of wave numbers, on each of the new components. Analysis of the loadings for principal component 2 would in this case identify the discriminatory wave numbers between the two species and perhaps make it possible to identify the class of chemical that

differed most between the two species. This very powerful way to look at chemical data will also reveal not just increases or decreases in given compounds but changes in the concentrations of several chemical components simultaneously.
PCA and other methods of this kind e.g. canonical correlation, cluster analysis and multidimensional scaling are therefore excellent tools for exploring structure relationships within the spectral data but they cannot be used to determine the concentration of components within the sample mixture. For that it is necessary to employ a multivariate regression approach. Many approaches have been developed over the years e.g. see Otto (2007), and currently many analytical chemists rely on regression by partial least squares (PLS). Other methods such as principal components regression and multivariate linear regression offer subtle differences in capability but the way in which they are employed is similar. Multivariate regression requires the analyst to provide a training set of data from samples that have been also been analysed using standard chemical methods. The variance in these training data is arranged by PLS into new orthogonal components (latent variables) but unlike PCA these new latent variables not only capture variance but also achieve correlation with the analytical data. To ensure that the correlation is genuine and based on the chemical components being measured rather than to some selective fitting of noise the process is controlled by cross validation. The mathematics underlying these calculations is complex and many analytical scientists rely on software from specialist suppliers or instrument manufacturers, although it is also possible to find free-ware packages on the internet. The PLS regression is judged for predictive accuracy using an independent test set of spectral data from chemically analysed samples that were excluded from the PLS regression fitting process. This gives the user confidence that predictions made with the model will be accurate. Confidence in predictions made using the model is obtained by chemically analysing a small percentage of all samples and checking the agreement between te real and predicted values, and by monitoring the variance presented by new samples to ensure that they are adequately explained by the existing model. Figure 3B shows a plot of predicted vs. measured nitrogen content for the 194 samples of reed canary grass and switchgrass. Theoretical perfect fit is displayed as a dotted line in the figure.

6. Example of process

In this example using previously unpublished data we show how FTIR spectra can be used to develop predictive PLS models for several combustion parameters in samples of coal. This may seem to be a curious choice for illustration but the procedure by which the data are analysed and models developed are essentially identical to those employed for developing models to predict aspects of cell wall composition in less ancient biomass. The purpose of this study was to explore the possibility of using FTIR for the process monitoring of coal and biomass destined for co-firing. Two similar studies were published in 2009 which used NIRS (Kim et al., 2009) and FTIR (Geng et al., 2009) and these provide a useful comparison for these findings. Sixty nine samples of powdered coal were provided by E.ON. It was known that the coal samples were from a number of locations but details of their origins were not disclosed. The coal samples had been subjected to proximate analysis and values were supplied of the mass fractions of moisture, volatiles, ash and sulphur, see Table 1. With only minor exceptions the data were normally distributed. Triplicate IR spectra were taken from each sample from 600 cm^{-1} to 4000 cm^{-1} using a Golden Gate ATR accessory fitted with a single bounce diamond crystal (Specac, U.K.) placed in the sample compartment of an Equinox 55 FTIR spectrophotometer (Bruker Optik GmbH, Germany).

Each spectrum represented an average of 32 scans at a resolution of 4 cm^{-1}. Spectra were averaged, smoothed and derivatised to the first Savitsy- Golay derivative using a window of 25 wave numbers and a second order polynomial fit before being normalised to unit variance. All spectral processing was performed using MatLab (version 7.10; Mathworks) and the PLS toolbox (version 6.0; Eigenvector Research, Inc.).

	Xa	Xm	Xs	Xv
Mean	10.81	11.93	1.124	31.01
Minimum	2.30	4.00	0.120	24.10
Maximum	41.10	25.90	2.390	35.60
SD	4.70	3.72	0.648	2.56

Table 1. Values, range and standard deviation (SD) of Q, calorific value (corrected for ash and moisture content); Xa, ash mass fraction; Xcl, chlorine mass fraction; Xm, moisture mass fraction; Xs, sulphur mass fraction and Xv, volatile mass fraction for 69 samples of powdered coal. Data supplied by E.ON

Figure 4A shows the complete set of 207 spectra taken from the 69 samples before averaging, derivatisation and normalisation. Derivatisation is a commonly employed practice in spectroscopy as it improves the separation of non-resolved peaks. In NIRS it is not unusual to derivatise spectra to the 3rd or 4th derivative. In FTIR however, the peaks are better resolved and derivatisation beyond the 1st or 2nd derivative typically offers no advantage. The spectra shown in Figure 4B are much more tightly grouped and it is apparent that this regime of pre-processing has improved the noise in the data considerably. Other processes which might have been applied include scatter correction, baseline correction offsets and general least squares regression which serves to decrease noise between chemically similar data points. The 69 spectra were divided into two groups; one group of 59 spectra that were used to develop the partial least squares regression model and an independent test set of 10 spectra which were excluded from the model but which were used to assess model predictive accuracy. The PLS modes were developed using SIMPLS algorithms (de Jong, 1993) and the PLS toolbox software. Models were cross validated using venetian blinds cross validation protocol (7 data splits) and developed to an optimal number of latent variables to ensure that the models were based on variance explaining the parameter of interest rather than on noise in the data. In each case the best number of latent variables to include was indicated by a minimal value of the root mean square error of cross validation (RMSECV). This measure of error can be interpreted as the standard deviation of the unexplained variance in the cross validated regression, and has the useful property of being in the same units as the response variable. This measure of error is a much better indication of model fit than root mean square error of calibration (RMSEC) as the latter does not indicate when the model is over-fitted to the data. The models were tested for predictive accuracy by measuring the root mean square error of prediction (RMSEP) using the independent data test set. For RMSECV, n relates to the number of spectra in the training data set, y_i is an observed value obtained by chemical analysis and $\hat{y}_i$ a value predicted from the cross validated regression model. RMSEC can be calculated using the same equation by substituting the values of $\hat{y}_i$ with the predicted values from the non-cross validated. Similarly, RMSEP is calculated by using values from the independent data test set.

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (8)$$

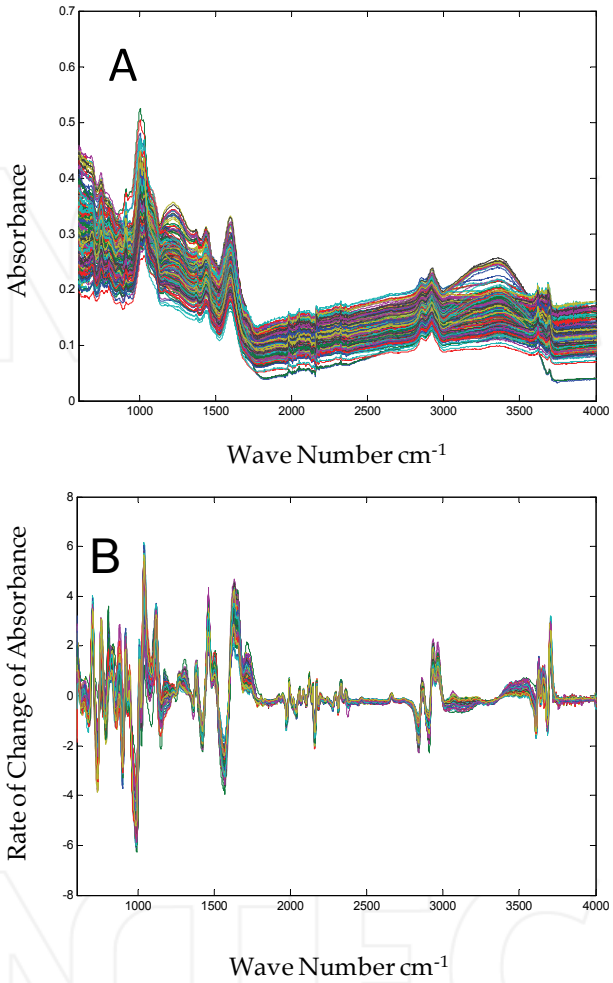


Fig. 4. (A) Triplicate ATR-FTIR spectra (207 spectra) of 69 the samples of powdered coal. (B) First derivative averaged spectra after Savitsky-Golay smoothing and normalisation (69 spectra)

Plots of regression fits for these four models are shown in Figure 5 and the parameters of the four models in Table 2. All of the models are sufficiently good to be used for prediction of these parameters in unknown samples. Ideally, the variance in spectra from the unknown samples would be compared with those used to develop the model in question. This ensures that the model is capable of making an accurate prediction. Some of the samples detected as dissimilar would be analysed chemically and these data incorporated into a new model better able to make predictions. In all circumstances a proportion of the unknown samples, perhaps 5 – 10% of the total number should be analysed chemically to validate model predictions.

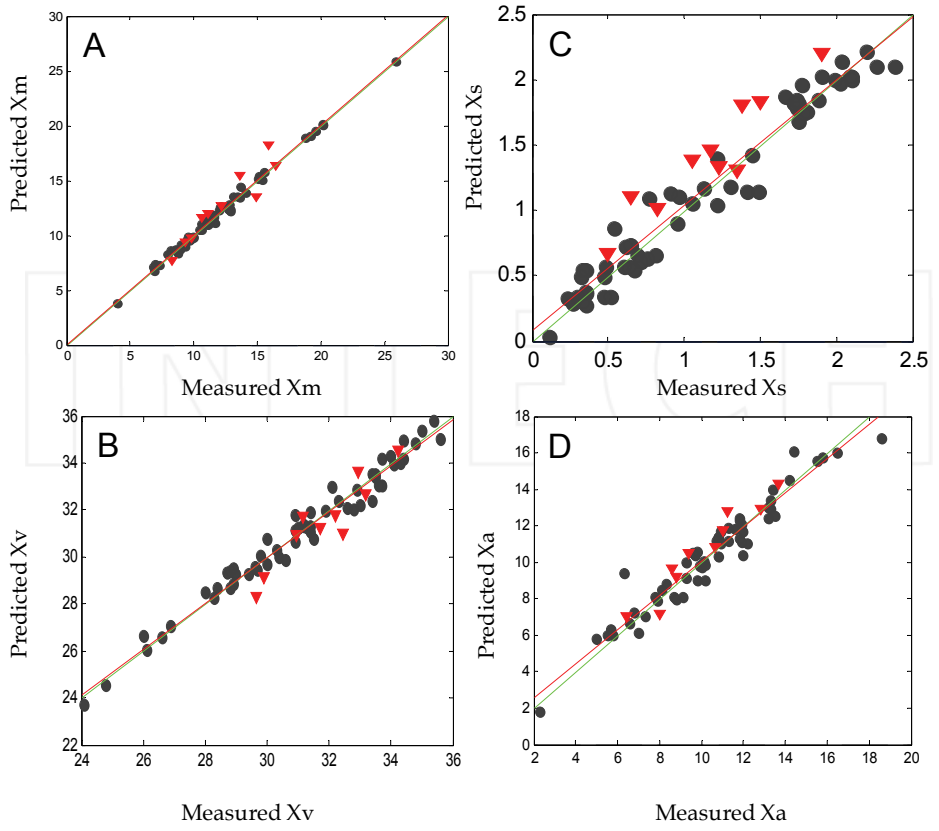


Fig. 5. PLS regression fits (red line) for the spectral data from the coal samples for A, moisture; B, volatiles; C, sulphur and D, ash mass fractions. Black and red symbols indicate samples in the training and test data sets respectively. The green line denotes a 1:1 regression fit

Model	Calibration							Validation		
	LV	% Var x	% Var y	R^2	RMSEC	R^2_{cv}	RMSECV	PredBias	R^2_p	RMSEP
Xm	9	78.7	99.5	0.995	0.228	0.781	1.802	-0.08	0.888	1.181
Xv	7	76.3	97.4	0.974	0.435	0.714	1.490	-0.35	0.869	0.782
Xs	5	62.3	95.9	0.935	0.137	0.827	0.286	0.05	0.887	0.290
Xa	5	69.5	93.2	0.932	0.785	0.762	1.504	0.02	0.928	0.831

Table 2. Summary of the PLS models developed to measure moisture (Xm), volatiles (Xv), sulphur (Xs) and ash (Xa) mass fractions. Details are shown of the number of latent variables in each model (LV), the percentage variation included from the spectral (% Var x) and non-spectral data (% Var y), predictive bias (Pred Bias) and model goodness of fit is shown for correlation, cross validation and prediction as root mean square error (RMSEC, RMSECV and RMSEP) and as determination coefficients (R^2, R^2_{cv} and R^2_p)

7. Conclusions

Mid-IR spectrophotometry has become a potentially useful predictive tool for the plant breeder, biologist and engineer as it offers high rates of sample through-put, low unit cost and robust and accurate analysis. Key to this approach has been the development of modern Fourier transform spectrophotometers, the advent of modern personal computers and the availability of chemometrics software. Whilst many still prefer to use near-IR spectral analysis for the prediction of compositional parameters mid-IR gives the analyst additional information on the molecules within the sample which may be more difficult to discern using NIRS.

8. Acknowledgements

This work was funded by the Engineering and Physical Sciences Research council (EPSRC) and carried out as part of the Supergen Consortium in Biomass, Biofuels and Energy Crops (GR/S28204). The author wishes to thank Catherine Morris for making FTIR spectral analysis of the coal samples and Hugh Burnham-Slipper of E.ON for supplying the samples.

9. References

Agarwal, U.P. (2006). Raman imaging to investigate ultrastructure and composition of plant cell walls: distribution of lignin and cellulose in black spruce wood (*Picea mariana*). *Planta*, 224, 1141-1153.

Allison, G.G., Morris, C., Hodgson, E., Jones, J., Kubacki, M., Barraclough, T., Yates, N., Shield, I., Bridgwater, A.V., Donnison, I.S. (2009). Measurement of key compositional parameters in two species of energy grass by Fourier transform infrared spectroscopy. *Bioresource Technology*, 100, 2428-2433.

Allison, G.G., Robbins, M.P., Carli, J., Clifton-Brown, J., Donnison, I. (2010). Designing biomass crops with improved calorific content and attributes for burning: a UK perspective. in: *Plant Biotechnology for Sustainable Production of Energy and Co-Products*, P. Mascia, J. Schefrran, J.M. Widholm (Eds.). Springer, pp. 25-56, IBSN 978-3-642-13439-5.

Allison, G.G., Thain, S.C., Morris, P., Morris, C., Hawkins, S., Hauck, B., Barraclough, T., Yates, N., Shield, I., Bridgwater, A.V., Donnison, I.S. (2009). Quantification of hydroxycinnamic acids and lignin in perennial forage and energy grasses by Fourier-transform infrared spectroscopy and partial least squares regression. *Bioresource Technology*, 100, 1252–1261.

Bridgwater, A.V. (2003). Renewable fuels and chemicals by thermal processing of biomass. *Chemical Engineering Journal*, 91, 87-102.

Carroll, A., Somerville, C. (2009). Cellulosic biofuels. *Annual Review of Plant Biology*, 60, 165-182.

Chang, M.C.Y. (2007). Harnessing energy from plant biomass. *Current Opinion in Chemical Biology*, 11, 677-684.

Chen, L., Carpita, N.C., Reiter, W.-D., Wilson, R.H., Jeffries, C., McCann, M.C. (1998). A rapid method to screen for cell-wall mutants using discriminant analysis of Fourier transform infrared spectra *The Plant Journal*, 16, 385-392.

Clifton-Brown, J., Robson, P., Allison, G., Lister, S., Sanderson, R., Hodgson, E., Farrar, K., Hawkins, S., Jensen, E., Jones, S., Huang, L., Roberts, P., Youell, S., Jones, B., Wright, A., Valantine, J., Donnison, I. (2008). *Miscanthus*: breeding our way to a better future. in: E. Booth, M. Green, A. Karp, I. Shield, D. Stock, D. Turley (Eds.), Biomass and Energy Crops III, pp. 199-206

Cooley, J.W., Tukey, O.W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19, 297-301.

de Jong, S. (1993). SIMPLS: an alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18, 251-263.

Fahmi, R., Bridgwater, A.V., Darvell, L.I., Jones, J.M., Yates, N., Thain, S., Donnison, I.S. (2007). The effect of alkali metals on combustion and pyrolysis of Lolium and Festuca grasses, switchgrass and willow. *Fuel*, 86, 1560-1569.

Fahmi, R., Bridgwater, A.V., Donnison, I., Yates, N., Jones, J.M. (2008). The effect of lignin and inorganic species in biomass on pyrolysis oil yields, quality and stability. *Fuel*, 87, 1230-1240.

Faix, O. (1991). Classification of lignins from different botanical origins by FTIR spectroscopy. *Holzforschung*, 45, 21-27.

Foster, C.E., Martin, T.M., Pauly, M. (2010). Comprehensive Compositional Analysis of Plant Cell Walls (Lignocellulosic biomass) Part I: Lignin. *Journal of Visualised Experiments*, 37, http://www.jove.com/index/Details.stp?ID=1745.

Foster, C.E., Martin, T.M., Pauly, M. (2010). Comprehensive Compositional Analysis of Plant Cell Walls (Lignocellulosic biomass) Part II: Carbohydrates. *Journal of Visualised Experiments*, 37, http://www.jove.com/index/Details.stp?ID=1837.

Fukushima, R.S., Dehority, B.A. (2000). Feasibility of using lignin isolated from forages by solubilization in acetyl bromide as a standard for lignin analyses. *Journal of Animal Science*, 78, 3135-3143.

Fukushima, R.S., Hatfield, R.D. (2001). Extraction and Isolation of Lignin for Utilization as a Standard to Determine Lignin Concentration Using the Acetyl Bromide Spectrophotometric Method. *Journal of Agricultural and Food Chemistry*, 49, 3133-3139.

Geng, W., Nakajima, T., Takanashi, H., Ohki, A. (2009). Analysis of carboxyl group in coal and coal aromaticity by Fourier transform infrared (FTIR) spectrometry. *Fuel*, 88, 139-144.

Gierlinger, N., Schwanninger, M. (2006). Chemical imaging of poplar wood cell walls by confocal Raman microscopy. *Plant Physiology*, 140, 1246-1254.

Giger-Reverdin, S. (1995). Review of the main methods of cell wall estimation: interest and limits for ruminants. *Animal Feed Science and Technology*, 55, 295-334.

Gosselink, R.J.A., Abacherli, A., Semke, H., Malherbe, R., Kauper, P., Nadif, A., van Dam, J.E.G. (2004). Analytical protocols for characterisation of sulphur-free lignin. *Industrial Crops and Products*, 19, 271-281.

Grabber, J.H. (2005). How do lignin composition, structure, and cross-linking affect degradability? A review of cell wall model studies. *Crop Science*, 45, 820-831.

Hatfield, R.D., Fukushima, R.S. (2005). Can lignin be accurately measured? *Crop Science*, 45, 832-839.

Hatfield, R.D., Ralph, J., Grabber, J.H. (1999). Cell wall structural foundations: molecular basis for improving forage digestibilities. *Crop Science*, 39, 27-37.

Hodgson, E.M., Fahmi, R., Yates, N., Barraclough, T., Shield, I., Allison, G., Bridgwater, A.V., Donnison, I.S. (2010). Miscanthus as a feedstock for fast-pyrolysis: Does agronomic treatment affect quality? *Bioresource Technology*, 101, 6185-6191

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417-441; 498-520.

Hsu, C.-P.S. (1997). Infrared spectroscopy. In: *Handbook of instrumental techniques for analytical chemistry Prentice Hall*, F.A. Settle (Ed.), pp. 247-283, ISBN 0-131-77338-0.

Hu, S., Xiang, J., Sun, L., Xu, M., Qiu, J., Fu, P. (2008). Characterization of char from rapid pyrolysis of rice husk. *Fuel Processing Technology*, 89, 1096-1105.

Huang, A., Zhou, Q., Liu, J., Fei, B., Sun, S. (2008). Distinction of three wood species by Fourier transform infrared spectroscopy and two-dimensional correlation IR spectroscopy. *Journal of Molecular Structure*, 883-884, 160-166.

Johnson, D.B., Moore, W.E., Zank, L.C. (1961). The spectrophotometric determination of lignin in small wood samples. *TAPPI*, 44, 793-798.

Jung, H.-J.G., Himmelsbach, D.S. (1989). Isolation and characterisation of wheat straw lignin. *Journal of Agricultural and Food Chemistry*, 37, 81-87.

Kacurakova, M., Capek, P., Sasinkova, V., Wellner, N., Ebringerova, A. (2000). FTIR study of plant cell wall model compounds: pectic polysaccharides and hemicelluloses. *Carbohydrate Polymers*, 43, 195-203.

Kim, D., Lee, J., Kim, J. (2009). Application of near infrared diffuse reflectance spectroscopy for on-line measurement of coal properties. *Korean Journal of Chemical Engineering*, 26, 489-495.

Knauer, M., Ivleva, N.P., Liu, X., Niessner, R., Haisch, C. (2010). Surface-Enhanced Raman Scattering-Based Label-Free Microarray Readout for the Detection of Microorganisms. *Analytical Chemistry*, 82, 2766-2772.

Laird, D.A. (2008). The charcoal vision: A win-win-win scenario for simultaneously producing bioenergy, permanently sequestering carbon, while improving soil and water quality. *Agronomy Journal*, 100, 178-181.

Liu, C.F., Xu, F., Sun, J.X., Ren, J.L., Curling, S., Sun, R.C., Fowler, P., Baird, M.S. (2006). Physicochemical characterization of cellulose from perennial ryegrass leaves (Lolium perenne). *Carbohydrate Research*, 341, 2677-2687.

McCann, M.C., Defernez, M., Urbanowicz, B.R., Tewari, J.C., Langewisch, T., Olek, A., Wells, B., Wilson, R.H., Carpita, N.C. (2007). Neural network analyses of infrared spectra for classifying cell wall architectures. *Plant Physiology*, 143, 1314-1326.

McKendry, P. (2002). Energy production from biomass (part 1): overview of biomass. *Bioresource Technology*, 83, 37-46.

Mohan, D., Pittman, C.U., Steele, P.H. (2006). Pyrolysis of wood/biomass for bio-oil: A critical review. *Energy & Fuels*, 20, 848-889.

Monties, B. (1989). Lignins. *Methods in Plant Biochemistry: Plant Phenolics,* J.B. Harborne (Ed.), 1, pp. 113-157.

Mouille, G., Robin, S., Lecomte, M., Pagant, S., Hofte, H. (2003). Classification and identification of Arabidopsis cell wall mutants using Fourier-Transform InfraRed (FT-IR) microspectroscopy. *The Plant Journal*, 35, 393-404.

Movasaghi, Z., Rehman, S., Rehman, I.u. (2008). Fourier transform infrared (FTIR) spectroscopy of biological tissues. *Applied Spectroscopy Reviews*, 43, 134 - 179.

Nakanishi, R., Kawakami, H. (1991). Determination of lignin in needles by diffuse reflectance Fourier-transform infrared spectrometry. *Mokuzai Gakkaishi*, 37, 841-846.

Otto, M. (2007). Chemometrics. Wiley-VCH, IBSN 3-527-31418-0.

Pearson, K. (1901). On lines and planes of closest fit to a system of points in space. *Philosophical Magazine*, 2, 557-572.

Ptasinski, K.J., Prins, M.J., Pierik, A. (2007). Exergetic evaluation of biomass gasification. *Energy*, 32, 568-574.

Schmidt, M., Schwartzberg, A.M., Carroll, A., Chaibang, A., Adams, P.D., Schuck, P.J. (2010). Raman imaging of cell wall polymers in Arabidopsis thaliana. *Biochemical and Biophysical Research Communications*, 395, 521-523

Sims, R.E.H., Hastings, A., Schlamadinger, B., Taylor, G., Smith, P. (2006). Energy crops: current status and future prospects. *Global Change Biology*, 12, 2054-2076.

Stewart, B. (2005). *Infrared spectroscopy: Fundamentals and application*s John Wiley & Sons, Ltd., ISBN 0-470-85428-6.

Stewart, D., Yahiaoui, N., McDougall, G.J., Myton, K., Marque, C., Boudet, A.M., Haigh, J. (1997). Fourier-transform infrared and Raman spectroscopic evidence for the incorporation of cinnamaldehydes into the lignin of transgenic tobacco (*Nicotiana tabacum* L.) plants with reduced expression of cinnamyl alcohol dehydrogenase. *Planta*, 201, 311-318.

Sun, L., Simmons, B.A., Singh, S. (2010). Understanding tissue specific compositions of bioenergy feedstocks through hyperspectral Raman imaging. *Biotechnology and Bioengineering*.

Sun, R.C., Sun, X.F., Fowler, P., Tomkinson, J. (2002). Structural and physico-chemical characterization of lignins solubilized during alkaline peroxide treatment of barley straw. *European Polymer Journal*, 38, 1399-1407.

Sun, R.C., Sun, X.F., Wang, S.Q., Zhu, W., Wang, X.Y. (2002). Ester and ether linkages between hydroxycinnamic acids and lignins from wheat, rice, rye, and barley straws, maize stems, and fast-growing poplar wood. *Industrial Crops and Products*, 15, 179-188.

Sun, R.C., Tomkinson, J. (2002). Characterization of hemicelluloses obtained by classical and ultrasonically assisted extractions from wheat straw. *Carbohydrate Polymers*, 50, 263-271.

Woolf, D., Amonette, J.E., Street-Perrott, F.A., Lehmann, J., Joseph, S. (2010). Sustainable biochar to mitigate global climate change. *Nat Communications*, 1, 56.

Xu, F., Geng, Z.C., Sun, J.X., Liu, C.F., Ren, J.L., Sun, R.C., Fowler, P., Baird, M.S. (2006). Fractional and structural characterization of hemicelluloses from perennial ryegrass (Lolium perenne) and cocksfoot grass (Dactylis glomerata). *Carbohydrate Research*, 341, 2073-2082.

Xu, F., Sun, J.X., Geng, Z.C., Liu, C.F., Ren, J.L., Sun, R.C., Fowler, P., Baird, M.S. (2007). Comparative study of water-soluble and alkali-soluble hemicelluloses from perennial ryegrass leaves (*Lolium perenne*). *Carbohydrate Polymers*, 67, 56-65.

Xu, F., Zhou, Q.-A., Sun, J.X., Liu, C.-F., Ren, J.L., Sun, R.-C., Curling, S., Fowler, P., Baird, M.S. (2007). Fractionation and characterization of chlorophyll and lignin from de-juiced Italian ryegrass (Lolium multifolrum) and timothy grass (Phleum pratense). *Process Biochemistry*, 42, 913-918.

5

Applications of Fourier Transform Infrared Spectroscopy to Study Cotton Fibers

Noureddine Abidi, Eric Hequet and Luis Cabrales
Texas Tech University
USA

1. Introduction

Fourier Transform Infrared microscopy (FTIR) has become an essential analytical tool available to scientists to study various materials. Specifically FTIR has been increasingly used to study cell wall developments in plants, investigate the efficiency of the surface modification of polymers, identifying contaminants, and predicting the physical properties of certain polymers and biopolymers, etc. This chapter reviews some selected application of the FTIR to study cellulose development in cotton fibers and to predict cotton fiber physical properties. Cotton fiber maturity is a major yield component and an important fiber quality trait that is directly linked to the quantity of cellulose deposited during the secondary cell wall (SCW) biogenesis. Cotton fiber development consists of five major overlapping stages: differentiation, initiation, polar elongation, secondary cell wall development, and maturation. The transition period between 16 and 21 dpa (days post anthesis) is regarded to represent a major developmental stage between the primary cell wall and the SCW. Fourier Transform Infrared spectroscopy was used to investigate the structural changes that occur during the different developmental stages. The IR spectra of fibers harvested at different stages of development (10, 14, 17, 18, 19, 20, 21, 24, 27, 30, 36, 46, and 56 dpa) show the presence of vibrations located at 1,733 cm^{-1} (C=O stretching originating from esters or amides) and 1,534 cm^{-1} (NH_2 deformation corresponding to proteins or amino acids). The results converge towards the conclusion that the transition phase between the primary cell wall and the secondary cell wall occurs between 17 and 18 dpa in fibers from TX19 cultivar, while this transition occurs between 21 and 24 dpa in fibers from TX55 cultivar. The Universal Attenuated Total Reflectance FTIR (UATR-FTIR) was used to evaluate the cotton fiber properties. One hundred and four cotton samples were selected. Thirty FTIR spectra were acquired from each sample and analyzed. Partial Least Square (PLS) analysis of the FTIR spectra was performed and the results showed that micronaire and surface area (calculated from the AFIS data) could be predicted from the FTIR measurements with very high coefficients of determination. However, the prediction of fiber maturity is probably not possible with the UATR-FTIR. It was concluded that, to be able to predict the fiber maturity with the FTIR, it would be necessary to perform the measurements in the transmission mode rather than in the reflectance mode.

2. Fourier transform infrared spectroscopic approach to the study of the secondary cell wall development in cotton fiber

"With kind permission from Springer Science + Business Media: Cellulose, Fourier transform infrared spectroscopic approach to the study of the secondary cell wall development in cotton fiber, 17, 2010, 309-320, Noureddine Abidi, Luis Cabrales, Eric Hequet".

2.1 Introduction

Fiber maturity is a major yield component and an important fiber quality trait that is directly linked to the quantity of cellulose deposited during the Secondary Cell Wall (SCW) biogenesis, and to the organization and orientation of crystalline microfibrils. It is inuitively obvious to hypothesize that immature fibers (having a thin, poorely developed secondary wall) will be fragile, and therefore, are likely to break during the multiple mechanichal stresses involved in transforming fibers into yarns. Immature fibers generate short fibers and neps (entanglement of fibers) that result in yarn defects and decreased productivity in the spinning mills. Therefore, studying cotton fiber maturity and understanding the link between SCW biogenesis and cotton fiber maturity is very important. This study was designed to investigate the structural changes occurring during the growth and development of cotton fibers.

Cotton fiber development consists of five major overlapping developmental stages (Wilkin & Jernstedt, 1999): differentiation, initiation, polar elongation, secondary cell wall deposition, and maturation. The day of flowering is referred to as anthesis and the term "days post-anthesis" (dpa) is often used to describe the cotton fiber development. Fiber initiation, which commences at 0 dpa, signals the onset of fiber morphogenesis. Fiber growth is characterized by the synthesis of the primary cell wall and an increase in fiber length up to ~30 mm within 3 weeks after anthesis. The stage of secondary cell wall development commences in general around 21 dpa and continues for a period of ~3 to 6 weeks post-anthesis. This phase is marked by a massive deposition of a thick cellulosic wall (Wilkin & Jernstedt, 1999). The transition period between 16 and 21 dpa is considered to represent a developmental switch in emphasis from primary to secondary cell synthesis during cotton fiber development. During these developmental stages, important structural changes occur leading to cellulose macromolecules formation (β(1→4) glucopyranose).

Tokumoto et al. (2002) reported on the changes in the sugar composition and molecular mass distribution of matrix polysaccharides during cotton fiber development. The results showed that the extractable matrix (pectic and hemicellulosic) polysaccharides accounted for 30 - 50% of total sugar content during the elongation stage and less than 3% during the cell thickening stage. With respect to the amount of cellulose present during fiber development, it was reported that the secondary wall thickening and maturation stages are characterized by a dramatic increase in the amount of cellulose (Tokumoto et al., 2002). The primary cell wall was reported to contain between 35 and 50% cellulose (Timpa & Triplett 1993; Meinert & Delmer, 1977). However, the secondary cell wall is composed of nearly 100% cellulose (Haigler et al., 2005).

The composition of the cotton fiber cell wall exhibits a continuous change throughout the development of the fiber (Meinert & Delmer, 1977). Several studies have been focused on

the analysis of the composition of the cell wall extracts (Tokumoto et al., 2002; Timpa & Triplett 1993; Meinert & Delmer 1977; Maltby et al., 1979; Huwyler et al., 1979, Gokani et al., 1998). However, limited research has been conducted on intact fibers. In previous research, we reported on the usefulness of Fourier Transform Infrared (FTIR) and Thermogravimetric Analysis (TGA) to investigate cotton fiber development (Abidi et al., 2008). The results showed that these two analytical techniques could be used to evaluate the cell wall composition and structure during fiber development. In addition, because of the nondestructive character of the FTIR analysis other testing could be performed on the same set of samples.

Fourier Transform Infrared spectroscopy (FTIR) has emerged as a key technique for the study of plant growth and development (McCann et al., 2007; Yong et al., 2005; Carpita et al., 2001; Zeier & Schreiber, 1999; Chen et al., 1998; McCann et al., 1997; Séné et al., 1994; McCann et al., 1993; McCann et al., 1992). Zeier and Schreiber (1999) used FTIR to characterize isolated endodermal cell walls from plant roots and assigned FTIR frequencies to functional groups present in the cell wall, including the relative amounts of the cell wall biopolymers suberin and lignin, as well as cell wall carbohydrates and proteins. FTIR absorption spectra indicated structural differences for three developmental stages of the endodermal cell wall under study. The authors concluded that FTIR could be used as a direct and non-destructive method suitable for the rapid investigation of isolated plant cell walls. The approach has since been successfully applied to screen large numbers of mutants for a broad range of cell wall phenotypes using FTIR of leaves of *Arabidopsis thaliana* and flax (*Linum usitatissimum*) (Chen et al., 1998). In this study, Chen and co-workers reported that principal component analysis (PCA) of FTIR spectra can distinguish between mutants that are deficient in cell wall sugars. Also, FTIR and Fourier-Transform Raman spectroscopy have been successfully used to investigate the primary cell wall architecture at a molecular level (Séné et al. 1994). Dynamic changes in cell wall composition of hybrid maize coleoptiles (*Zea mays*) were investigated by FTIR (McCann et al., 2007). The authors reported that neural network algorithms could correctly classify infrared spectra from cell walls harvested from individuals differing at one-half-day interval of growth.

The aim of the work reported in this paper has been 2-fold: 1) to investigate the structure and composition of fiber during different phases of development; and 2) to elucidate the effect of cultivar on the developmental stages of the cotton fiber.

2.2 Experimental

2.2.1 Materials

For this study, two independent replications (10 plants each) of two cotton cultivars (*Gossypium hirsutum* L. cv. TX19 and TX55) were planted in a greenhouse with day/night cycles varying from 13/11 to 11/13 hours and day/night temperatures of about 31°C / 24°C. Plants were grown in 20-L (5 gallons) pots of Sungrow SB 300 potting mix that had been amended with Peters 15-9-12 slow release fertilizer prior to potting. Plants were watered as needed. On the day of flowering (0 dpa), individual flowers were tagged, and 14 developing bolls per cultivar and per replication were harvested at 10, 14, 17, 18, 19, 20, 21, 24, 27, 30, 36, 46, and 56 dpa. The pericarp was immediately removed (excised with scalpel) and isolated ovules were transferred into cryogenic vials and stored in a Cryobiological Storage System filled with liquid nitrogen for analyses. Each replication was tested independently.

2.2.2 Methods

2.2.2.1 Sample dehydration

Frozen cotton fiber samples were dehydrated using the procedure described in (Abidi et al., 2008; Muller & Jackset al., 1975; Rajasekaran et al., 2006). Frozen samples were first rinsed with water and then washed with acidified solution of 2,2-dimethoxypropane (one drop of HCl in 50 ml of 2,2-dimethoxypropane), followed by five exchanges for 15 minutes each in 100% acetone. In a slightly acidic solution, 2,2-dimethoxypropane is instantly hydrolyzed by water to form methanol and acetone (Muller & Jacks, 1975).

2.2.2.2 FTIR measurements

The FTIR spectra of cotton fiber samples were recorded in an environmentally-controlled laboratory maintained at relative humidity of 65±2% and 21±1°C using the Spectrum-One equipped with an UATR (Universal Attenuated Total Reflectance) accessory (Perkin-Elmer, USA). The UATR-FTIR was equipped with a ZnSe-Diamond crystal composite that allows collection of FTIR spectra directly on a sample without any special preparation. The instrument is equipped with a "pressure arm" which is used to apply a constant pressure to the cotton samples positioned on top of the ZnSe-Diamond crystal to ensure a good contact between the sample and the incident IR beam and prevent the loss of the IR beam. The amount of pressure applied is monitored by the Perkin-Elmer FTIR software.

Thirty FTIR spectra per sample were acquired for each developmental stage to produce a total of 180 spectra (30 spectra x 3 replications per dpa x 2 greenhouse replications). All FTIR spectra were collected at a spectrum resolution of 4 cm^{-1}, with 32 co-added scans over the range from 4,000 to 650 cm^{-1}. A background scan of clean ZnSe-Diamond crystal was acquired before scanning the samples.

2.2.2.3 FTIR spectra analysis

The Perkin-Elmer software was used to perform spectra normalization, baseline corrections, and peak integration. FTIR spectra were then exported to Excel and were subjected to Principal Component Analysis (PCA) with leverage correction and mean-center cross validation boxes checked using Unscrambler V. 9.6 Camo Software AS (CAMO Software AS, Norway).

2.2.2.4 Scanning electron microscope

Hitachi Scanning Electron Microscopy (TM-1000, Hitachi Japan) with an accelerating voltage of 15kv was used to visualize the frozen and dehydrated samples. Fibers were placed either on a carbon disc or on a microscopy slide and no coating was performed prior to visualization.

2.3 Results and discussion

2.3.1 Fourier transform infrared analysis

Figures 1a and b show a series of FTIR spectra of fibers respectively from TX19 and TX55 cultivars during the elongation stage (10, 17, and 19 dpa), during the SCW formation stage (20, 24, and 30 dpa), and during the maturation stage (56 dpa). Compared to the FTIR spectra of mature cotton fibers, several additional vibration bands are noticed in the FTIR spectra of developing cotton fibers.

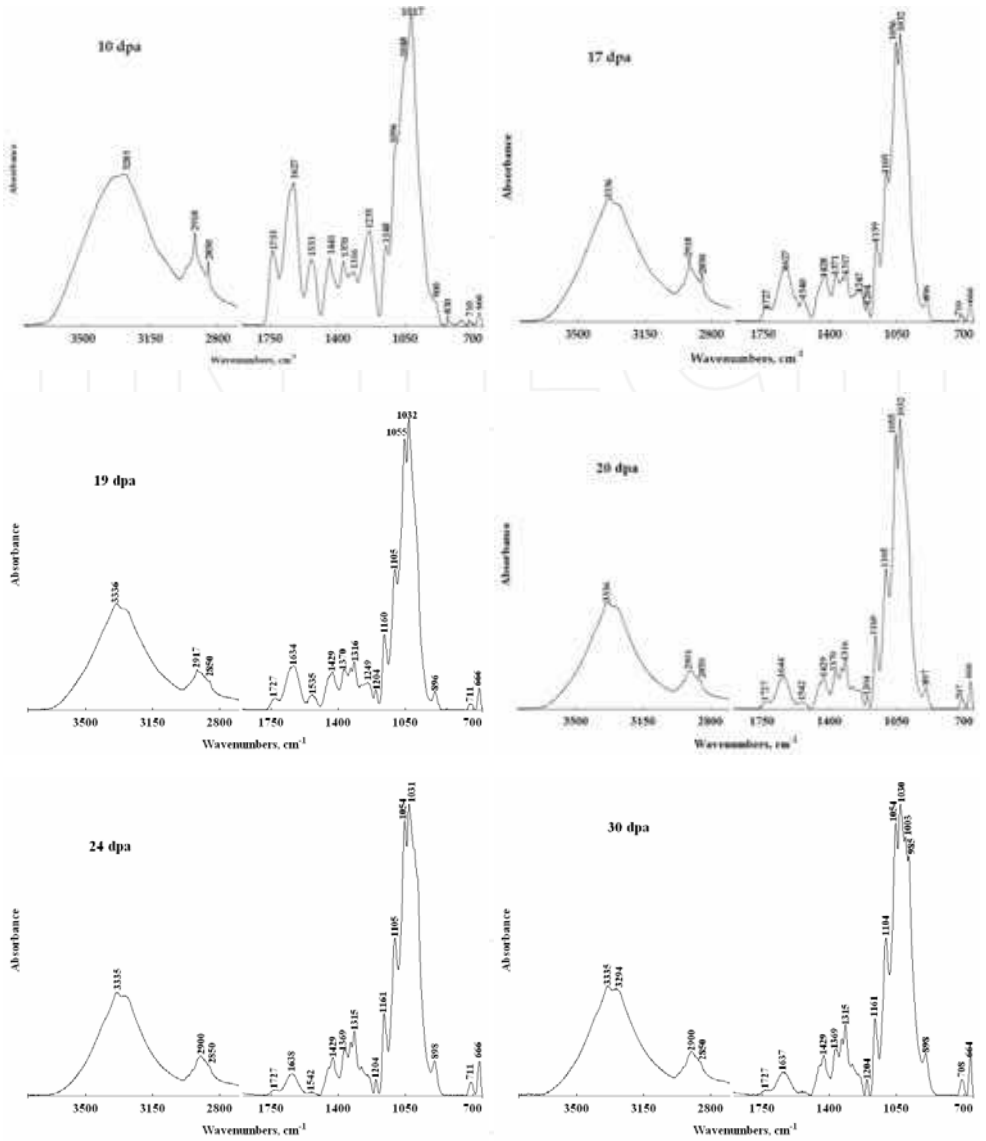


Fig. 1a. FTIR spectra of developing cotton (*Gossypium hirsutum* L. cv. TX19) fibers at different days post-anthesis (dpa)

The vibrations located at 2,918 and 2,850 cm^{-1} are attributed to $-CH_2$ asymmetric vibrations and could originate from the presence of wax substances present on the surface of the primary cell wall (Abidi et al., 2008). The intensities of these two peaks start decreasing at 19 dpa for both cultivars. This result is in agreement with our previous findings (Abidi et al.,

2008). We reported that the percent contribution of the primary cell wall to the total weight of the fiber decreased as wall thickness increases (increased fiber maturity, thus secondary cell wall). Therefore, the relative importance of the vibration bands attributed to noncellulosic substances (e.g. waxes which are located essentially on the primary cell wall) is less.

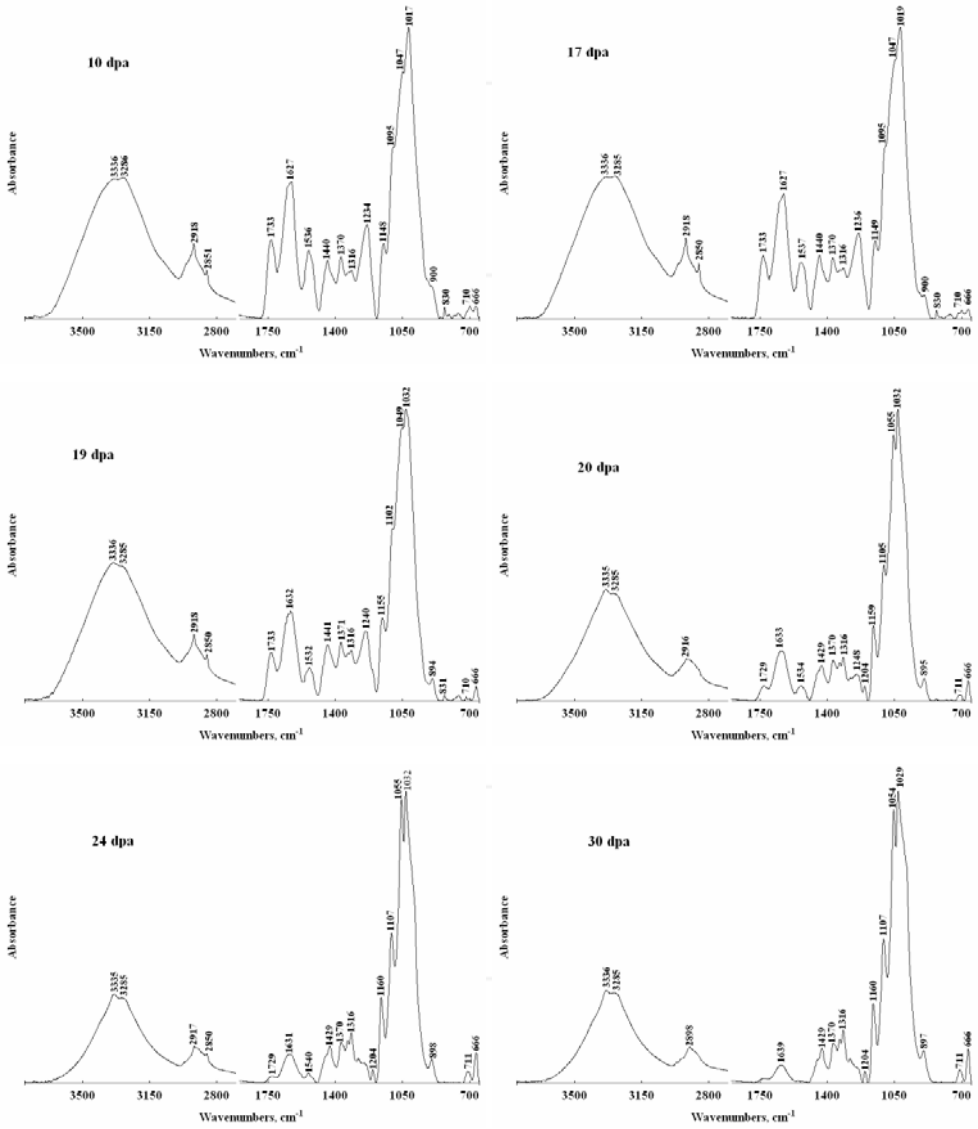


Fig. 1b. FTIR spectra of developing cotton (*Gossypium hirsutum* L. cv. TX55) fibers at different days post-anthesis (dpa)

In addition to O-H bending vibration of water molecules located at 1,627 cm^{-1}, the FTIR spectra of fibers showed strong absorption between 3,000 cm^{-1} and 3,600 cm^{-1} (which is attributed to O-H stretching vibration). This absorption band is composed of two vibrations located at 3,285 cm^{-1} (attributed to intermolecular hydrogen bonds) and 3,335 cm^{-1} (attributed to intra-molecular hydrogen bonds) (Liang & Marchessault, 1959). The intensity of the vibration 3,280 cm^{-1} decreased in the FTIR spectra of fibers older than 17 dpa from TX19 cultivar while for fibers from TX55 the decrease occurred at 19 dpa. The decrease in intensity of 3,280 cm^{-1} band is associated with an increase in the intensity of the vibration located at 3,335 cm^{-1}.

The vibration located at 1,733 cm^{-1} is attributed to C=O stretching vibration and could originate from esters or amides (Abidi et al., 2008). The integrated intensity of this peak (I_{1733}) was calculated between 1,780 cm^{-1} and 1,701 cm^{-1} and was reported as function of dpa for both cultivars TX19 and TX55 (Fig. 2a). As exhibited in this chart, for fibers from TX19 cultivar I_{1733} decreased sharply between 10 and 18 dpa and leveled-off at 19 dpa. However, for fibers from TX55 cultivar a continuous decrease is observed between 10 and 24 dpa. The statistical analysis (analysis of variance) showed significant effects of the developmental stage (dpa) and the cultivar on the integrated intensity (Table 1). These results indicate that the structural changes that occur during fiber development are influenced by the cultivar.

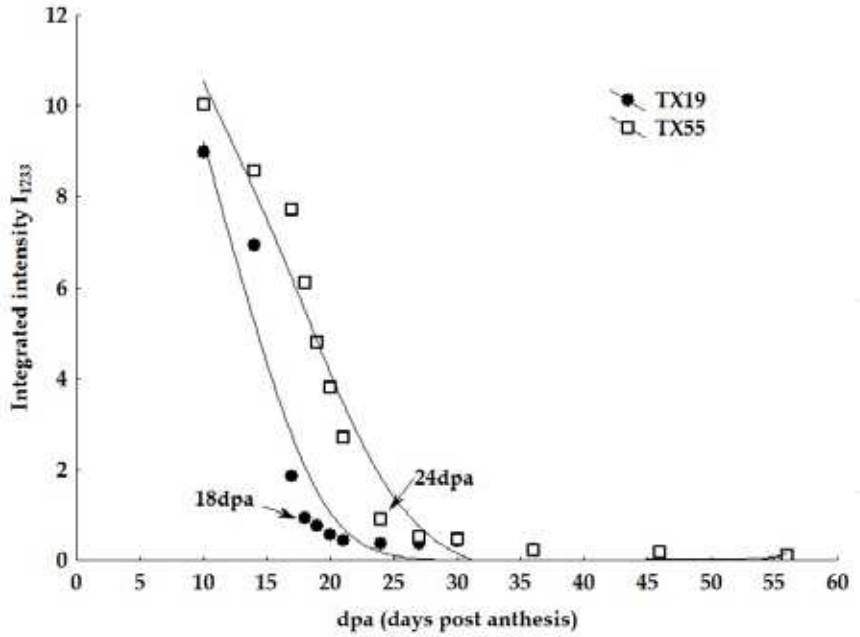


Fig. 2a. Evolution of the integrated peak intensity (I_{1733}) as function of developmental stages (dpa)

Parameter	df	F	Probability	Integrated intensity I_{1733} ¥
Intercept	1	574.4213	0.000001	
Cultivar	1	71.1905	0.000001	
dpa (days post anthesis)	12	60.0900	0.000001	
10				9.5049 a
14				7.7578 b
17				4.7894 c
18				3.5184 d
19				2.7784 de
20				2.1827 de
21				1.5740 ef
24				0.6320 f
27				0.4391 f
30				0.4388 f
36				0.2263 f
46				0.1651 f
56				0.0881 f
Cultivar*dpa	12	7.0895	0.000016	
Error	26			

df, degrees of freedom; F, variance ratio; ¥ Values not followed by the same letter are significantly different with $\alpha = 5\%$ (according to Newman-Keuls tests).

Table 1. Variance Analysis: Effect of developmental stage (day post-anthesis) and cultivars on the integrated intensity of the peak 1,733 cm^{-1} (I_{1733})

The vibration located at 1,627 cm^{-1} is attributed to O-H bending of adsorbed water molecules (Abidi et al., 2008). The integrated intensity of this peak (I_{1627}) was calculated between 1,701 cm^{-1} and 1,576 cm^{-1}. The change in the integrated intensity I_{1627} as function of dpa is reported in Fig. 2b for both cultivars. The statistical analysis (analysis of variance) showed a significant effect of the developmental stage and the cultivar on the integrated intensity (Table 2). For fibers from TX19 cultivar, the amount of adsorbed water decreased between 10 and 18 dpa, no significant changes are observed between 19 and 56 dpa. However, for fibers from TX55 cultivar the amount of adsorbed water decreased linearly until the fibers reached 24 dpa. The decrease in the amount of adsorbed water could be attributed to the decrease of the surface area and to reduced accessibility of water molecules to the internal hydroxyl groups to establish hydrogen bonding. This could be the result of increased polymerization reactions of glucose units to form cellulose macromolecules followed by increased crystallinity. Hsieh et al. reported that the degree of crystallinity of two cotton fiber cultivars (Maxxa and SJ-2) increases beginning 24 dpa (Hsieh et al., 1997). The results indicated that the big change in crystallinity occurred between 24 and 28 dpa, and no significant changes occurred thereafter. These results are in agreement with our FTIR results, which indicate that no change in the amount of adsorbed water is noticed between 27 and 56 dpa. In this developmental stage, fibers from both cultivars have nearly the same amount of adsorbed water. This could indicate that the remaining water molecules are those which are strongly bonded to cellulose macromolecules via hydrogen bonding.

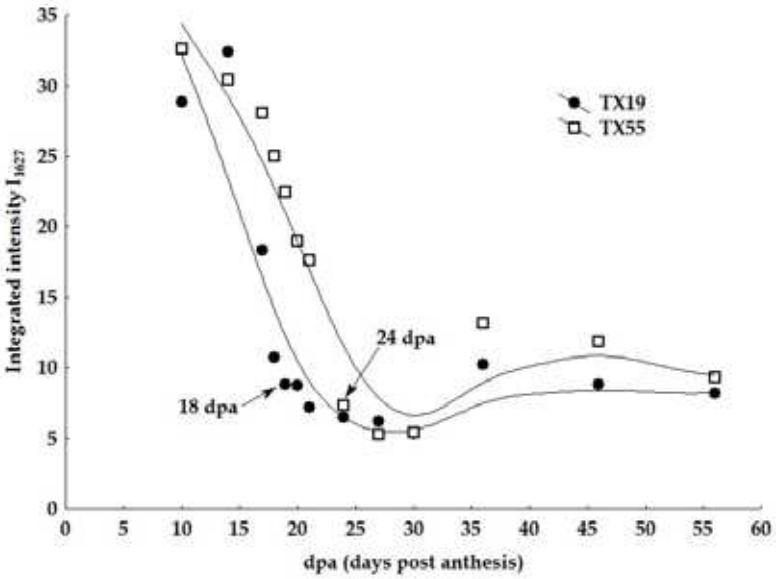


Fig. 2b. Evolution of the integrated peak intensity (I_{1627}) as function of developmental stage (dpa)

Parameter	df	F	Probability	Integrated intensity I_{1627} ¥
Intercept	1	1755.569	0.000001	
Cultivar	1	52.774	0.000001	
dpa (days post anthesis)	12	46.508	0.000001	
10				30.7341 a
14				31.4025 a
17				23.1798 b
18				17.8545 c
19				15.6207 cd
20				13.8401 cde
21				12.3703 ed
24				6.8814 fg
27				5.7146 g
30				5.3511 g
36				11.6973 def
46				10.3251 defg
56				8.7280 efg
Cultivar*dpa	12	4.916	0.000334	
Error	26			

df, degrees of freedom; F, variance ratio, ¥ Values not followed by the same letter are significantly different with $\alpha = 5\%$ (according to Newman-Keuls tests).

Table 2. Variance Analysis: Effect of developmental stage (day post-anthesis) and cultivars on the integrated intensity of the peak 1,627 cm^{-1} (I_{1627})

The vibration located at 1,534 cm^{-1} is attributed to NH_2 deformation and likely indicative of proteins or amino acids (Abidi et al., 2008). The integrated intensity of this peak (I_{1534}) was calculated between 1,575 cm^{-1} and 1,487 cm^{-1} and the evolution of I_{1534} as function of dpa for both TX19 and TX55 showed a behavior similar to I_{1733} (Fig. 2c). An abrupt decrease is observed between 10 and 18 dpa for fibers from TX19 cultivar and a continuous decrease between 10 and 24 dpa for fibers from TX55 cultivar. The statistical analysis (analysis of variance) showed a significant effect of both the developmental stage (dpa) and the cultivar on the integrated intensity (Table 3).

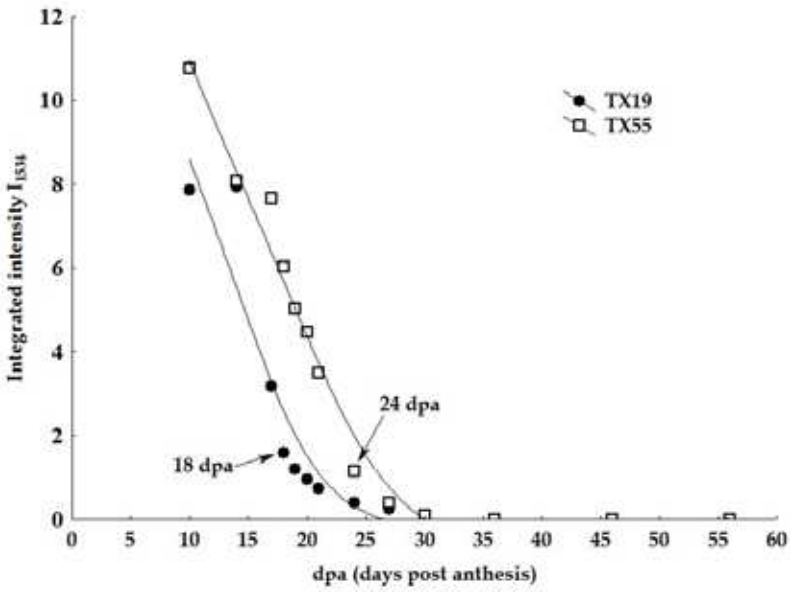


Fig. 2c. Evolution of the integrated peak intensity (I_{1534}) as function of developmental stage (dpa)

The vibration located at 1,236 cm^{-1} is attributed to C=O stretching or NH_2 deformation (Abidi et al., 2008). This vibration decreased in intensity during fiber development. The disappearance of this band is accompanied by the appearance of a vibration at 1,204 cm^{-1} at 20 dpa for fibers from TX55 cultivar and as sharp band at 17 dpa for fibers from TX19 cultivar. It is important to point out that the vibration located at 1,204 cm^{-1} appears only beginning 17 dpa for fibers from TX19 cultivar and beginning 20 dpa for fibers from TX55 cultivar. This vibration has been attributed by Ilharco et al. (1997) to C-O-C stretching mode of the pyranose ring. This vibration appears almost simultaneously with the vibration located at 900 cm^{-1} (attributed to β-linkage). Consequently, these two vibrations could be considered as the secondary cell wall's fingerprint.

Other vibrations located at 1,148, 1,096, and 1,048 cm^{-1} are present only as shoulders in the spectra of fibers from TX19 cultivar at 10 and 14 dpa and become sharper beginning at 17 dpa. It is also noteworthy the shift at 17 dpa of the vibration 1,148 cm^{-1} to 1,161 cm^{-1}, the vibration 1,096 cm^{-1} to 1,105 cm^{-1}, and the vibration 1,048 cm^{-1} to 1,056 cm^{-1}. However, for fibers from TX55 cultivar the shift of these vibrations to higher wavenumbers occurred at 20

dpa. The vibration located at 1,161 cm^{-1} is assigned to the anti-symmetric bridge C-O-C stretching vibration (Ilharco et al., 1997). The vibration located at 1,105 cm^{-1} is assigned to anti-symmetric in-plane ring stretching band (Ilharco et al., 1997). The vibration located at 1,056 cm^{-1} is attributed to C-O stretching mode (Abidi et al., 2008).

Parameter	df	F	Probability	Integrated intensity I_{1534} ¥
Intercept	1	474.3505	0.000001	
Cultivar	1	49.8492	0.000001	
dpa (days post anthesis)	12	48.3051	0.000001	
10				9.3057 a
14				8.0100 a
17				5.4175 b
18				3.8020 c
19				3.1142 c
20				2.7056 c
21				2.1114 c
24				0.7674 d
27				0.3300 d
30				0.0455 d
36				0.0000 d
46				0.0000 d
56				0.0000 d
Cultivar*dpa	12	4.3089	0.000888	
Error	26			

df, degrees of freedom; F, variance ratio, ¥ Values not followed by the same letter are significantly different with $\alpha = 5\%$ (according to Newman-Keuls tests).

Table 3. Variance Analysis: Effect of developmental stage (day post-anthesis) and cultivars on the integrated peak intensity of the peak 1,534 cm^{-1} (I_{1534})

The vibration located at 1,017 cm^{-1}, attributed to C-O stretch, is shifted to 1,031 cm^{-1} at 17 dpa for fibers from TX19 cultivar. However, for fibers from TX55 cultivar, this shift occurs at 19 dpa.

The vibrations located at 1,003 cm^{-1} and 985 cm^{-1} appeared at 30 dpa in the spectra of fibers from TX19 cultivar but only at 56 dpa in the spectra of fibers from TX55 cultivar. These vibrations are attributed to C-O and ring stretching modes (Liang & Marchessault, 1959).

The vibration located at 900 cm^{-1} is attributed to β-linkage (Abidi et al., 2008). This vibration is present only as a small shoulder in the FTIR spectra of fibers from both cultivars at 10 dpa but becomes sharper in the FTIR spectra of fibers from TX19 cultivar at 17 dpa and in the FTIR spectra of fibers from TX55 cultivar at 19 dpa.

The vibration located at 710 cm^{-1} is attributed to CH_2 rocking vibration in cellulose I_β (Schwanninger et al., 2004). Using 2D FTIR spectroscopy, Salmén et al. (2005) reported that the peak around 710 cm^{-1} (which is characteristic of cellulose I_β found in native cotton) has a linear correlation with the percentage of cellulose I_β of the crystalline part. We calculated the peak height of the vibration 710 cm^{-1} from all FTIR spectra and we reported the data as function of dpa for both genotypes as shown in Fig. 2d. The statistical analysis shows significant effects of the type of cultivar [$F(1,26)=122.318$, $p=0.000001$] and the developmental stage [$F(12,26)=121.334$, $p=0.00001$]. There is also an interaction

developmental stage * type of cultivar [F(12,26)=5.986, p=0.00008]. For fibers from TX19 cultivar the peak height of the vibration 710 cm^{-1} increases linearly between 10 and 30 dpa. No major change is noticed between 30 and 56 dpa. This indicates that the cellulose synthesis during the secondary cell wall development is accompanied by structural organization (increased crystallinity). However, for fibers from TX55 cultivar, the major change of the peak height occurred only beginning 21 dpa. This indicates that for fibers from TX55 cultivar, the structural organization of the cellulose is initiated few days later than in fibers from TX19 cultivar.

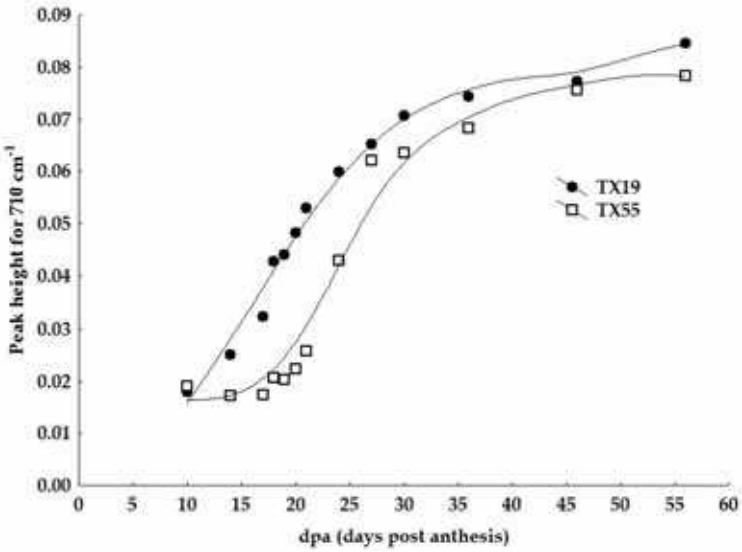


Fig. 2d. Evolution of the peak height for 710 cm^{-1} as function of developmental stage (dpa)

Figures 2e and f show the relationships between the integrated intensity of the peak 1,627 cm^{-1} (corresponding to adsorbed water) and the peak height of the vibration 710 cm^{-1} for fibers from TX19 and TX55, respectively. This relationships show that the decrease of the amount of adsorbed water is associated with an increase in the structural organization of the cellulose (increased crystallinity).

Principal Components Analysis (PCA) was performed on the FTIR spectra in order to identify distinct groups of spectra. The effect of this process is to concentrate the sources of variability in the data into the first 2 PCs (PC1 and PC2). The plots of PC1 against PC2 scores are depicted in Fig. 3a and b, respectively for fibers from TX19 and TX55 cultivars. For fibers from TX19 cultivar, two groups of FTIR spectra could be identified: group 1 includes FTIR spectra of fibers at 10, 14, and 17 dpa; and group 2 includes FTIR spectra of fibers from 18 to 56 dpa. For fibers from TX55 cultivar, two groups of FTIR spectra could also be identified: group 1 includes FTIR spectra of fibers from 10 to 21 dpa and group 2 includes FTIR spectra of fibers from 24 to 56 dpa. It is important to note that the transition between the two groups of FTIR spectra is different for these cultivars: for TX19, the transition occurs at 17 dpa and for TX55 the transition occurs at 21 dpa. Several studies have reported that the transition period between 16 and 21 dpa represents the

developmental switch in emphasis from primary to secondary cell wall synthesis during cotton fiber development (Wilkin & Jernstedt, 1999).

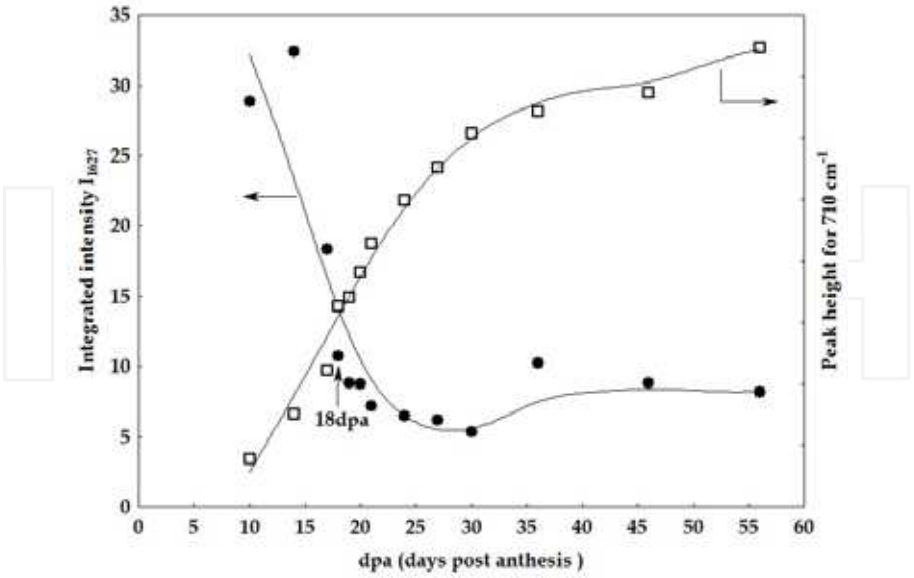


Fig. 2e. Relationship between the integrated peak intensity (I_{1627}) and the peak height for 710 cm^{-1} (*Gossypium hirsutum* L. cv. TX19)

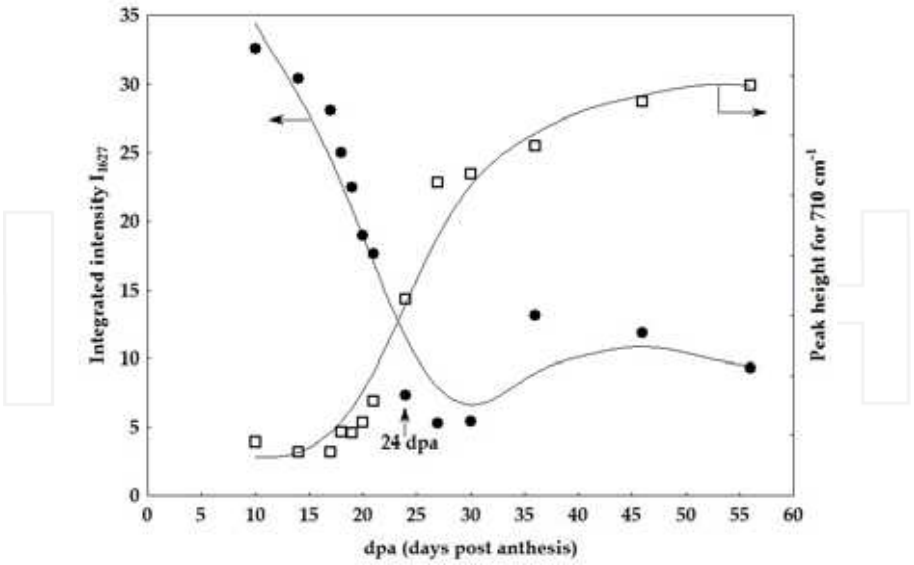


Fig. 2f. Relationship between the integrated peak intensity (I_{1627}) and the peak height for 710 cm^{-1} (*Gossypium hirsutum* L. cv. TX55)

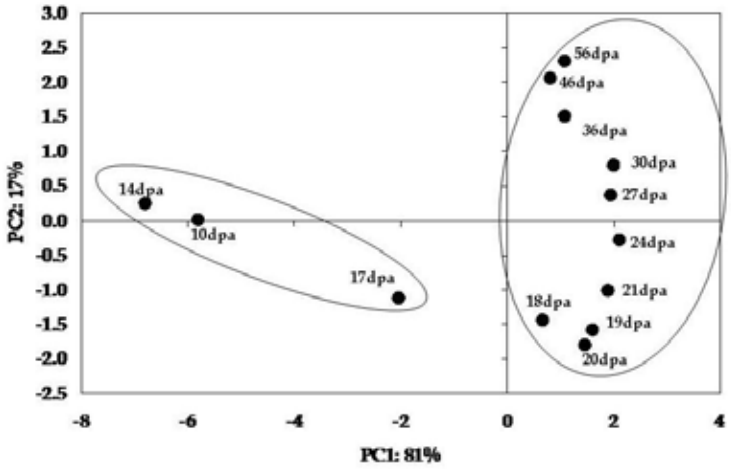


Fig. 3a. Principal Component Analysis of FTIR spectra of fibers from TX19 in the range 4,000 – 650 cm^{-1}

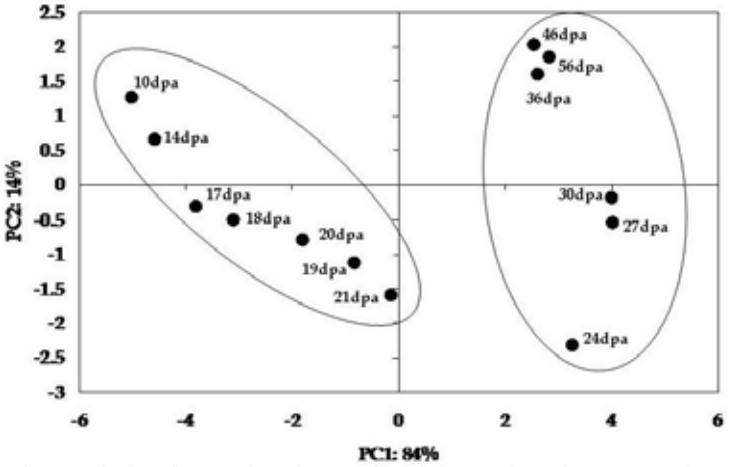


Fig. 3b. Principal Component Analysis of FTIR spectra of fibers from TX55 in the range 4,000 – 650 cm^{-1}

The FTIR data seem to indicate that the secondary cell wall synthesis in fibers from TX19 cultivar could start between 17 and 18 dpa. However, it could start between 21 and 24 dpa in fibers from TX55 cultivar. These data indicate that the FTIR spectroscopy technique has a potential use as screening tool for cotton fiber cultivars that have potentials for early maturation.

2.3.2 Scanning electron microscopy

Scanning electron microscopy micrographs of single fibers from selected developing stages are exhibited in Fig. 4. As illustrated in these micrographs, single cotton fibers at 17 dpa

from both cultivars appear transparent consisting mainly of primary cell walls. Single fiber from TX19 at this developmental stage appears slightly thicker than that of TX55. At 20 dpa, fibers from TX19 are thicker than those from TX55, signaling that the secondary cell wall synthesis in TX19 cultivar is well underway at this stage. Fibers from TX55, however, appear thicker only beginning at 24 dpa.

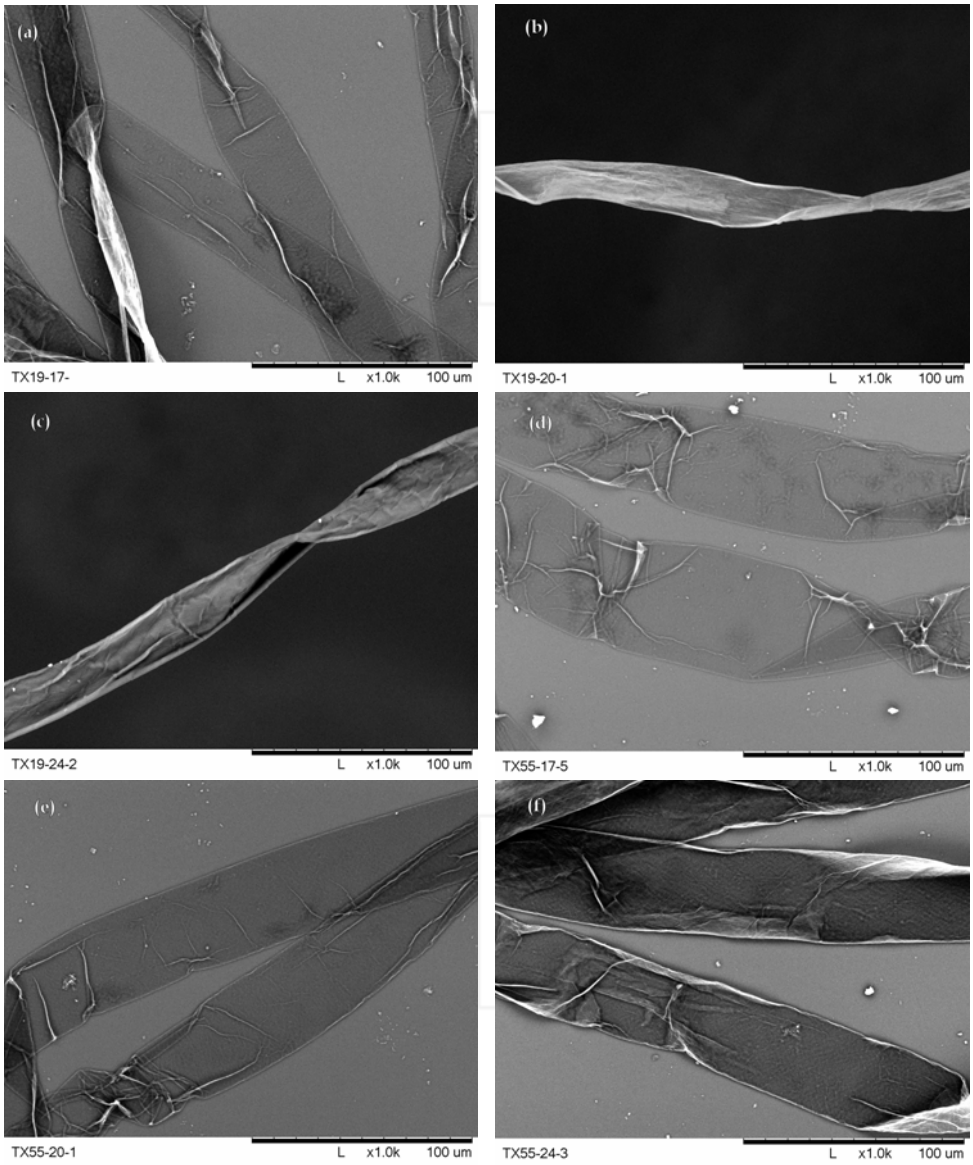


Fig. 4. Scanning electron microscopy micrographs of developing cotton fibers (A) TX19 17 dpa, (B) TX19 20 dpa, (C) TX19 24 dpa, (D) TX55 17 dpa, (E) TX55 20 dpa, (F) TX55 24 dpa

2.4 Conclusion

Fourier Transform Infrared spectroscopy was used to investigate the structural changes that occur during cotton fiber development starting at 10 days post-anthesis. FTIR spectra show pronounced differences during fiber development. The evolution of the integrated intensities of specific vibration bands located at 1,733, 1,534, and 1,627 cm^{-1} as function of developmental stages could be used to monitor the development of the secondary cell wall. FTIR results indicate that the two cultivars investigated (TX19 and TX55) exhibited different structural evolution. The results converge towards the conclusion that the transition phase between the primary cell wall and the secondary cell wall occurs between 17 and 18 dpa in fibers from TX19 cultivar, while this transition occurs between 21 and 24 dpa for fibers from TX55 cultivar. FTIR findings were supported by thermogravimetric characterization of fibers at different stages of development, changes in sugar composition as measured by High Performance Liquid Chromatography, and cellulose content as determined by the anthrone method (data not shown).

3. Cotton fiber physical properties prediction using FTIR

3.1 Introduction

There is universal interest in measuring cotton fiber properties that are useful to predict the performance of the fiber as an industrial raw material. In addition, significant efforts are being made to develop, through breeding and biotechnology, new cotton varieties that provide superior fiber properties. The "missing link" in these efforts is a scientific understanding of the relationships between the desired properties and the fiber structure/morphology. Among fiber properties, maturity is paramount for several reasons. The effect of maturity on the dye uptake is well known and constitutes the basis of the Goldthwait test (Goldthwait et al., 1947). Similarly, it is known that fine and mature fibers make it possible to spin a finer yarn. But maturity and fineness of cotton fibers are also essential qualitative characteristics if one wants to better understand the propensity of rupture of fibers when they are subjected to stress. It is intuitively obvious to hypothesize that immature fibers (having a thin, poorly developed secondary wall) will be fragile. Thus, they are likely to break during multiple mechanical stresses involved in transforming the fibers from the field to the yarn. These generate short fibers and neps (entanglements of fibers), which will result in yarn defects and decreased productivity.

The dominant tools used today to evaluate fiber properties are the High Volume Instruments (HVI) and the Advanced Fiber Information System (AFIS). The HVI is a fast instrument while the AFIS is relatively slow. Therefore, most of the cotton breeders rely only on HVI measurements. The HVI provides micronaire, length, strength, and color of the lint while the AFIS provides length, maturity, and fineness distributions as well as the nep count and the trash content. While the information provided by the HVI is necessary, it is not sufficient to provide answers about the structure/morphology that are needed to achieve new breakthroughs. The AFIS would be more appropriate but its relatively slow speed and high cost makes it non affordable by most of the cotton breeding programs. In an earlier work, Thermogravimetric Analysis was used to study the relationship between cotton fiber thermal properties and maturity and fineness (Abidi et al., 2007). The results showed that, low micronaires (immature and/or finer fibers), and low maturity ratios (immature fibers) are associated with high weight losses in the region 225 - 425°C. Both low micronaire and low maturity fibers exhibit high surface area for a given weight of fibers.

Therefore, the quantity of the non-cellulosic compounds, mostly present on and in the primary cell wall, is quite large. This leads to higher weight losses in the region 225 - 425°C because in addition to the pyrolysis reactions of cellulose, the decomposition of non-cellulosic compounds also takes place between 225 - 425°C. Furthermore, it was showed that it is possible to estimate the width of the primary cell wall, by comparing the weight losses of two cotton fibers that are identical except for having different maturities (i.e. different degrees of secondary cell wall development).

Our hypothesis is that the FTIR spectroscopy has the potential to assess quickly and accurately some fiber properties related to cotton fiber maturity and fineness. Ramey (1982) used Near-infrared reflectance (NIR) to estimate some quality components of natural fibers. The author explored the use of the NIR to predict the cross-sectional area, the specific surface, micronaire, and causticaire maturity index of cotton fibers. A Neotec Model 41 Grain Quality Analyzer was used to conduct the study. This instrument was equipped with 1.53, 1.97, and 2.32 μm central wavelength filters. The approximate wavelengths for data collection were 1.49 to 1.51, 1.90 to 1.93, 2.16 to 2.19, and 2.26 to 2.30 μm. The author concluded that meaningful estimates of the four mentioned fiber properties could be obtained with near-infrared reflectance. In addition, the author indicated that, except for micronaire, when performing near-infrared reflectance measurements, technician and instrument time per specimen is less than the usual procedures.

Montalvo and Von Hoven (2004) made a comprehensive review on both the reference methods and the application of NIR to predict cotton fiber properties (fineness, maturity, and micronaire). The authors indicated that successful application of NIR spectroscopy to cotton fiber quality measurements has been limited by shortcuts in the development and validation of NIR instruments and methodologies.

In this study, we used the Universal Attenuated Total Reflectance Fourier Transform Infrared (UATR-FTIR) to predict cotton fiber properties. The FTIR spectra were acquired in the mid-infrared range (4,000 – 650 cm^{-1}). After baseline correction and spectra normalization, the Partial Least Square was used to predict micronaire, maturity, and surface area.

3.2 Materials

One hundred and four cotton bales representing the two principal cultivated species were selected. These cotton samples were the same cotton samples used in our previous study (Hequet et al., 2006). The vast majority of the bales originated in the U.S.A., but some foreign-grown cotton bales were also selected (Egypt, Uzbekistan, Pakistan, Cameroon, Syria, Benin, and Australia). The bales were opened and ten samples per bale were taken. Each sample was tested using a HVI Uster 900A. For each bale, a total of 40 micronaire tests and 100 length and tenacity tests were done. This allowed us to conclude that the intra-bale variability was acceptable and that we had a wide range of fiber properties (Table 4). The same samples were also tested on the AFIS with 5 replications and 3,000 fibers tested per replication. This totaled 150,000 fibers per bale.

A representative sample of approximately 30 kg (70 pounds) was taken from each bale. Each sample was homogenized according to the protocol used by the ICCSC (International Cotton Calibration Standard Committee, 1999 (USDA, 1999) to produce reference cottons. From the card web produced, 20 samples were taken. Samples 1 to 5 were re-sampled (8 pinches per sample). This new sample was delicately mixed manually then 2 fibrograph

combs were formed. We chose to sample with the fibrosampler because this method, unlike Lord's method, is not length biased. This procedure was repeated for samples 6-10, 11-15 and 16-20. A sample was then taken from each of the 8 combs produced. For each comb, a minimum of 500 cross sections were analyzed following the procedure described in (Hequet et al., 2006). When the CV% of the averages between combs was higher than 2%, eight other combs were produced and 500 additional cross sections (per comb) were analyzed. The original CV% was confirmed in almost every case. From this point forward the 104 cottons homogenized according to the ICCS protocol will be called reference cottons.

	Micronaire	UHML (inch)	UI (%)	Tenacity (cN/tex)	Elongation (%)
Average :	4.2	1.08	81.9	29.4	5.4
Minimum	2.6	0.96	78.6	21.6	3.6
Maximum	5.7	1.31	84.5	41.3	8.3
Intra-bale CV% :					
Average	0.7	0.9	0.6	1.9	3.5
Minimum	0.0	0.4	0.2	0.8	1.4
Maximum	2.2	2.2	1.1	3.1	7.7

Table 4. HVI fiber properties of the 104 bales: Basic statistics

3.3 Methods

3.3.1 FTIR measurements

Spectrum-One equipped with an UATR (Universal Attenuated Total Reflectance) accessory was used to acquire the FTIR spectra of the cotton fiber samples (PerkinElmer, USA). The UATR-FTIR was equipped with a ZnSe-Diamond crystal composite that allows collecting the FTIR spectra directly on a sample without any special preparation. The FTIR spectrometer was placed in a conditioned laboratory at 65±2%RH and 21±1°C and all the FTIR spectra were acquired in this environment. Cotton samples were placed on top of the ZnSe-Diamond crystal and a pressure was applied on the sample to ensure a good contact between the sample and the incident IR beam and to prevent loss of the IR beam. The amount of pressure applied was kept the same and it was monitored through the FTIR software included in the Perkin-Elmer software package. Therefore, all samples were subjected to the same pressure.

A representative sample was taken from each reference cotton. Thirty spectra from each sample were acquired and analyzed. All the FTIR spectra were collected at a spectrum resolution of 4 cm^{-1}, with 32 co-added scans, over the range of 4,000 - 650 cm^{-1}. A background scan of clean ZnSe-Diamond crystal was acquired before acquiring the spectra of the sample. The Perkin-Elmer software was used to perform the baseline corrections of the spectra.

3.3.2 FTIR spectra analysis

The raw FTIR spectra were normalized (centered and scaled) and averaged. Then, the PLS1 (Partial Least Square) was run with a full cross-validation (The Unscrambler V. 9.6 Camo Software AS, USA). An uncertainty test was performed for each test. Full cross validation means that the same samples are used both for model estimation and testing. One sample is left out from the calibration data set and the model is calibrated on the remaining data points. Then the value for the left-out sample is predicted and the prediction residuals are computed. The process is repeated with another subset of the calibration set, and so on until every sample has been left out once; then all prediction residuals are combined to compute the validation residual variance and RMSEP (Root Mean Square Error of Prediction). RSMEP can be interpreted as the average prediction error. Therefore, we can give the predicted Y-values an estimated precision of two times the RSMEP.

3.4 Results and discussion

For each sample in addition to the FTIR measurements, the following fiber properties were selected: HVI micronaire, image analysis of the cross-sections (perimeter and theta), and the AFIS data (calculated Area mm^2/mg). PLS1 (Partial Least Square) procedure, was used to predict these fiber maturity-related parameters. Partial Least Square regression is a method that relates the variations in one response variable (micronaire for example) to the variations of several predictors (X variables = absorbance for the different wavelengths) with explanatory or predictive purposes. This method performs particularly well when the various X-variables express common information, i.e. when there is a large amount of correlation, or even collinearity between them, which is the case with the FTIR spectra. Partial Least Square regression is a bilinear modeling method where information in the original FTIR data is projected onto a small number of underlying ("latent") variables called PLS components. The Y-data are actively used in estimating the "latent" variables to ensure that the first components are those that are most relevant for predicting the Y-variables. Interpretation of the relationship between X-data and Y-data is then simplified as this relationship is concentrated on the smallest possible number of components.

3.4.1 Prediction of the micronaire

Cotton fiber micronaire is a function of fineness and maturity. It is based on the measurement of an air flow that passes through a porous plug of cotton fibers. The micronaire is proportional to the inverse of the specific surface of the cotton fibers. It should be noted that the definition of fiber fineness in cotton does not relate directly to fiber perimeter. Indeed, fiber fineness (expressed in millitex) is the weight in mg of 1,000 meters of fibers. Therefore a fine fiber may have a small perimeter and a high maturity ratio. Conversely, a fine fiber may have a large fiber perimeter and a low maturity ratio (which implies a large lumen in the fiber). In a similar manner, high micronaire readings indicate coarse fibers (high weight per unit length), while low micronaire readings indicate fine fibers (low weight per unit length).

Figure 5a shows the relationship between the FTIR and the HVI micronaire. With a coefficient of determination of 0.9252, the FTIR measurements lead to very good prediction of the micronaire (FTIR prediction = 0.9253*Micronaire + 0.3349). It should be noted that the coefficient of determination of the validation (R^2 = 0.8677) set is nearly as good as the coefficient of determination of the calibration set (fig. 5b).

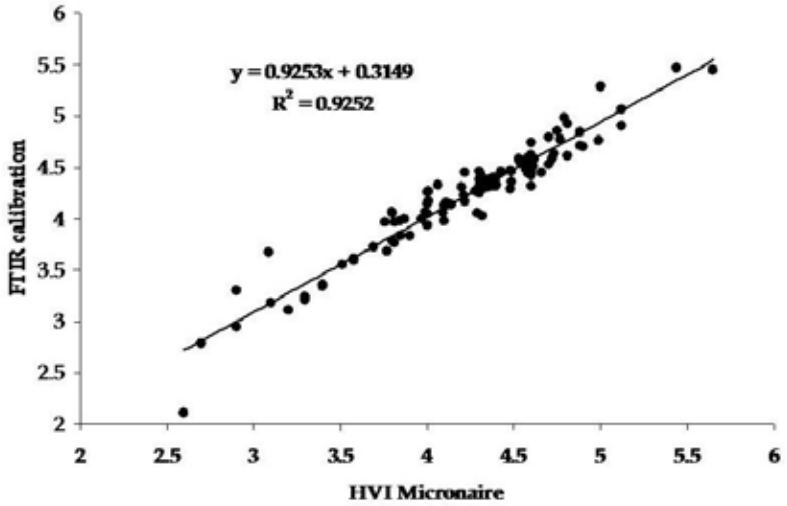


Fig. 5a. FTIR HVI micronaire prediction (calibration) versus HVI micronaire

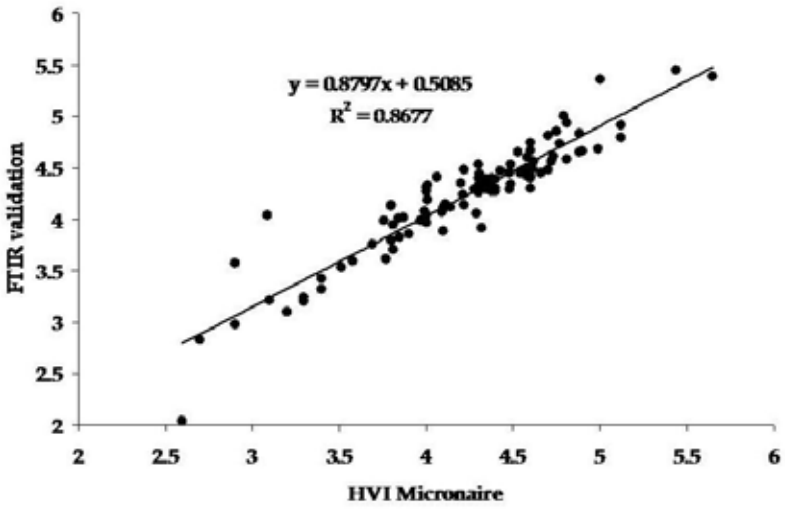


Fig. 5b. FTIR HVI micronaire prediction (validation) versus HVI micronaire

3.4.2 Prediction of fiber perimeter and theta

Fiber perimeter and theta were determined by image analysis (Hequet et al., 2006). Theta, degree of secondary cell wall thickening, is defined as the ratio of the area of the cell wall to the area of a circle having the same perimeter as the fiber cross-section. The measurement of theta leads to the determination of the maturity ratio M ($\theta = M*0.577$). Figure 6a shows the prediction of the fiber perimeter with FTIR (FTIR prediction = 0.6529*Perimeter + 18.308).

The coefficient of determination (R^2 = 0.6529) of the calibration set is not very good. For the validation set, the coefficient of determination (R^2 = 0.4548) is much lower (Fig. 6b). These results reveal that the PLS1 model is not valid.

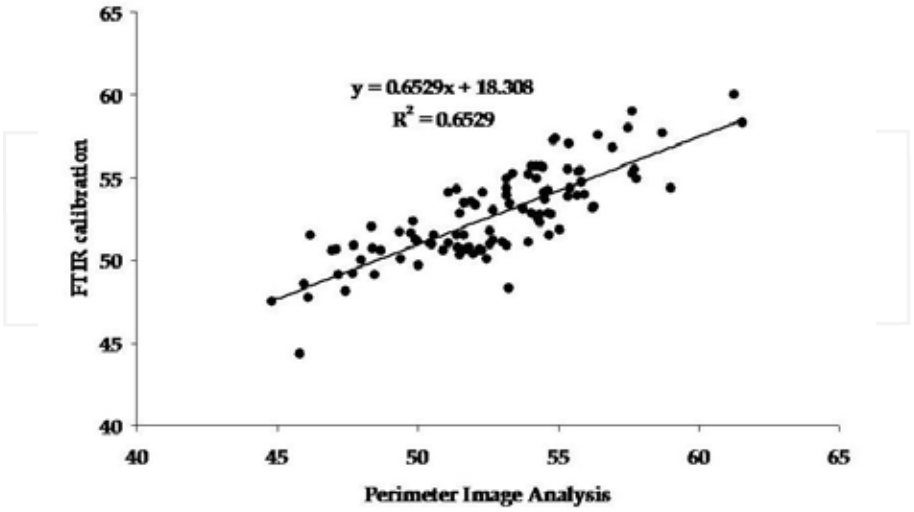


Fig. 6a. FTIR Perimeter Image Analysis prediction (calibration) versus Perimeter Image Analysis

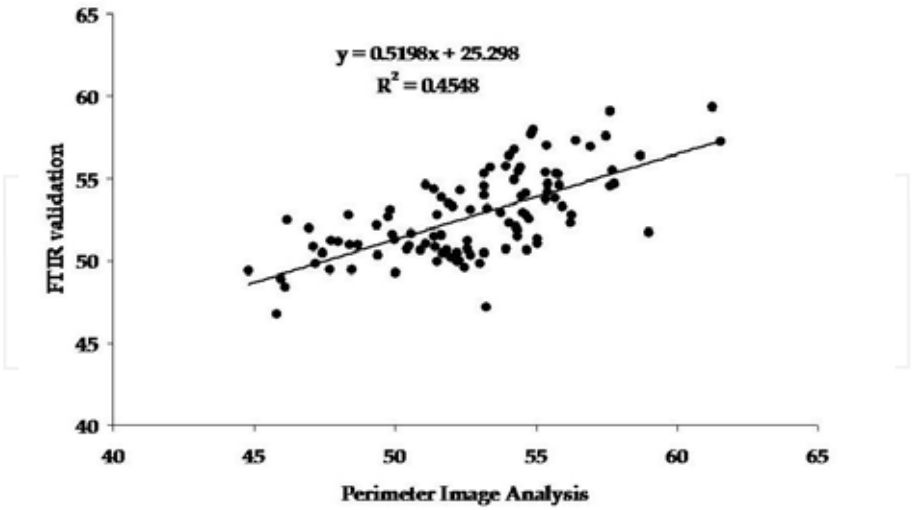


Fig. 6b. FTIR Perimeter Image Analysis prediction (validation) versus Perimeter Image Analysis

Figure 7a shows the prediction of theta (maturity) with FTIR (FTIR prediction = 0.8107*Theta + 0.0955). The coefficient of determination (R^2 = 0.8106) of the calibration set is acceptable. However, the coefficient of determination (R^2 = 0.6981) of the validation set is much lower than the one of the calibration set, revealing that a good prediction of maturity with this method (reflectance) is probably not possible (Fig. 7b).

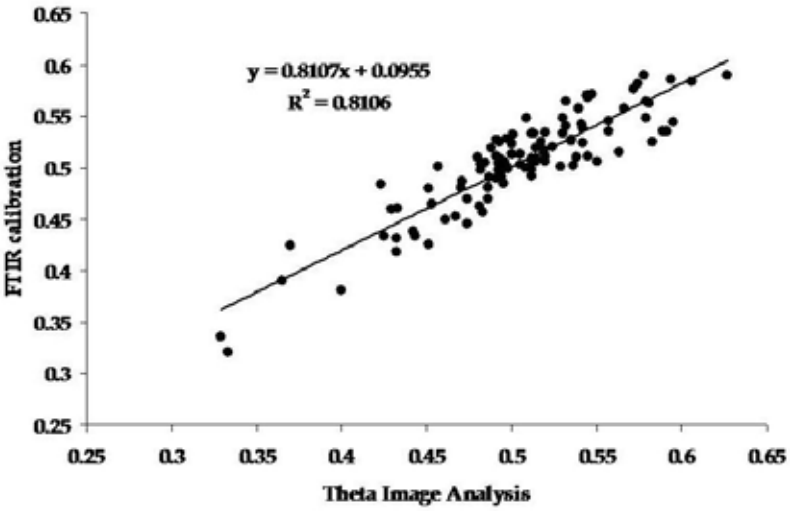


Fig. 7a. FTIR Theta Image Analysis prediction (calibration) versus Theta Image Analysis

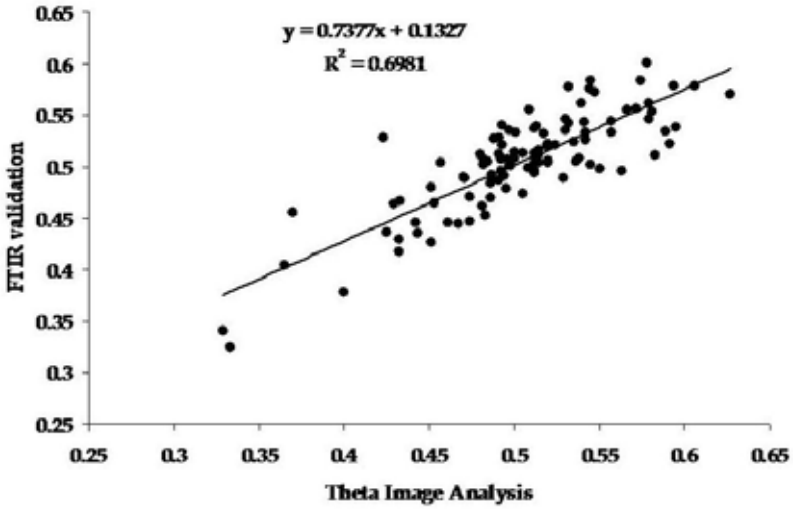


Fig. 7b. FTIR Image Analysis Theta prediction (validation) versus Theta Image Analysis

3.4.3 Prediction of the surface area

The results obtained with the AFIS for both Standard fineness and maturity are very similar to the one obtained with image analysis. From AFIS data, the surface area (expressed in mm^2/mg) was calculated. Fig. 8a and b shows the prediction of the surface area by FTIR (FTIR prediction = 0.9191*surface area + 26.264). Both the coefficient of determination of calibration set ($R^2 = 0.9191$) and the validation ($R^2 = 0.8645$) are good.

Estimating the specific surface with two drastically different methods (the micronaire and the AFIS) led to the same conclusion. It is possible to predict with the FTIR the Area in mm^2 per mg of fibers.

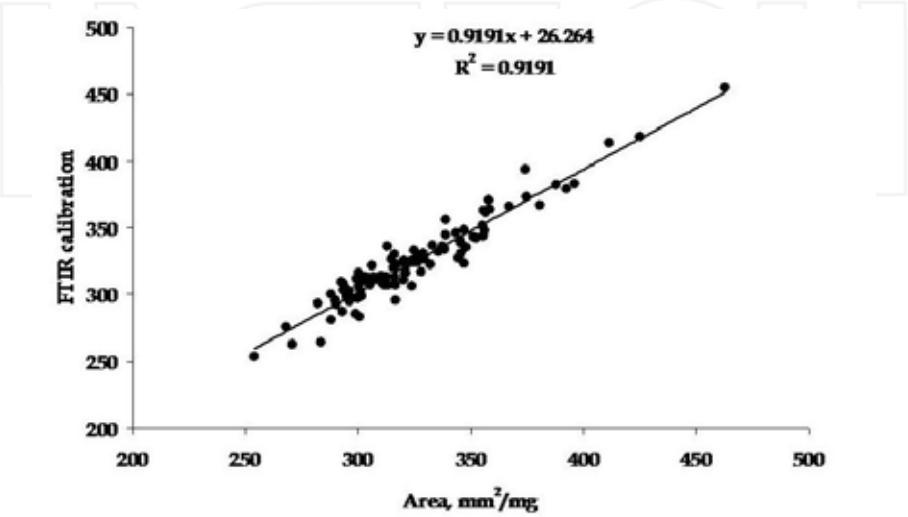


Fig. 8a. FTIR Area (mm^2/mg) prediction (calibration) versus Area (calculated from AFIS data)

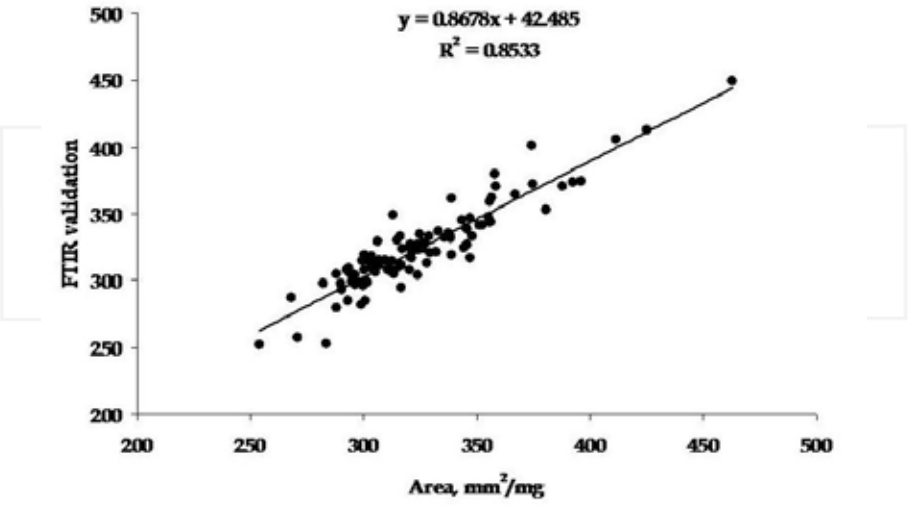


Fig. 8b. FTIR Area (mm^2/mg) prediction (validation) versus Area (calculated from AFIS data)

3.5 Conclusion

In this study we investigated the use of the Universal Attenuated Total Reflectance Fourier Transform Infrared spectroscopy to evaluate the cotton fiber properties. One hundred and four cotton samples were tested and 30 FTIR spectra were acquired and analyzed from each sample. The Partial Least Square (PLS) procedure was performed on the FTIR spectra. The results showed that the micronaire and the surface area (calculated from the AFIS data) could be predicted from the FTIR measurements with very high coefficient of determination. Two drastically different techniques to estimate the surface area of the cotton fiber (micronaire and AFIS data) led to the same conclusion. However, the prediction of fiber maturity is probably not possible with FTIR because of the low coefficient of the determination. It was concluded that, to be able to predict the fiber maturity with the FTIR, it is necessary to perform the measurement in the transmission mode.

4. Acknowledgments

The authors would like to thank the Texas Department of Agriculture/Food and Fibers Research Grant Program, Cotton Incorporated, and the International Cotton Research Center/USDA for the financial support.

5. References

Abidi, N., Hequet, E., & Ethridge, D. (2007). Thermogravimetric analysis of cotton fibers: relationships with maturity and fineness. *J Appl Polym Sci* , 103: 3476-3482

Abidi, N., Hequet, E., Cabrales, L., Gannaway, J., Wilkins, T., & Wells, LW. (2008) Evaluating cell wall structure and composition of developing cotton fibers using Fourier transform infrared spectroscopy and thermogravimetric analysis. *J Appl Polym Sci* 107:476-486

Carpita , NC., Defernez, M., Findlay, K., Wells, B., Shoue, DA., Catchpole, G., Wilson, RH., & McCann, MC. (2001) Cell wall architecture of the elongating Maize Coleoptile. *Plant Physiol*, 127: 551-565

Chen, L., Carpita, NC., Reiter, W-D., Wilson, RH., Jeffries, C., & McCann, MC. (1998) A rapid method to screen for cell-wall mutants using discriminant analysis of Fourier transform spectra. *Plant J*16(3): 385-392

Goldthwit, C.F.; Smith, H.O.; & Barnett, M.P. New dye technique shows maturity of cotton. Textile World, July 1947, p 105.

Gokani, SJ., Kumar, R., & Thaker, VS. (1998). Potential Role of Abscisic Acid in Cotton Fiber and Ovule Development. *JPlant Growth Regul* 17:1-5

Haigler, CH., Zhang, D., & Wilkerson, CG. (2005). Biotechnological improvement of cotton fibre maturity. *Physiol Plant* 124:285-294

Hequet, E., Wyatt, B., Abidi, N., & Thibodeaux, D.P. (2006). Creation of a set of reference materials for cotton fiber maturity measurements. Text Res J 76(7): 576-586.

Huwyler, HR., Franz, G., & Meier, H. (1979). Changes in the composition of cotton fibre cell walls during development. *Planta* 146:635-642

Hsieh, Y-L., Hu, X-P., & Nguyen, A. (1997). Strength and crystalline structure of developing Acala cotton. *Textile Res J*67(7):529-536

Ilharco, LM., Garcia, AR., Lopez, da Silva J., & Vieira Ferreira, LF. (1997), Infrared approach to the study of adsorption on cellulose: influence of cellulose crystallinity on the adsorption of benzophenone. *Langmuir* 13:4126-4132

Liang, C-Y., & Marchessault, RH. (1959). Infrared spectra of crystalline polysaccharides. II. Native celluloses in the region from 640 to 1700 cm^{-1}. *JPolym Sci* XXXIX:269-278

Maltby, D., Carpita, NC., Montezinos, D., Kulow, C., & Delmer, DP. (1979). β-1,3-glucan in developing cotton fibers, structure, localization, and relationship of synthesis to that of secondary wall cellulose. *Plant Physiol* 63:1158-1164

McCann, MC., Chen, L., Roberts, L., Kemsley, EK., Sene, C., Carpita, NC., Stacey, NJ., & Wilson, RH. (1997). Infrared microspectroscopy: sampling heterogeneity in plant cell wall composition and architecture. *Physiol Plant* 100:729-738

McCann, MC., Hammouri, N., Wilson, R., Belton, P., & Roberts, K. (1992). Fourier transform infrared microspectroscopy is a new way to look at plant cell walls. *Plant Physiol* 100:1940-1947

McCann, MC., Defernez, M., Urbanowicz, BR., Tewari, JC., Langewisch, T., Olek, A., Wells, B., Wilson, RH., & Carpita, NC. (2007). Neural network analyses of infrared spectra for classifying cell wall architectures. *Plant Physiol* 143:1314-1326

McCann, MC., Stacey, NJ., Wilson, R., & Roberts, K. (1993). Orientation of macromolecules in the walls of elongating carrot cells. *JCell Sci* 106:1347-1356

Meinert, MC., & Delmer, DP. (1977). Changes in biochemical composition of the cell wall of the cotton fiber during development. Plant Physiol. 59:1088-1097

Montalvo, J G.& Von Hoven, T. (2004). Analysis of Cotton. In Near-Infrared Spectroscopy in Agriculture. *Agronomy Monograph* no. 44. 2004. Crop Science Society of America.

Muller, LL. & Jacks TJ. (1975). Rapid chemical dehydration of samples for electron microscopic examinations. *JHistochem Cytochem* 23(2):107-110

Rajasekaran, K., Muir, AJ., Ingber, BF., & French, AD. (2006). A dehydration method for immature or wet cotton fibers for light and electron microscopy. *Textile Res. J* 76(6):514-518

Ramey, H.H. (1982) Estimating quality components of natural fibers by Near-infrared reflectance: Part 1: Cotton fiber cross-sectional area and specific surface. *Text Res J*, 52, 20-25

Timpa, JD. & Triplett, BA. (1993). Analysis of cell-wall polymers during cotton fiber development. *Planta* 189:101-108

Tokumoto, H., Wakabayashi, K., Kamisaka, S., & Hoson, T. (2002). Changes in the sugar composition and molecular mass distribution of matrix polysaccharides during cotton fiber development. *Plant Cell Physiol* 43(4):411-418

Salmén, L., Åkerholm, M., & Hinterstoisser, B. (2005). Two-dimensional Fourier transform infrared spectroscopy applied to cellulose and paper. In: Dumitriu S (ed) Polysaccharides, structural diversity and functional versatility, 2nd edn. Marcel Dekker, New York, pp159-187

Schwanninger, M., Rodrigues, JC., Pereira, H., & Hinterstoisser, B. (2004). Effects of short-time vibratory ball milling on the shape of FT-IR spectra of wood and cellulose. *Vib Spectrosc* 36:23-40

Séné, CFB., McCann, MC., Wilson, RH., & Grinter, R. (1994). Fourier-transform Raman and Fourier-transform infrared spectroscopy an investigation of five higher plant cell walls and their components. *Plant Physiol* 106:1623-1631

USDA, International Calibration Cotton Standards Program, USDA-AMS, Cotton Program, Memphis, TN, March 1999.

Wilkins, TA., & Jernstedt, JA. (1999). Molecular genetics of developing cotton fibers. In: In: Basra AS (Ed) Cotton fibers, developmental biology, quality improvement, and textile processing, edn. Food Products Press, New York, pp 231-269

Yong, W., Link, B., O'Malley, R., Tewari, J., Hunter, CT., Lu, C-An., Li, X., Bleecker, AB., Koch, KE., McCann, MC., McCarty, DR., Patterson, SE., Reiter, W-D., Staiger, C., Thomas, SR., Vermerris, W., & Carpita, NC. (2005). Genomics of plant cell wall biogenesis. *Planta* 221: 747-751

6

Medical Image Registration with Fourier basis Functions

Roberto A. Isoardi[1], Amílcar R. Osorio[2] and Germán Mato[3]
[1]*Fundación Escuela de Medicina Nuclear, Mendoza*
[2]*Fundación Centro Diagnóstico Nuclear, Buenos Aires*
[3]*Grupo Física Estadística, Centro Atómico Bariloche, Bariloche*
Argentina

1. Introduction

Registration is one of the most interesting, yet challenging computer-aided tasks in medical image processing, aimed at bringing two or more data sets into spatial and/or temporal correlation. If the represented data are medical images, there are countless situations where it is of interest to attain such correlation, as it has become routine practice in many diagnostic and image-guided therapeutic procedures.

For example, one of the most frequent clinical applications is to align two scans of a given patient, e.g. for easy identification of equivalent structures on both registered images, follow-up of disease, etc. Another use of image registration is commonplace during brain activation studies, where several functional Magnetic Resonance (*f*MRI) or Positron Emission Tomography (PET) scans are repeated on the same or different subjects while receiving sensorial or cognitive stimulations. In order to perform a statistical analysis on the brain images, it is often necessary to register all data sets with respect to a brain atlas. In this way, both statistical power and signal-to-noise ratio are enhanced in what is known as intra-modality registration (Friston et al, 2006).

One of the most exciting applications is inter-modality registration, e.g. correlating PET and MR (Magnetic Resonance) scans of a same subject, or PET vs. CT (Computed Tomography). PET imaging provides distinctive functional and metabolic information, but lacks the high anatomical resolution which is in turn provided by conventional MR or CT. Thus PET/CT registration is synergic, since it facilitates the location of malignancies in their anatomical context. Moreover, this modality combination is often useful for Radiotherapy Planning (Townsend, 2008).

An automatic image registration algorithm normally includes a *floating image* to be aligned to the coordinate system of a *reference image*. To do this, a spatial transformation function –containing a number of parameters-, must be proposed and applied to the former data set. The parameters are chosen in such a way that a proposed *similarity measure* between both data sets is optimized. This measure may be –for example-, the Normalized Mutual Information (NMI) or the Cross Correlation Coefficient.

Methods which apply rigid-body transformations perform reasonably well for registering images of the head (brain) and extremity portions. In a 3D space, a rigid-body transformation involves only six parameters, e.g. three translations (Δx, Δy, Δz) and three

rotation angles (θ_x, θ_y, θ_z). A natural extension of the rigid approach is the use of *affine* transformations that include scaling and shearing (Maintz & Viergever, 1998).

However, those simple transformations often yield unacceptable results when dealing with deformable body regions such as thorax and abdomen. Even with a careful repositioning on the scanner table, the body region of interest may displace between different scans with respect to the scanner coordinate system, due to various reasons, which may include voluntary and involuntary motion (e.g. respiration, cardiac function), misplacement errors, variations in morphology and physiology, etc.

Several non-rigid or deformable methods have been widely explored to treat this problem (Crum et al, 2004)(Rueckert & Aljabar, 2010). This is an open field of research, both for its complexity and situations of interest. Unlike rigid models, deformable registration is particularly difficult to validate with clinical images, where internal fiducial markers are not available. Besides, the optimization of the objective function may be non-unique or even physically meaningless.

Nevertheless, the main complexity of a deformable system lies on its many degrees of freedom. In an extreme situation, it would be necessary to propose a 3D independent transformation for each image voxel. In this case, the number of transformations would be three times the number of voxels. To treat this problem a set of approximations is implemented, otherwise the optimization would be very costly in such a parameter space. Most methods apply a rigid transformation as a first approximation, and then improve registration with a non-rigid approach with a small set of *basis functions*.

These functions form the basis for a vector space, where any function can be represented as a linear combination of those basis functions. They can be polynomial $\{1, t, t^2\}$, just like those used to build splines (Bookstein 1989). For instance, spline basis functions are used when fiducial markers are available or with physical model-based algorithms (e.g. elastic deformations, viscous fluids, etc) (D'Agostino et al, 2003).

One particular kind of functions that make up an orthonormal basis are trigonometric $\{sin(n\pi x), cos(n\pi x)\}$, where $n \geq 0$ is the order of the basis function. In this chapter we present a systematic analysis for the deformable registration problem using trigonometric Fourier basis functions (Ashburner & Friston, 1999). With these functions, the typical size of the deformation field can be easily controlled, ranging from long to short wavelengths. By combining this approach with a volume subdivision scheme, we expect to apply a small set of basis functions at a given stage of the algorithm. In principle, given that Fourier functions make a complete set, any arbitrary deformation may be approximated only by increasing the number of basis functions.

In our study, we analyze intra-modality registration (CT-CT, MR-MR, PET-PET) in 3D. For each modality, we determine the optimum number of coefficients (transformation order) for the basis functions and the number of subvolumes to attain a satisfactory registration within a reasonable computing time.

As a similarity measure, we calculate the Normalized Mutual Information (NMI) for different transformation orders and applying the algorithm in the thoraco-abdominal region. Each clinical volume data set was artificially deformed using a known displacement field, in order to simulate an inhalation expansion, and then co-registered to its original counterpart, taken as the reference image. In order to evaluate its clinical usefulness, we also apply this method to co-register two scans of the same subject, each one acquired on different dates. In a previous work, we established and tried this methodology in 2D (Osorio

et al, 2007). In this chapter, we extend it to 3D, including parameter optimization analysis (Osorio et al, 2010).

2. The image registration process

The aim of a registration procedure is to attain spatial alignment of one or more images – which we shall call "floating" (B), with respect to another one taken as the "reference" image (A). In a very general framework, given image A with coordinates $\mathbf{r_A}$, and image B with coordinates $\mathbf{r_B}$, we search the transformation:

$$\mathbf{T} : \mathbf{r}_B \rightarrow \mathbf{r}_A$$

such that: $$\mathbf{T}(\mathbf{r}_B) = \mathbf{r}_A \text{ and } max_T[S(A,B)]$$

where $S(A,B)$ is a *similarity measure* between A and B, usually represented by an objective function. For most cases, the similarity measure is critical for the success and general performance of the registration process. Its choice depends on the use of extrinsic or intrinsic image properties (e.g. fiducial markers or pixel intensity values), body region, modalities, matrix dimensions, etc. If the registration algorithm is iterative, another essential component is the *optimizer* for function S, which necessarily involves multiple evaluations in a multidimensional parameter space (Hill & Batchelor, 2001). The process is complete when a termination criterion is met.

Transformation $\mathbf{T}$ describes a spatial mapping from $\mathbf{r}_B$ to $\mathbf{r}_A$. In terms of elasticity, transformations may be rigid, affine, projective or curve (van den Elsen et al, 1993). The simplest transformation is the rigid one, requiring only 6 parameters in 3D (3 translations + 3 rotation angles). An affine transformation keeps parallelism between lines, introducing scaling and shearing (12 parameters in 3D). After a general projective transformation, line straightness is conserved, though parallelism is lost. Finally, a curved transformation produces a deformation that may be arbitrarily complex, though some constraints may apply to preserve smoothness and topology (e.g. each voxel should keep the same neighbours after applying the deforming transformation). As introduced in the previous section, the rigid model provides satisfactory results when the body region is limited to the head and some limb portions. To account for non-rigid displacements that may happen in the body, we adopt the scheme described in the next section.

3. Fourier-based registration: theoretical grounds

We have devised the following algorithm for non-rigid registration of image volumes A and B, each consisting of a stack of tomographic slices:

1. Pre-processing: segmentation, resampling, filtering.
2. An initial registration is carried out by applying an affine transformation (rigid+scaling), B ->A, optimizing the Normalized Mutual Information as a similarity measure (to be described later on).
3. B is divided into k sub-volumes B_r (k=8 in 3D).
4. An independent transformation is applied to each portion B_r , with a rigid and a non-rigid component.
5. The global transformation $T_{global}=S\ (T_1 \cup T_2 ... \cup T_k\ )$ is obtained after assembling the transformations T_r .

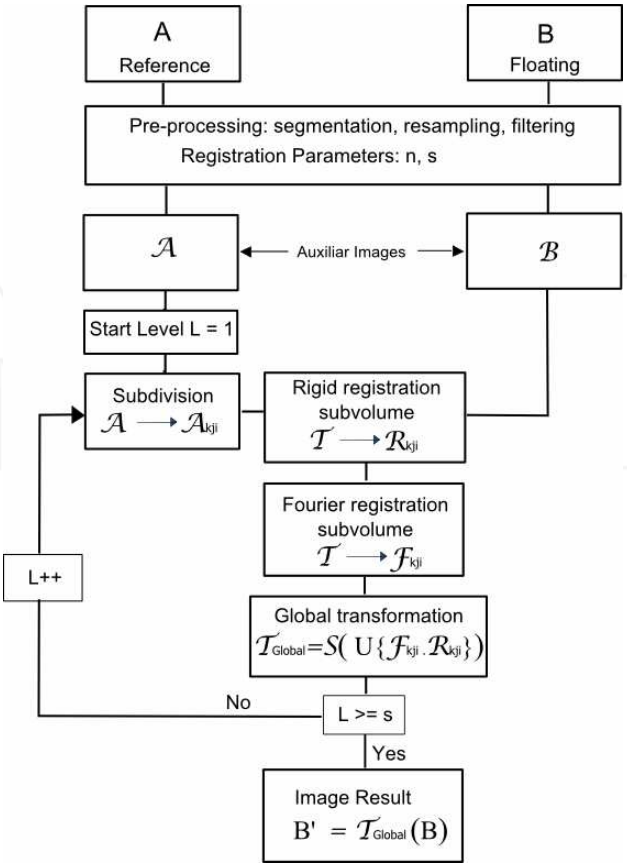


Fig. 1. Flowchart of the whole registration algorithm

Fig. 1 shows a diagram of this algorithm. After the initial affine approach in steps 2 and 4, the non-rigid stage is performed by a *n*-order Fourier expansion along each dimension:

$$T_x(\mathbf{r}) = x + \sum_{i,j,k=1}^{n} \left[a_{ijk}\varphi_{ijk}(\mathbf{r}) + b_{ijk}\psi_{ijk}(\mathbf{r}) \right] \quad (1)$$

$$T_y(\mathbf{r}) = y + \sum_{i,j,k=1}^{n} \left[c_{ijk}\varphi_{ijk}(\mathbf{r}) + d_{ijk}\psi_{ijk}(\mathbf{r}) \right] \quad (2)$$

$$T_z(\mathbf{r}) = z + \sum_{i,j,k=1}^{n} \left[e_{ijk}\varphi_{ijk}(\mathbf{r}) + f_{ijk}\psi_{ijk}(\mathbf{r}) \right] \quad (3)$$

where:

$$\varphi_{ijk}(\mathbf{r}) = \sin\left(\frac{\pi i x}{X}\right)\sin\left(\frac{\pi j y}{Y}\right)\sin\left(\frac{\pi k z}{Z}\right),$$

and:

$$\psi_{ijk}(\mathbf{r}) = \cos\left(\frac{\pi i x}{X}\right)\cos\left(\frac{\pi j y}{Y}\right)\cos\left(\frac{\pi k z}{Z}\right).$$

X,Y,Z are the image dimensions and *a, b, c, d, e, f* the coefficients to find up to order *n*. The initial condition is randomly chosen with a Gaussian distribution $N(0,\sigma^2)$.
With this approach, functions φ_{ijk} represent the transformation subset that keeps the volume boundary invariant, whereas functions ψ_{ijk} represent the transformations with null gradient on that boundary.

4. Subdivision scheme

Once the rigid transformations are applied, the independent subvolumes were assembled using quaternion interpolation (Walimbe et al, 2004). Since there is no standard interpolation method adopted for non-rigid transformations, in this work we propose the strategy shown on Fig. 2. It consists of a hierarchical scheme where each volume is divided into 8 equal subvolumes, this process being repeated *s* times. In what follows, we shall refer to the variable *s* as the subdivision number. In this way we build up a global transformation which is smooth, continuous and differentiable. A detailed description of this strategy is developed in the Appendix of Osorio et al (2010).

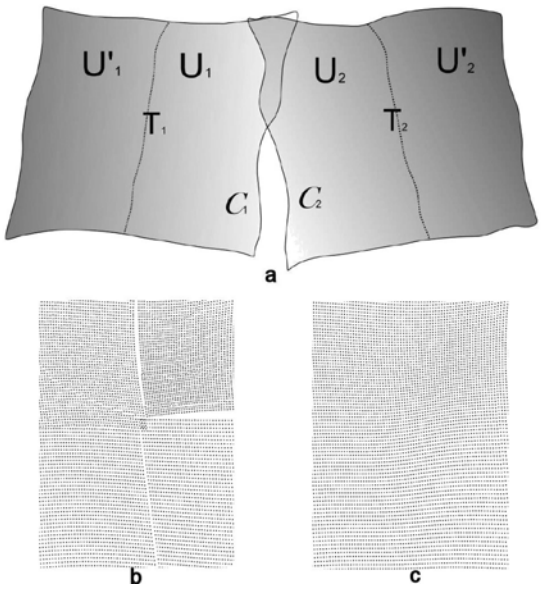


Fig. 2. a) Sub-volume assembling method. b) subvolumes before and c) after assembling. Figures show a 2D section of the 3D volume

5. Similarity measure and optimization

The metric adopted for this project is the Normalized Mutual Information, widely explored in the literature (Pluim et al, 2003), (Maes et al, 1997), (Studholme et al, 1999):

$$NMI = \frac{\sum_i p_i \log(p_i) + \sum_j p_j \log(p_j)}{\sum_{ij} p_{ij} \log(p_{ij})} \tag{4}$$

where p_i , p_j are the marginal probability distributions and p_{ij} is the joint distribution, using the partial volume method as interpolator (Chen et al, 2003).
The extended Downhill Simplex method was chosen as optimizer of the objective function (Press et al, 1992), (Zagrodsky et al, 2001). The program terminates when the change in the NMI is less than 10^{-4} between consecutive iterations.

6. Image data

Three tomographic modalities were studied: CT, MR and FDG[1]-PET of thorax and abdomen. For each one of those, we carried out a systematic analysis of the performance of intra-modality registration. CT studies were acquired with a HiSpeed scanner (GE, Milwaukee, USA) (matrix size: 512x512x47, voxel size: 0.7x0.7x7 mm^3). MRI studies were performed with a Signa Advantage 0.5 scanner (GE, Milwaukee, USA) (matrix size: 256x256x24) (voxel size 1.7x1.7x9 mm^3). The PET scanner used was a Quest 250 (UGM, Philadelphia, USA) (matrix size: 128x128x50, voxel size: 2x2x4 mm^3). All studies were completed at the Fundación Escuela de Medicina Nuclear (Mendoza, Argentina).
With the purpose of evaluating algorithm registration performance in a systematic way and for different initial conditions, we devised the following strategy. Each selected data set was slightly deformed using TPS (Thin-Plate Splines) (Bookstein, 1989)(Rohr et al, 2001) with a regular grid of 432 control points and average displacement of 27 mm (max. 40 mm), simulating a moderate thorax expansion during inhalation. This artificially deformed volume was selected as the floating image to be registered to the reference image (original, non-deformed version). This procedure was repeated for each modality, in order to find the parameters which maximize the NMI in a reasonable computing time.
Once the optimal Fourier order and subdivision number were found, we applied the algorithm to co-register pairs of thoraco-abdominal studies of a selected subject, which were acquired on different dates. In this way we had two data sets corresponding to the same body region, but displaying so evident anatomical discrepancies that would justify the application of a non-rigid model.

7. Error estimation

The error measure was chosen as the absolute average displacement over the whole image volume:

$$\varepsilon = \frac{1}{N} \sum_{i=1}^{N} \left| T(u_i) - \mathrm{v}_i \right| \tag{5}$$

where u_i and v_i are the voxel coordinates for the floating and reference images, respectively, and N is the total number of voxels.

[1] FDG: ^{18}F-labelled FluoroDeoxiGlucose

8. Registration of TPS-deformed image pairs

Fig. 3 shows the results corresponding to CT-CT registration after artificial deformation of the floating image. Shown are overlays of initial rigid registration and Fourier registration for transaxial slices. Coronal and sagittal overlays are shown in Fig. 4, which provide a visual assessment of our non-rigid approach. Bone structures remain unaltered during the process, which is consistent with a real situation.

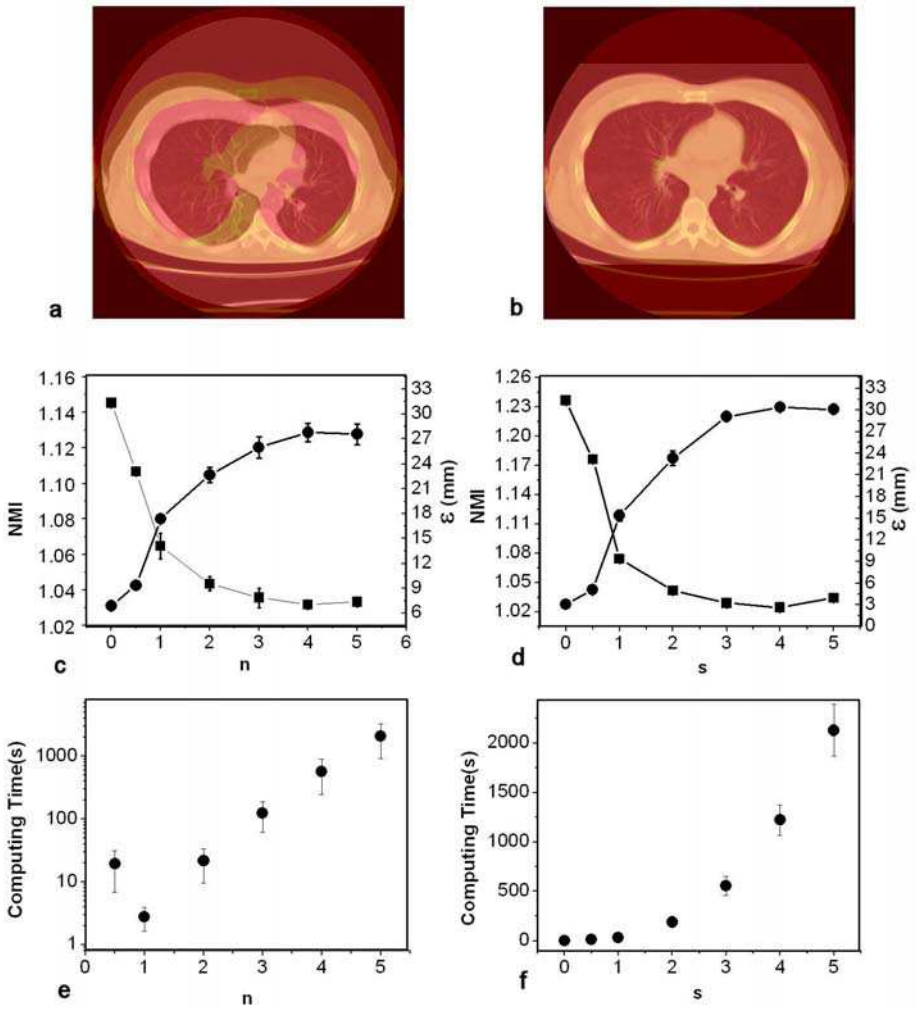


Fig. 3. CT vs. CT deformed with TPS (a) Rigid registration. (b) Fourier registration, n=2, s=4 (c) NMI (circles) and error ε (squares) vs. Fourier order for s=1. (d) NMI and error ε vs. number of subdivisions for n=2. (e) Computing time for Fig. 3c. (f) Computing time for Fig. 3d

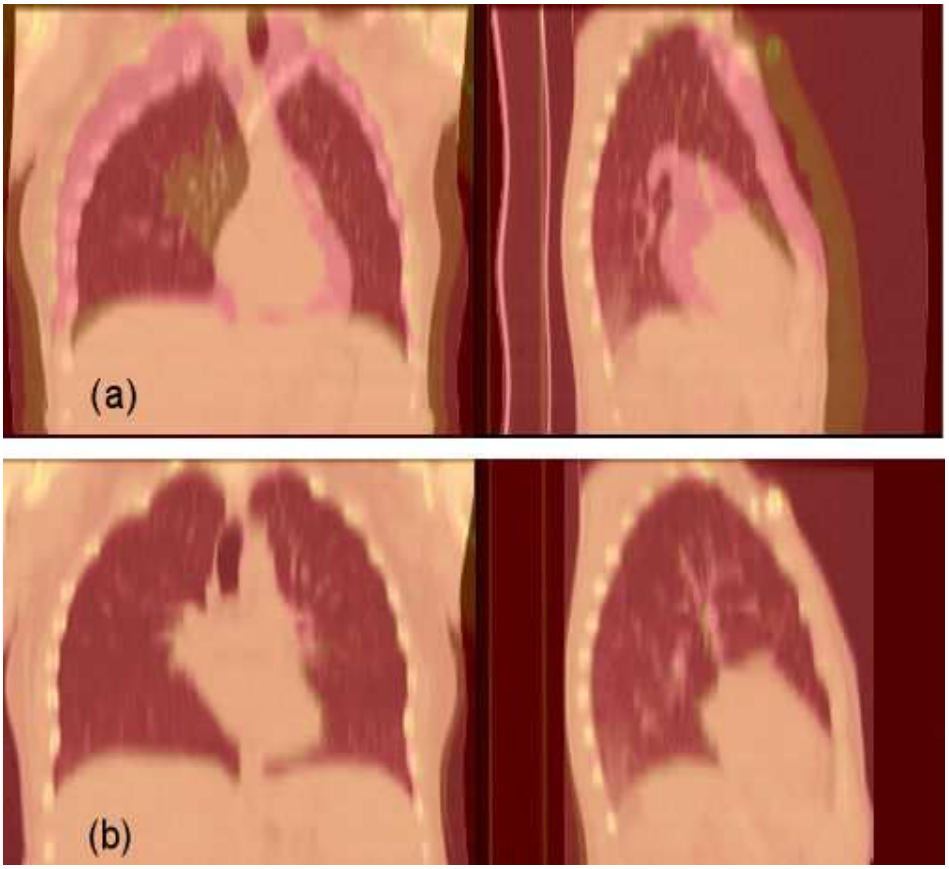


Fig. 4. Coronal and sagittal views of CT vs. TPS-deformed CT before (a) and after (b) Fourier registration

The graphs in Fig. 3 show the NMI, ε and computing time as a function of both the transformation order n and the subdivision number s. For all graphs, n,s=0 and n,s=0.5 refer to initial centre-of-mass alignment and rigid registration, respectively. Let us keep in mind that non-rigid registration comes into play when $n \geq 1$. As expected, the NMI increases and ε decreases at higher n. Looking for further optimization in the parameter space in a reasonable computing time, we set n=2 and plotted NMI vs. s. In this way, maximum NMI was attained at (n=2, s=4) for CT-CT registration. Each registration cycle was run 12 times, starting with different deformations, and the graphs show average values with their corresponding error bars. The computing time shows an exponential increase for n>1.

From the above results, we chose n=2 as the most appropriate transformation order, since at higher n values, error ε hardly decreases at the expense of long execution times. For the same reason, a suitable subdivision number was chosen at s=3,4. These results were also confirmed by visual inspection of the registered data sets.

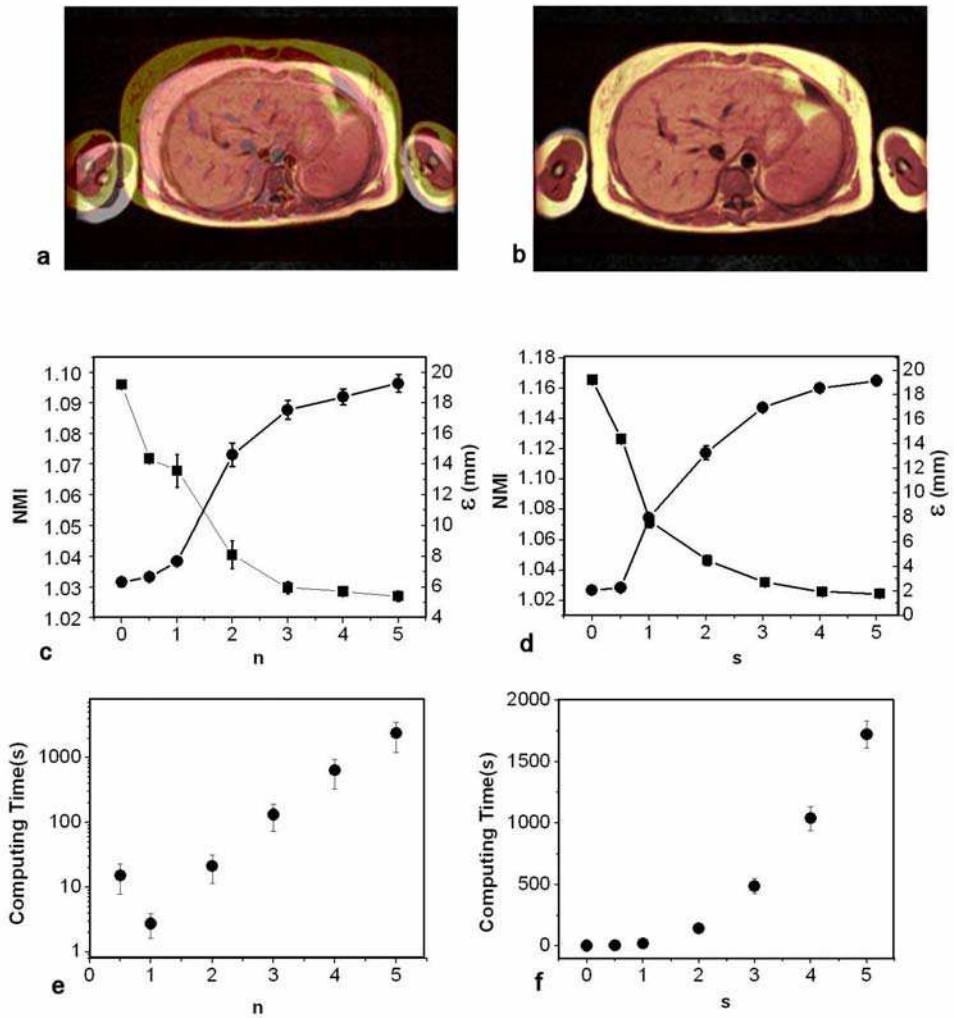


Fig. 5. MRI vs. MRI deformed with TPS (a) Rigid registration. (b) Fourier registration, n=2, s=4 (c) NMI (circles) and error ε (squares) vs. Fourier order for s=1. (d) NMI and error ε vs. number of subdivisions for n=2. (e) Computing time for Fig. 5c. (f) Computing time for Fig. 5d

Similar results were obtained for the MR-MR situation (Fig. 5). As for PET vs. PET, since it is a smaller dataset, no further improvement is obtained for s>3, when few voxels remain in each subvolume to calculate the NMI with enough statistical power.

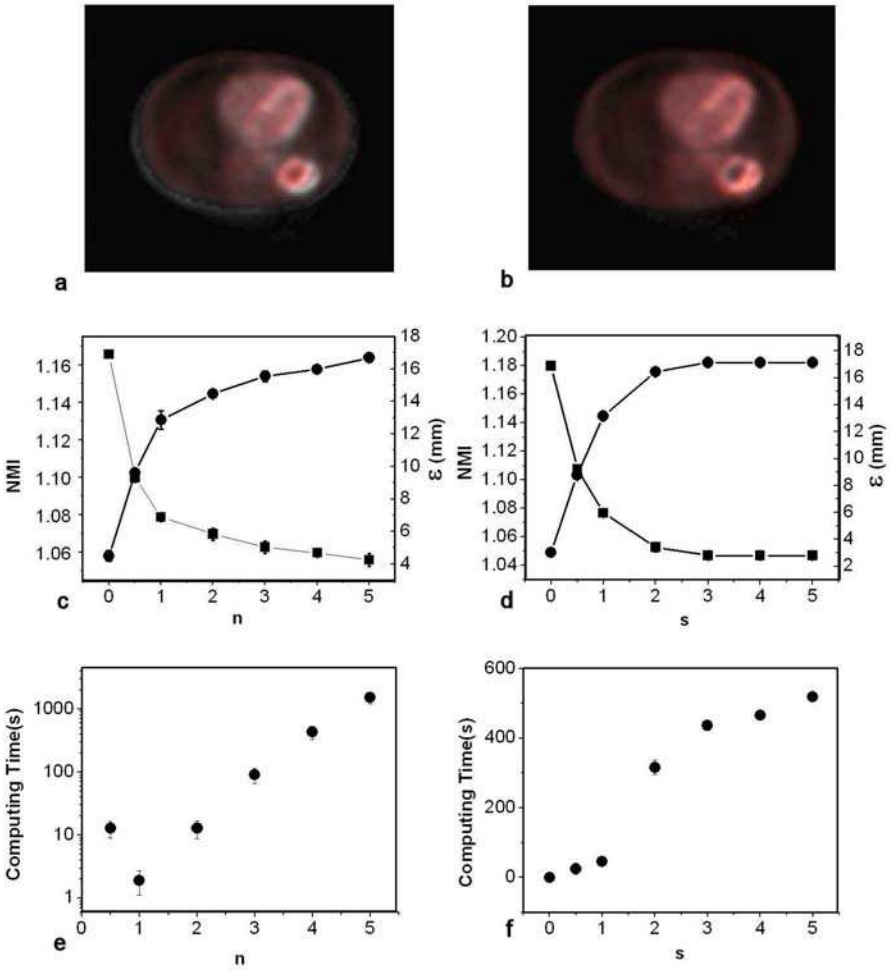


Fig. 6. PET vs. PET deformed with TPS (a) Rigid registration. (b) Fourier registration, n=2, s=4 (c) NMI (circles) and error ε (squares) vs. Fourier order for s=1. (d) NMI and error ε vs. number of subdivisions for n=2. (e) Computing time for Fig. 6c. (f) Computing time for Fig. 6d

9. Clinical studies

Once the optimum parameters were chosen from the previous analysis with TPS-deformed datasets, we selected three pairs of clinical studies to evaluate our method for intra-subject, intra-modality registration in real situations.

For each modality, the first study was taken as reference to which the second study was registered. Fig. 7 shows the results for all modalities, which were qualified by expert

radiologists as "acceptable" by visual inspection. The reader should keep in mind that in this case we lack from a "gold standard" for error calculation for obvious reasons, so this evaluation cannot be considered a validation test.

10. Error estimation with digital fiducial markers

Our choice for error estimation as described in Section 7, includes large portions of the field-of-view that are not relevant to registration process itself, such as the background that surrounds the body region of interest. To check if this has a significant impact on error calculation, 10 digital spheres were inserted in selected anatomical locations inside the CT, MR and PET volume data sets, so as to simulate easily identifiable internal fiducials. Next we applied a TPS deformation on such "marked" volumes and carry out the registration cycle as described in Section 8.

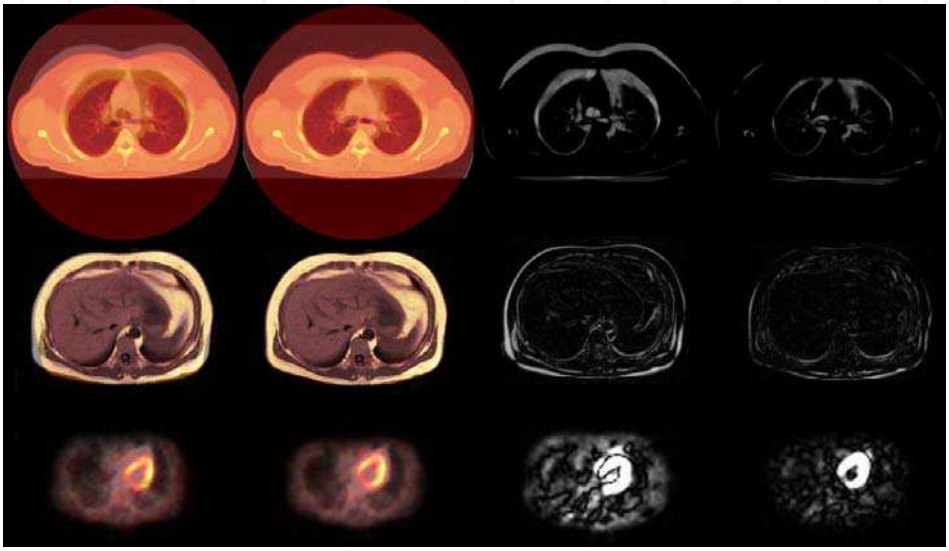


Fig. 7. Intra-modality registration of two clinical scans (same subject, different sessions). From left to right: Rigid registration, Fourier registration, Image difference after rigid registration, Image difference after Fourier registration. From top to bottom: CT, MRI and PET

After registration of both data sets, we measured the distance between center-of-masses for homologous spheres. Finally, the Mean Root Square Error (MRSE) was calculated using Eq. 5. In this case the summation is limited to the available sphere centers. In order to control for the effect of the additional deformation induced by the adjustments of the spheres on the registration, two different sphere sizes were studied: big (8 mm ϕ) and small (2 mm ϕ). The results are summarized in Table I, showing that the effect of the sphere size on the global error is quite small. This means that the spheres themselves are not affecting the registration process in a significative fashion.

We can see from Table 1 that the typical errors are between 1.5 mm and 2.5 mm, which are slightly smaller than the average displacements calculated for the whole volumes, as shown

in Figs. 3d, 5d and 6d. These results are of the same order as the ones found in recent approaches (Andronache et al, 2008), (Sohn et al, 2008) , using more complex methods such as a combination of cross-correlation and mutual information in the former and local matching of anatomical features in the latter. The computing time (as seen in Figs, 3, 5 and 6) is between 400 secs and 1000 secs, which is also of the same order as the time reported in the former reference.

	2 mm ϕ Sphere	8 mm ϕ Sphere	Error whole image
MRSE (CT)	2.55 mm	2.44 mm	2.62 mm
MRSE (MRI)	1.29 mm	1.55 mm	2.99 mm
MRSE (PET)	1.98 mm	1.58 mm	2.80 mm

Table 1. Mean Root Square Error (MRSE) for the three modalities (CT, MRI, PET). The errors in the second and third columns are evaluated after the registration with images that includes a set of 6 to 10 spheres. Sphere diameters are 2 mm and 8 mm. The error in the last column was evaluated over the whole image. The parameters of the algorithm were n=2, s=4 for CT and n=2, s=3 for MRI and PET

11. Discussion and conclusions

Our systematic analysis shows how the similarity measure (NMI) behaves with both the order of the transformation n and the number of subdivisions s. For intra-modality registration, and for the three modalities studied, we found that maximum NMI is attained for $n \geq 3$ (s=1). For n=2 the NMI is only slightly inferior, however the computing time becomes an order of magnitude longer if calculation is performed up to n = 3 (Figs. 3e, 4e, 5e), as the number of Fourier parameters to be optimized is $6n^3$ in 3D (Eqs. 1-3).

As expected, the similarity measure increases with the number of subdivisions s, accompanied by a decrease in error ε (Figs. 3d-5d). The improvement in registration quality was also confirmed by visual assessment for both CT-CT and MRI-MRI by comparing results using s=3 versus s=4. For PET-PET, NMI and ε does not get better for s>3, because further subdivision results in sub-volumes with too sparse data for that modality. Only slight improvements in the NMI and ε were recorded for n>3 and s>3, but at a very high computational cost and providing negligible visual improvements. One issue of concern is that rigid structures such as bone in CT should remain so after registration. Since the characteristic size of the deformation applied is greater than typical bone structures, they do not deform noticeably (Fig. 4).

Regarding calculation time, by setting optimal parameters (n=2, s=3), and using an ordinary computer[2], the time for co-registering two CT volumes (matrix dimensions 512×512×47) is ~ 600 secs. (~ 200 secs. if s=2). For MRI-MRI, somewhat shorter times were measured, whereas for PET-PET, execution took slightly over 400 secs. (Figs. 3f,5f,6f). Such computation times were attained without any specific optimization technique. Let us note that the algorithm leads naturally to parallelization because the subvolumes can be processed independently. In that way, the computing time can be substantially reduced.

In principle, the use of Fourier basis functions allows arbitrary deformations on any given image volume; the combination of this method with a subdivision scheme allows to

[2] Core 2 Duo, 3GHz, no multithreading (2008).

accommodate small image portions in a progressive fashion, without affecting the rest of the image data (Walimbe et al, 2004), (Likar & Pernus, 2001). Obviously, as the subdivision creates small subvolumes with fewer and fewer pixels, similarity measures like the NMI or the CCC become affected in their performance (Andronache et al, 2008).

In general, the proposed registration method rendered acceptable results for small and moderate deformations (~ 25mm). A preliminary study suggests that it is fairly robust, even in the presence of Gaussian noise (Osorio et al, 2007). We evaluated its performance using clinical images after deformation with Thin-Plate Splines, as well as image pairs corresponding to different scan sessions for a same subject. The selected studies were thoracic and abdominal scans for three common tomographic modalities. Obviously, not only organ deformations and displacements may come about between scan sessions, but also significant variations in anatomy and function, due to normal or pathological conditions. In these cases, the outcome of any non-rigid registration method offers an approximation whose usefulness must be assessed for each particular situation.

12. References

Andronache A.; von Siebenthal M.; Székely G. & Cattin P. (2008). Non-rigid registration of multi-modal images using both mutual information and cross-correlation. *Med. Imag. Anal.* 12, pp. 3-15.

Ashburner J. & Friston K. (1999). Nonlinear spatial normalization using basis functions. *Hum Brain Mapp*. Vol. 7, June 1999, pp. 254–266.

Bookstein F. (1989). Principal warps: Thin-plate splines and decomposition of deformations. *IEEE transactions on Pattern Analysis and Machine Intelligence,* 11:567-585.

Chen M. & Varshney P. K. (2003). Mutual information-based CT-MR brain image registration using generalized partial volume joint histogram estimation. *IEEE Trans. Med. Imaging* 22, pp. 1111-1119.

Crum W.; Hartkens T. & Hill D. (2004). Non-rigid image registration: theory and practice. *Br J Radiol*. Vol. 77: S140–S153.

D'Agostino E.; Maes F.; Vandermeulen D. & Suetens P. (2003). A viscous fluid model for multimodal non-rigid registration using mutual information. *Med Image Anal*. Vol. 7, December 2003, pp. 565-575.

Friston K.; Ashburner J.; Kiebel S.; Nichols T. & Penny W. (2006). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. ISBN-10: 0123725607 Academic Press, London.

Hill D. & Batchelor P. (2001). Registration Methodology: Concepts and Algorithms. En: Medical Image Registration (eds: Hajnal J.V. et al), CRC Press.

Likar B. & Pernus F. (2001). A Hierarchical Approach to Elastic Registration Based on Mutual Information. *Image Vis. Comput.* 19, pp. 33-44.

Maes F.; Collignon A.; Vandermeulen D.; Marchal G. & Suetens P. (1997). Multimodality image registration by maximization of mutual information. IEEE *Trans. Med. Imaging* 16, pp. 187-198.

Maintz J. & Viergever M. (1998). A survey of medical image registration. *Med Image Anal*. Vol. 2, March 1998, pp. 1–36.

Osorio A.; Isoardi, R. & Mato G. (2007). Non-rigid registration of tomographic images using Fourier transforms. *Journal of Physics: Conference Series. Vol.* 90, Issue 1, 2007, 012058

Osorio A.; Isoardi, R. & Mato G. (2010). Deformable CT registration using Fourier basis functions in 3D. In: *XXIII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI '2010), Gramado, IEEE Computer Society Press* (in press, September 2010).

Pluim J. P. W.; Maintz J. B. A. & Viergever M. A. (2003). Mutual-Information-Based Registration of Medical Images: A Survey. *IEEE Trans. Med. Imaging* 22, pp. 986-1004 .

Press W. H.; Flannery B. P.; Teukolsky S. A. & Vetterling W. T. (1992). Numerical Recipes in C: The Art of Scientific Computing, 2nd ed. (Cambridge University Press, Cambridge).

Rohr K.; Stiehl H. S.; Buzug T. M.; Weese J. & Kuhn M. H. (2001). Landmark-Based Elastic Registration Using Approximating Thin-Plate Splines. *IEEE Trans. Med. Imaging* 20, pp. 526-534.

Rueckert D. & Aljabar P. (2010). Non-rigid registration of medical images: theory, methods and applications. *IEEE Signal Proc. Mag*. Vol. 27, July 2010, pp. 113-119.

Sohn M.; Birkner M.; Chi Y.; Wang J.; Yan D.; Berger B. & Alber M. (2008). Model-independent, multimodality deformable image registration by local matching of anatomical features and minimization of elastic energy. *Med. Phys.* 35, pp. 866-878.

Studholme C.; Hill D. L. G. & Hawkes D. J. (1999). An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognit.* 32, pp. 71-86.

Townsend D. (2008). Multimodality imaging of structure and function, *Phys. Med. Biol.*, Vol. 53, February 2008: R1-R39.

van den Elsen P.; Pol E-J. D. & Viergever M. (1993). Medical image matching-a review with classification. *IEEE Eng. Med. Biol. Mag*. Vol. 12, March 1993, pp. 26-39.

Walimbe V. S.; Zagrodsky V.; Raja S.; Bybel B.; Kanvinde M. & Shekhar R. (2004). Elastic registration of three-dimensional whole body CT and PET images by quaternion-based interpolation of multiple piecewise linear rigid-body registrations, *Proceedings of the SPIE* (SPIE, San Diego, CA, 2004), pp. 1191–1228.

Zagrodsky V.; Shekhar R. & Cornhill J. F. (2001). Multi-function extension of simplex optimisation method for mutual information based registration of ultrasound volumes. *Proc. SPIE* 4322, pp. 508-515.

7

Fourier Transform Infrared Microspectroscopy for Cancer Diagnostic of C6 Glioma on Animal Model

Beljebbar A. and Manfait M.
Unité MéDIAN, MEDyC, CNRS UMR 6237, IFR 53, UFR de Pharmacie, Université de Reims Champagne-Ardenne
France

1. Introduction

Cancer is one of the leading causes of death in the world. Currently, the gold standard in most cancer diagnosis is histopathological evaluation, which involves the removal of tissue biopsies and examination by pathologists. This process includes tissue staining and morphological pattern recognition. During tissue transformation, it is expected that substantial modifications occur at molecular level before visible morphological changes become apparent. Histological examination requires extensive human observations to recognize both the constitutive histologic entities and the pathologic state. Early detection of cancer is the most important factor in the prevention of cancer and a guarantee in most cases of an effective treatment and in some cases for a complete cure. Moreover, the histological assessment must be performed while surgery is ongoing to determine whether the tissue has been removed or spared. For example, brain tissue cannot be removed with a large safety distance from the tumor. The resection of brain tumors is strongly limited to the border of the tumor. The main objective is a detection of the source of the pathological variation at molecular level in order to further understand the molecular carcinogenic process in a range of cancers. Indeed, tumor tissues are mostly heterogeneous in nature, and this heterogeneity further depends on the stage of disease and its aggressiveness. The emergence of a novel technique, complementary to histopathology and immunohistochemistry, can thus help in the early diagnostic of tissue transformation during carcinogenesis.

Fourier-transform infrared microspectroscopy (FTIRM) has emerged as a powerful tool to study molecular structure and structural interactions in biological systems. When this technique is applied to tissues, the resulting spectra is composed of characteristic absorption bands originating from all infrared-active vibrational modes of biological macromolecules present in the tissue, such as proteins, lipids, and nucleic acids (Parker, 1971). Each of these molecules provides a unique absorption spectral pattern named fingerprint through the entire infrared spectrum. This property offers a way to identify the molecule type (qualitative analysis) and the amount or quantity of this molecule in the sample (quantitative analysis) (beljebbar et al., 2008). This method can be used as a diagnostic tool, complementary to histopathology or immunochemistry (Fernandez et al., 2005). As the

image contrast is based on the intrinsic vibrational signature of the tissue components, spectral images does not require the use of added dyes or labelling methods for visualization of different chemical components in the sample (Bates, 1976). Indeed, FT-IRM imaging combined a high spatially resolved morphological and biochemical information that offer a number of advantages for ex-vivo assessment of tissue and aid the histopathologist in the identification and classification of subtle biochemical changes related to carcinogenesis (Petibois & Déléris, 2006; Cohenford & Rigas, 1998; Kneipp et al., 2000; Yano et al., 2000). With the fast image acquisition provided by modern mid-infrared imaging systems, it is now envisaged to analyze tumor biopsies in delays compatible with surgery (Levin & Bhargava, 2005). Other advantages of this method are that it is objective and provides reproducible diagnosis, minimize inter-observer variability. Indeed, IR spectroscopy can detect and monitor characteristic changes in molecular composition and structure that accompany transformation from normal to cancerous state (Afanasyeva et al., 1998, Diem et al., 1999, Franck et al., 1998). The identification and quantification of these specific molecular changes within tissues can provide diagnostic information for aiding in early detection of diseases and their optimized treatment. Correlations of morphologic and biochemical tissue differences could be used to identify variations that occur between healthy and diseased tissues. The development of clinical protocols for the routine examination of tissue histology or of localized tumors using IR microspectroscopic methods has been largely used in medical diagnostics to identify neoplasia in breast, (Ci et al, 1999) cervix, (Wong et al, 1991) colon, (Rigas et al, 1990) lung, (Yano et al, 2000) stomach, (Li et al., 2005) and glioma. (Krafft , 2006, 2007; Amharref et al, 2006; Beljebbar et al., 2008).

Infrared spectra contain many overlapping bands and so data interpretation cannot be made by simple visual inspection and alternative approaches are needed. Because of the high complexity of the FTIR spectra obtained from tissues, multivariate statistical methods are required to extract biochemical information related to tissue. This would permit to objectively differentiate distinct tissue structures and for identifying origin that gave rise to the specific tissue pathology. These methods have had a major impact on the quantitative and qualitative analysis of infrared spectral data. They have been shown to improve analysis precision, accuracy, reliability, and applicability for infrared spectral analyses relative to the more conventional univariate methods of data analysis. Rather than attempting to find and use only an isolated spectral feature in the analysis of spectral data, multivariate methods derive their power from the simultaneous use of multiple intensities (i.e. multiple variables) in each spectrum (Mourant, et al., 2003). During the last decade, it has been recognized that FT-IR, in combination with the appropriate multivariate analysis strategies, has considerable potential as a metabolic fingerprinting tool for the rapid detection and diagnosis of disease or dysfunction (Goodacre et al, 2004; Diem et al., 1999). Multivariate imaging techniques including Unsupervised Hierarchical Cluster Analysis (UHCA) (Jackson et al., 1998; Mohlenhoff et al., 2005), K-means clustering (Lasch et al., 2004; Zhang et al., 2003), Principal Components Analysis (PCA) (Lasch and Naumann, 1998), Linear Discriminant Analysis (LDA) (Mansfield et al, 1999), Fuzzy C-means clustering (Lasch et al., 2004; Mansfield et al., 1997) and neural networks (Lasch & Naumann, 1998) have proven to be invaluable in the identification of spectral groups or "clusters" which can be directly compared to stained tissue sections. In multivariate methods, the information of the entire spectrum can be utilized for the analysis. The high correlation of spectral clusters with anatomical and histopathological features has been conclusively demonstrated for a number of different tissue types including.

The aim of this chapter was the monitoring and interpretation of molecular changes associated to C6 glioma growth and invasion by micro-FTIR imaging. Micro-FTIR maps were recorded on normal brain tissue and on glioma growth after injection of C6 glioma cell suspension in brain parenchyma. Multivariate statistical analysis were used to i) identify the molecular changes associated with the development of the glioma tumor ii) definition of the tumor and peritumoral margins, iii) grading of malignancy and prognosis based on the presence of necrosis. We have investigated the spatial distribution of molecular changes associated with C6 glioma progression using integrated intensity ratios of some specific bands associated to lipids and proteins in order to determine spectroscopic markers to successfully monitor the changes in the molecular composition associated to C6 glioma progression.

2. FTIR characterization of normal brain tissues and C6 glioma progression using cluster analysis

To investigate the potential of FTIR spectroscopy for clinical application, experimental animals are necessary. Rat C6 glioma cells are an experimental cell line that when injected into neonatal Wistar rats grow into an intracerebral tumor with pathological similarities to human glioblastoma (GBM) (Auer et al., 1981). This glioblastoma model was used in a variety of studies related to brain tumor biology including tumor growth (Nagano et al., 1993; San-Galli et al., 1989), invasion (Nagano et al., 1993; Bernstein et al., 1991; Chicoine & Silbergeld, 1995), and evaluation of the therapeutic efficacy of cancer treatments (Barth 1998). The glioma tumors were obtained by injection of C6 glioma cells suspension in brain parenchyma as described elsewhere (Grobben et al., 2002). All animals with implanted C6 cells developed tumors with reproducible localization and size around the site of injection. These groups were sacrificed after 5, 7, 9, 12, 15, 19, days post-implantation (PI). After brain excision, tissue samples were snap-frozen by immersion in methyl-butane cooled down in liquid nitrogen and stored at - 80°C. Two adjacent sections were cut from each sample using a cryomicrotome. One section, 10 µm thick, was placed onto infrared transparent calcium fluoride (CaF_2) slides for infrared imaging. The second section, 7 µm thick, was placed on a microscope glass slide and stained with hematoxylin and eosin (H&E) for histopathological image. Spectra were collected using an FTIR imaging system (SPOTLIGHT, Perkin-Elmer, France) coupled to a FTIR spectrometer (Spectrum 300, Perkin-Elmer, France). This system is equipped with a liquid N_2 cooled Mercury-Cadmium-Telluride MCT line detector comprised of 16 pixel elements. The microscope was equipped with a movable, software-controlled x, y stage. In this study, FTIR images were collected from selected sites with a spatial resolution of 25 µm/pixel, in transmission mode, in the 4000–720 cm^{-1} range, with a final spectral resolution of 4 cm^{-1}, and 16 scans per pixel. After atmospheric correction, data were cut to high wavenumber fingerprint region (2600 to 3700 cm^{-1}), converted to their first derivative, and smoothed using a seven point Savitzky-Golay algorithm in order to minimise the influence of background scatter in the spectra (Savitzky & Golay, 1964). The resulting spectra were then normalized using a Standard Normal Variate (SNV) procedure. A multivariate statistical analysis (Principal Component Analysis (PCA) and K-Means (KM)) was performed on this dataset. K-means clustering was performed on these principal component scores. Pseudo-color maps based on cluster analysis were then constructed by assigning a color to each spectral cluster. The cluster spectra were calculated by averaging absorbance spectra associated to each group and used for the interpretation of the chemical

or biochemical differences between clusters. All data measured on normal brain and tumor development (from 7 to 21 days growth) were pooled in one dataset, processed at the same time and the results were displayed as pseudo-color maps with the same color scale. In this way, we can easily determine all their common and discriminating features by comparing their infrared maps. The rat brains analysed before 5 days after tumor implantation did not show any visible changes in the brain tissue (data not shown). Indeed, all animals died beyond 25 days post-implantation and therefore no data are available beyond this time point.

Histological and FTIR analysis revealed that the tumor size dependent on the period elapsed after injection of glioma cells. Fig. 1 displays FTIR pseudo-color maps of brain tissue and their histopathological images. 13 clusters describing both normal brain and cancer features were extracted and pseudo FTIR maps were constructed with the same color scale. Different clusters in the FTIR images were correlated with features of the histopathological images. White color represents the area where no tissue was present. In the pseudo-color map obtained from glioma tumor obtained 7 days after C6 cells injection (fig. 1B). This pseudo-color map displays some normal structures associated to white matter and grey matter. Cluster 4 described white matter corresponding to corpus callosum (CC) and commissura anterior (CA). CC and CA present important lipid content due to high myelin level in these structures involved in communication within and between hemispheres. Several clusters (6, 1, 2, 8, and 11) describe the transition from white matter to gray matter (cortex). In fact, the CC is the largest white matter structure in the brain, consisting mainly of interhemispheric fibers. Gray matter is distributed at the surface of the cerebral hemispheres (cerebral cortex) and of the cerebellum (cerebellar cortex). It is predominantly composed of neuron cell bodies and unmyelinated axons. Comparison between pseudo-color maps and the histopathology images (Fig. 1A and 1B) shows that the FT-IRM image provides more information on the cortex than standard histopathology. In fact, six layers were identified from the cortex in the pseudocolor FTIR map (fig. 1B), whereas, H&E staining did not allowed to discriminate between these layers in the cortex (fig. 2A). Luxol fast blue (LFB) staining was then used to visualize myelin distribution into brain tissues and to map particular sections within the cortex. In fig. 2B, LFB staining shows a gradation color density between brain structures. In stained preparations, myelin is intensely blue, so that white matter is well differentiated from gray matter. In fact, large fiber tracts like CC and some bundles in CP can be easily recognized. However, even with higher magnification it was very difficult to distinguish between all cortex layers because this staining is mainly restricted to fibers.

Thus, the individual cell type cannot be recognized within the tissue. This LFB staining was then combined with cresyl violet (CV) coloration to stain not only rough endoplasmic reticulum (Nissl substance) but chromatin and nucleoli as well. This coloration was used in our study to visualize all different cell types present in the cortex (Fig. 2C). This pseudo-color FTIR map (fig. 2D) was correlated with that obtained with LFB-CV staining (fig. 2C). With this coloration, six different layers were then identified in the cortex instead of five layers in the FT-IRM map (fig. 2D). On the other hand, when multivariate statistical analysis was applied only on FT-IRM data measured from normal brain tissue we were able to distinguish between five cortex layers. The cortex consists of a thin layer of gray substance which covers the two hemispheres and whose thickness varies between 2 and 4 mm. According to the cortical histological organization, one distinguishes 6 layers numbered from I to VI from top to bottom (Burwell, 2001). The outermost molecular layer (I)

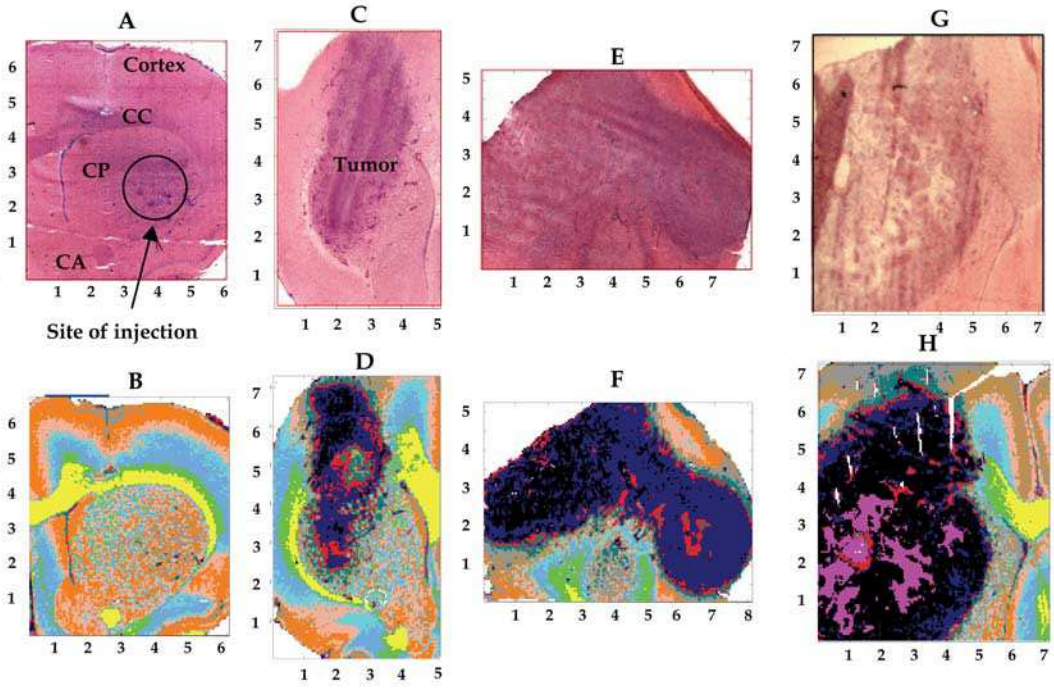


Fig. 1. Photomicrographs (H&E staining) associated to glioma progression (A, C, E, G, I) from respec
days post-implantation. All data measured on several brain tissues during tumor evolution were po
at the same time to extract all features describing both normal and tumor tissues. Data were cut into
3700 cm^{-1}) and cluster analysis was carried out on the first derivative spectra (to enhance the resolut
K-means was calculated the cluster-membership of spectra by assigning each color to one class. Pseu
corresponding histological images were constructed on tumor progression (maps B, D, F, H, and J) r

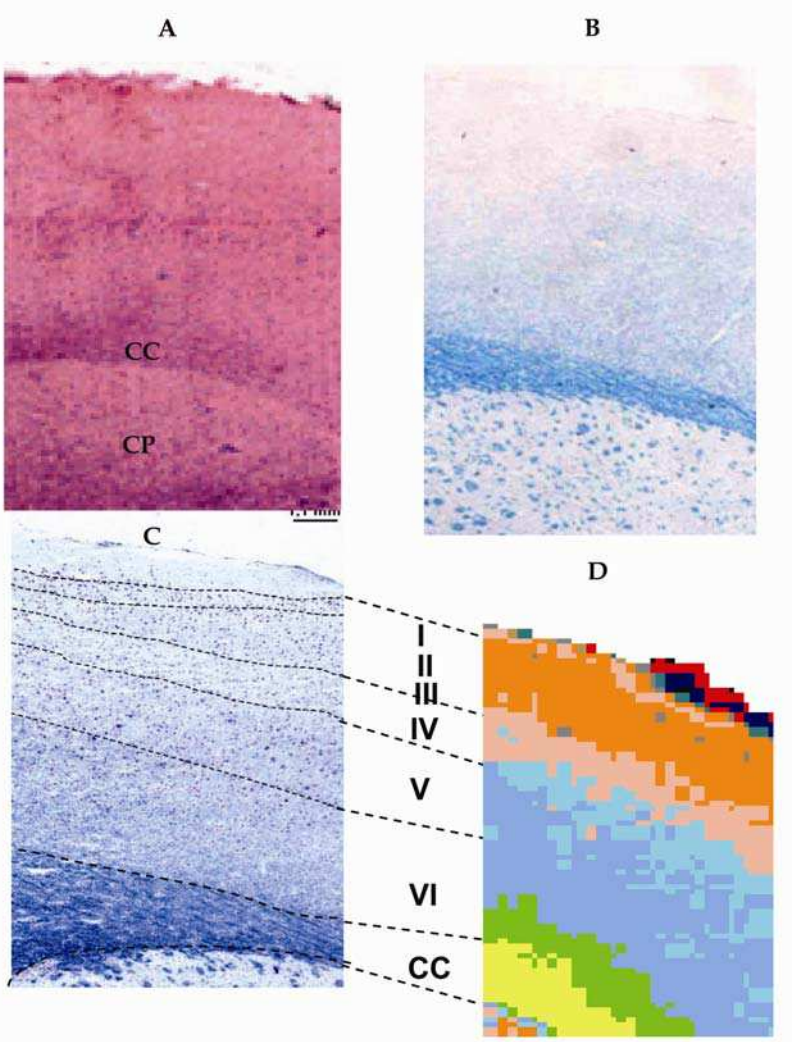


Fig. 2. Photomicrography of a brain tissue section (A) H & E staining, (B) stained with LFB; in this image CC and CP appears more lightly stained, (C) stained with LFB-CV, pointing out the cortical layers of the rat cortex. Pseudo-color FTIR map (D) based on K-means cluster analysis applied only on FTIR data measured from normal brain structures

containing non-specific fibers, corresponds to the yellow cluster. The external granular layer (II) is a rather dense layer composed of small cells. The external pyramidal layer (III) contains pyramidal cells, frequently in row formation. These three layers were encoded by cluster 11. The internal granular layer (IV) is usually a thin layer with cells similar to those in the external granular layer (cluster 8). The ganglionic layer (V) contains, in most areas, pyramidal cells that were fewer in number but larger in size than those in the external pyramidal layer (clusters 1 and 2). The fusiform layer (VI) consists of irregular fusiform cells

whose axons enter the adjacent white matter known as the CC (cluster 1 and 6). By correlating FT-IRM spectral maps with histopathology (H&E, LFB, and LFB-CV) of the adjacent tissue sections, we highlighted the potential of FT-IRM to identify the morphologic origin that gave rise to the specific spectral features found in this study. In fact, with standard staining (H&E), we were not able to discriminate between the different cortex layers. On the other hand, FT-IRM pseudo-color maps were clearly similar to LFB-CV staining for visualizing myelin distribution in healthy brain tissues (white and gray matters).

Fig. 1B exhibit a particular structure (cluster 7) located in the caudate putamen (CP) corresponding to C6 cells injection site. This feature was associated to C6 cell growth. Histological interpretation however showed small abnormalities in rats sacrificed seven days after implantation of C6 cells (fig. 1A). Indeed, this abnormality is characterized by a diffuse structure within the brain parenchyma. The tumor cells appear scattered in the form of cells or grouped in clusters in the vicinity of the blood vessels. At day 9, the viable tumor started to be visible in the implantation site, destroys the corpus callosum and grew into all the cortical layers (fig. 1D). The tumor area was encoded by clusters 12 and 13. On the other hand, clusters 5 and 7 were detected around the tumor. These clusters were associated to the proliferative and invasive character of glioma tumors. In fact, in our previous study, immunohistochemical Ki-67 and MT1-MMP staining were used to visualize the proliferative and invasive activities of glioma and were clearly correlated with the cluster that encoded the surrounding tumor area *(*Amharref et al., 2007). Histopathological staining confirms this FTIR result (fig. 1C), In fact, a Large focus of invasion (6mm) were well separated from the surrounding brain tissue. Tumor was hypercellular with cellular and nuclear pleomorphism and mitotic figures observed. From day 12, the tumor is fairly large and deeply situated within the cortex with massive infiltration into the brain tissue (fig. 1F, 1H, 1J). Most of cerebral cortex is destroyed by tumor tissue. At day 15, we observe the appearance of clusters 9 and 5 associated to the formation of necrosis and perinecrosis (fig. 1H). Histological image displayed oedematous zones as well as zones of necrosis with a pseudopalissading cells (> 8mm) (fig. 1G). Around necrotic zones, tumor shows a large cell density with an increase of the proliferation of endothelial cells and haemorrhagic zones. This necrotic zone increased until day 19 post-implantation (fig. 1J). Cluster 5 observed in the border of the necrotic zone seems to correlate with the pseudopalisading formation. Cluster 9 correlates to the center of the necrosis (full necrosis) of the tumor. The presence of necrosis is important for grading tumors and is often linked to a poorer clinical prognosis (Barker et al., 1996). Indeed, the most characteristic finding of glioblastoma is the necrotic foci surrounded by tumor cells (Kleihues et al., 1993). Pallisading cells delineate the foci of necrosis and lymphocytic infiltration, with the occasional formation of edema fluid (Auer et al., 1981). At day 19, we observed the transformation of the structure of the tumor (fig. 1J). In fact, cluster 12 was replaced partially by cluster 13. On the other hand, the tumor occupied almost the entire hemisphere as visualized on the tissue of section (fig. 1I).

Fig. 3 shows class average spectra associated to normal brain structures and tumor development. The lipids spectra contain a large proportion of methyl, methylene and carbonyl bands in the region 2600-3700 cm^{-1}. The bands at 2852 cm^{-1} and 2924 cm^{-1} were due to the symmetric CH_2 stretching mode of the membrane lipid which is directly related to the lipid acyl, primarily saturated, chains. The band at 2956 cm^{-1} is associated to asymmetric CH_3 stretching mode of the methyl group and less contribution of proteins. The =C–H stretching bands due to unsaturated acyl chains are found at 3014 cm^{-1}. The bands at 3292 cm^{-1} and 3080 cm^{-1} were linked to the Amide A (mainly N–H stretching) and Amide B (N–H

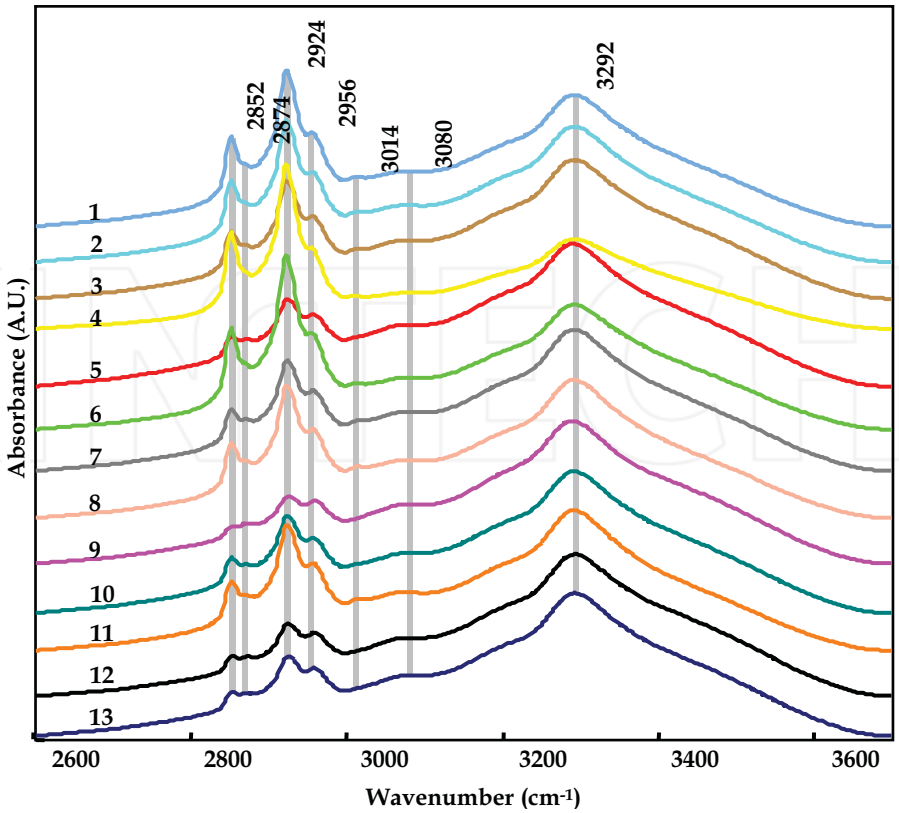


Fig. 3. Representative cluster mean FTIR spectra extracted from pseudocolor maps. Cluster averaged spectra were obtained by meaning absorbance spectra associated to each group. 13 models describing normal and glioma brain development. Each cluster averaged spectrum assigned to one class was plotted with the same color than in pseudo-color map

stretching) of proteins respectively. The band at 2874 cm^{-1} was related mainly to the symmetric stretch of proteins. The changes in frequencies, intensities and band shapes of these bands may provide further information about the structural changes associated with malignancy. In the malignant tissues, i) the intensity of the symmetric CH_2 band decreased compared to the corresponding bands in normal brain structures (white matter and grey matter) and ii) the band ratio 3292/2852 cm^{-1} was higher in the glioma tumor and reduced in the normal brain tissues. Indeed, the intensity ratio 2924/2956 cm^{-1} was higher in the spectra associated to white matter and decreased from grey matter to C6 glioma tumors. Band 2874 cm^{-1} became better resolved in the spectra of malignant tissues because of diminished intensities of CH_2 bands in this region. Therefore, the largest variances from spectra to spectra in IR spectroscopic maps of normal tissue were assigned to spectral contributions of lipids reflecting cell differentiation. This result confirms that the development of tumor was characterized by a reduction in total lipid content. This reduction was also observed in the invasive area (clusters 5 and 7), which is composed of healthy and tumoral cells. This result is in agreement with those obtained in brain diseases (Krafft et al., 2006; Kneipp et al., 2000).

This lipid reduction in malignant tissues could be related to the fast growth of tumor cells which need more energy (Wang et al., 2003). Indeed, it is known that, in developing brain tumors, structural and functional cell changes take place in which lipids play a crucial role. Yet, qualitative and quantitative aspects of lipid changes in brain tumors of different degree of malignancy are still the subject of numerous studies (Steiner et al., 2003; Krafft et al., 2006; Campanella, 1992).

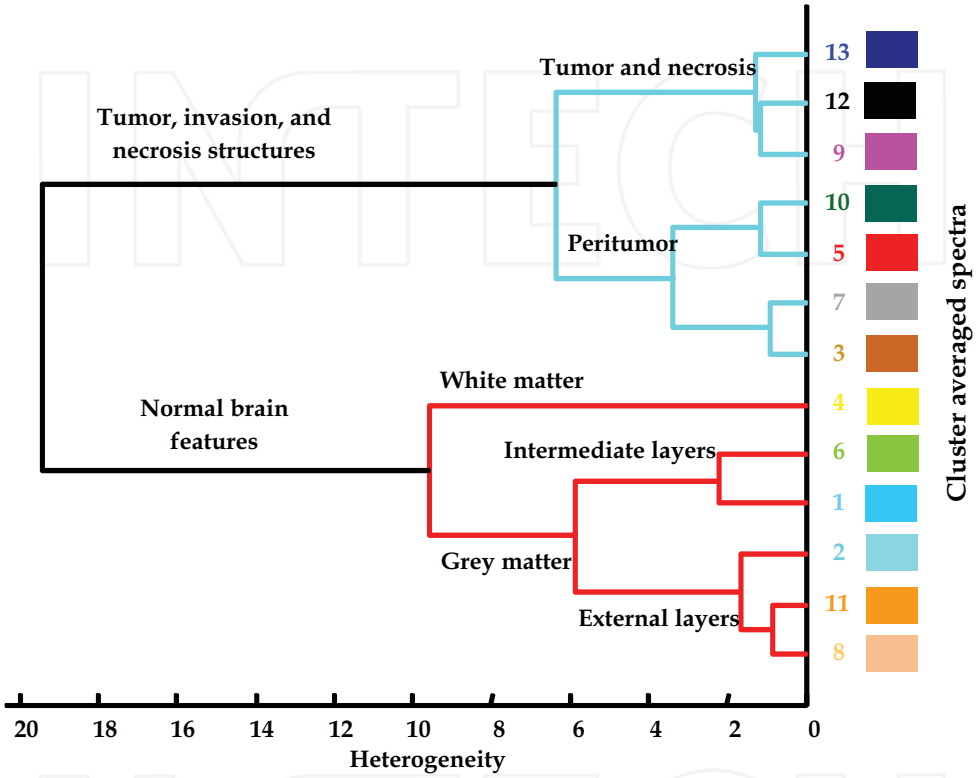


Fig. 4. Dendrogram obtained from hierarchical cluster analysis on spectral cluster averages associated to different tissue types. Heterogeneity represents the discriminating distance given by arbitrary units (au)

To distinguish between normal, tumor, and necrotic brain structures, cluster averaged spectra obtained from pseudo-color maps were input in the hierarchical cluster analysis using Ward's clustering algorithm and the square Euclidian distance measure. The result, as shown on the dendrogram in Fig. 4, showed a clear distinction between all normal and tumor brain structures. Indeed, the class related to normal brain structures was divided to two sub-clusters associated to white matter (clusters 4 corresponding to higher lipid content) and the second sub-cluster related to grey matter described by clusters 1 and 6 (intermediate cortex layers) and clusters 2, 8, and 11 (external cortex layers). We were also able to discriminate between glioma tumor (clusters 12, and 13) and necrosis (cluster 9) from peri-necrosis (clusters 3, 5, 7, and 10) in the second groups. In general, an increase in

malignancy is accompanied by a reduction in total lipids that involves all main classes of lipids found in plasma membranes (Campanella, 1992). Changes in lipid and in phospholipids contents, as seen in glioblastoma as compared with adjacent tissue, could indicate an evolution in the undergoing pathological process. This loss of lipids, correlated with demyelination, observed in different disorders, could be used as a spectroscopic marker. Increased levels of cholesterol esters (cholesterol oleate and linoleate) have also been reported in glioma tissue (Koljenovic et al., 2002). Koljenoviv et al. demonstrated that the difference between meningioma and dura is mainly related to lipids, cholesterol linoleate and linoleic acid levels. Steiner et al. studied the discriminating constituents between normal and tumoral tissues (astrocytoma and glioblastoma) by infrared spectroscopy (Steiner et al., 2003). They demonstrated that changes mainly arise from differences in lipid constituents. The potential use of lipid measurements for judging the stage, and hence the prognosis, of low grade tumors is suggested by the apparent gradual increase in lipid content over time. This increase, believed to be associated with necrosis, could thus be used in low grade tumors as an early marker of disease prior to the patient becoming symptomatic (Krafft et al., 2006; Koljenovic et al., 2005; Steiner et al., 2003). During tumor development, tissue composition and concentration of lipids decreased. Kraft et al. have investigated the lipid content of the white matter of human brain tissue using near infrared Raman spectroscopy (Krafft et al., 2005). They reported that the brain lipids can be divided into three principal classes: neutral lipids, phospholipids and sphingolipids.

3. Distribution of molecular changes in brain constituents associated to tumor growth and invasion

We have investigated the spatial distribution of molecular changes associated to C6 glioma progression using FT-IR micro-spectro-imaging in order to better understand the tissue transformation during carcinogenesis. Integrated intensity ratios bands were used in the region of 3700–2800 cm^{-1} to characterize differences between healthy and pathological brain. Maps of absorbance intensity ratios of bands in the region from 3000 to 2800 cm^{-1} due to CH_2 and CH_3 stretching vibrations (mainly due to membrane lipid which is directly related to the lipid acyl, primarily saturated) and those due to CH_2 and NH stretching vibrations (due to the protein-to-lipid ratios) were calculated and pseudo color maps scores were constructed (Fig. 5). These integrated absorbance intensity were correlated to molecular changes associated to tissue transformation. The comparison between the pseudo color scores maps and histological image shows that high scores described the white matter structures such as CC and CA (colorbar). These scores decreased in the grey matter and become null in the tumor tissues. This intensity ratio was correlated to myelin content. In fact, white matter presents important lipid content due to high myelin level. The concentration of myelin decreased from CC to cortex. In general, an increase in malignancy is accompanied by a reduction in total lipids that involves all main classes of lipids found in plasma membranes (Krafft et al., 2006). This loss of lipids was correlated with demyelination, observed in the scores maps associated to tumor. We have study the distribution of the intensity ratio 3292/2852 cm^{-1} corresponds to NH stretching vibration in proteins vs symmetric CH_2 stretching mode of the membrane lipid. The results show that high score value was related to tumor at days 7, 12, 15, and 19 days PI. These scores decreased from invasion zone to normal brain structures (white matter and grey matter). The common underlying effects of malignancy are changes in the constituents that lead to

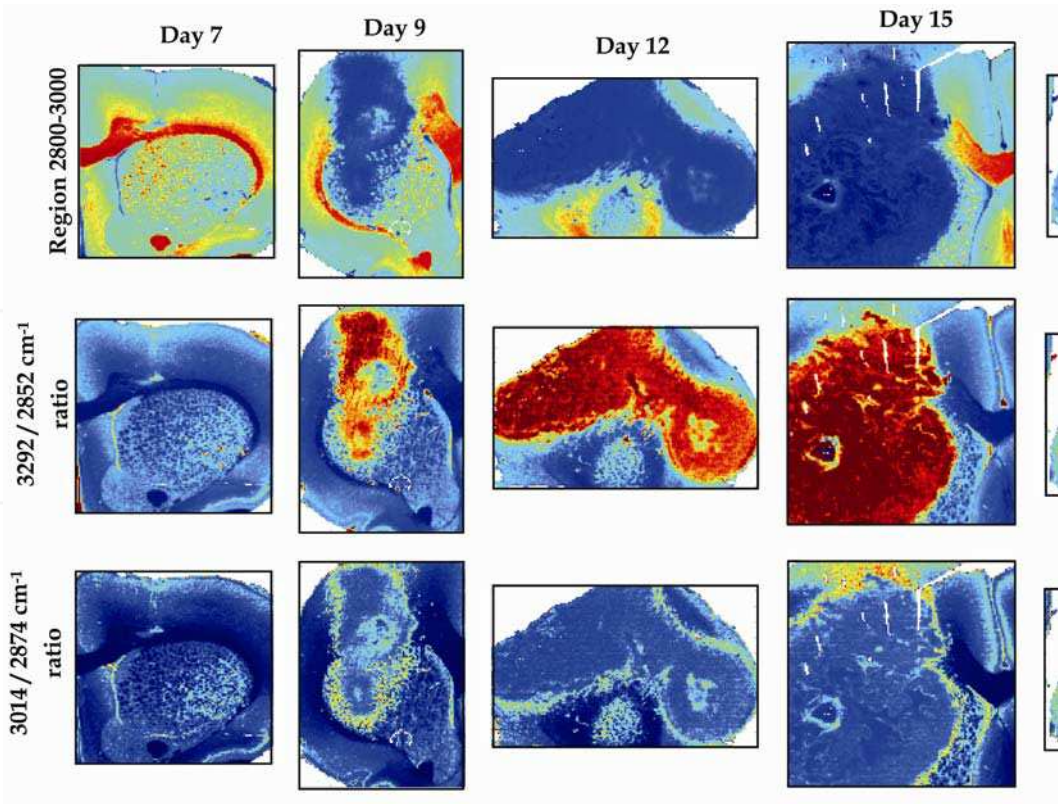


Fig. 5. Biochemical distribution of the changes in the molecular composition of tissues through the
Maps of absorbance intensity ratios of bands in the region from 3000 to 2800 cm^{-1} due to CH_2 and
(mainly due to membrane lipid which is directly related to the lipid acyl, primarily saturated) and
stretching vibrations (due to the protein-to-lipid ratio) were constructed and used to identify whic
be more potential indicators of such variations between normal and tumor development

an altered metabolism and biochemical composition in the malignant tissues. This study confirms the absorption intensity ratios were correlated with the histological state of the sample. Such a result may indicate the change of lipid and protein distribution in pathological tissues with respect to healthy one. There are some papers considering the pathological state of tissues by analysis of the changes in band intensities in the region from 3700 to 2800 cm^{-1}. Eckel et al. have analyzed breast cancer tissues (Eckel et al., 2001). They have considered the 3300 cm^{-1} band which corresponds to NH stretching vibration in proteins as a good criterion for distinguishing between non-cancerous and cancerous parts of tissue. Liu et al. applied IR absorption of human breast tissues in vitro and studied the CH stretching region of lipids and NH absorption region from 3600 to 2700 cm^{-1} (Liu et al., 2006). Gazi et al. used Synchrotron-FTIRM imaging to study prostate cancer cells and analyzed the distribution of lipid absorption intensities (CH_2 and CH_3) in the region from 3000 to 2800 cm^{-1} (Gazi et al., 2005) Krafft et al. have evaluated the usefulness of the lipid-to-protein ratio (2850/1655 cm^{-1}) as a spectroscopic marker to discriminate between normal and tumor tissue, as well as between low- and high-grade glioma tissues. They demonstrated that this ratio is maximal for normal brain tissue and decreases with the progression of the disease (Krafft et al., 2007). The intensity ratio 3014/2874 cm^{-1} was higher in the invasion zone and peritumor part of the necrosis and decreased in the tumor and normal brain tissues. This intensity ratio can be used to provide an objective method to delineate lesion margins that would reduce unnecessary tissue excisions.

4. Conclusion

This study demonstrated that FT-IRM imaging, with high spatially resolved morphological and biochemical information can be used as a diagnostic tool, complementary to histopathology in order to understand the molecular changes associated to C6 glioma progression. Cluster analysis allowed investigation of C6 glioma progression (from day 7 to day 19 post implantation). Different clusters in the FTIR images were correlated with features of the histopathological images such as white and grey matters, tumor, peritumor, and necrosis. Our results showed that 7 days after tumor implantation, FTIR investigations displayed a very small abnormal zone associated with the proliferation of C6 cells in the caudate putamen. From this day, rats developed solid and well-circumscribed tumors. Additionally, we have identified one peculiar structure all around the tumour. This structure was attributed to infiltrative events, such as peritumoral oedema observed during tumor development. The presence of necrotic areas was visible from day 15. In fact, the grade of malignancy and prognosis, in particular GBM multiforme, is based on the presence of necrosis. By combining intensity ratios of specific bands with imaging technique, we were able to take in account the variance due to the heterogeneity of brain tissues. We have monitored the changes in the intensity ratios of specific bands related to lipid and proteins. Our results reported that by correlating pseudocolor map scores with H&E staining it was possible to screen histological changes associated with tissue transformation. In fact, the integrated intensity in the 2800 to 3000 cm^{-1} spectral region described normal brain structures such as white matter (CC and CA) and some cortex layers (grey matter). The intensity decreased in the tumor tissues. Intensity ratio 3292/2852 cm^{-1} allowed the identification of tumor part of the tissue. The invasion zone was described by the 3014/2874 cm^{-1} ratio. These constituents can be used as spectroscopic markers for early detection of tissue abnormality and discrimination among normal, invasion, tumor and necrosis.

5. Acknowledgement

This work was supported by La Ligue de la Marne contre le Cancer, France.

6. References

Afanasyeva, N. I., Kolyakov, S.F., Artjushenko, S.G., Sokolov, V.V. & Frank, G.A. (1998). Minimally invasive and ex vivo diagnostics of breast cancer tissues by fiber optic evanescent wave Fourier transform IR (FEW-FT-IR) spectroscopy. *Proceedings of SPIE - The International Society for Optical Engineering*, pp. 140-146. 0277786X, San Jose, CA,

Amharref, N., Beljebbar, A., Dukic, S., Venteo, L., Schneider, L., Pluot, M., Vistelle, R.. & Manfait, M. (2006). Brain tissue characterisation by infrared imaging in a rat glioma model. *Biochim. Biophys. Acta*, Vol. 1758, No. 7, (07.2006) 892-899, ISSN 00052736

Amharref, N., Beljebbar, A., Dukic, S., Venteo, L., Schneider, L. Pluot, M., & Manfait, M. (2007). Discriminating healthy from tumor and necrosis tissue in rat brain tissue samples by Raman spectral imaging. *Biochim. Biophys. Acta*, Vol. 1768, No. 10, (10 2007) 2605-2615, ISSN 00052736.

Auer, R.N., Del Maestro, R.F., Anderson, R. (1981). A simple and reproducible experimental in vivo glioma model. *Can J Neurol Sci.,* Vol. 8, No. 4, (11 1981), pp. 325-331, ISSN 0317-1671.

Barker, F.G., Davis, R.L., Chang, S.M., & Prados, M.D. (1996). Necrosis as a prognostic factor in glioblastoma multiforme. *Cancer*, Vol. 77, No. 6, (03 1996), 1161-1166, ISSN 0008543X.

Barth, R.F. (1998). Rat brain tumor models in experimental neuro-oncology: the 9L, C6, T9, F98, RG2 (D74), RT-2 and CNS-1 gliomas. *Neurooncol*., Vol. 36, No. 1, (01 1998), pp. 91-102, ISSN 0167-594X.

Bates, J.B. (1976). Fourier transform infrared spectroscopy. *Science*, Vol. 9, No. 191, 4222, (01.1976), pp. 31-37, ISSN 00368075

Beljebbar, A., Amharref, N., Lévèques, A., Dukic, S., Venteo, L., Schneider, L., Pluot, M., & Manfait, M. (2008). Modeling and quantifying biochemical changes in C6 tumor gliomas by Fourier transform infrared imaging. *Anal. Chem*., Vol. 15, No 80, (11.2008), pp. 8406-8415, ISSN 00032700

Bernstein, J.J., Laws, E.R. Jr, Levine, K.V., Wood, L.R., Tadvalkar, G., Goldberg, W.J. (1991). C6 glioma-astrocytoma cell and fetal astrocyte migration into artificial basement membrane: a permissive substrate for neural tumors but not fetal astrocytes. *Neurosurgery*, Vol. 28, No. 5, (5 1991), pp. 652-658, ISSN 0148-396X.

Burwell, R.D. (2001). Borders and cytoarchitecture of the perirhinal and postrhinal cortices in the rat, *J. comparative neurol.*. Vol. 437, No. 1, (08 2001), pp. 17-41, ISSN 00219967.

Campanella, R., (1992). Membrane lipids modifications in human gliomas of different degree of malignancy. *J. Neurosurg. Sci.,* vol. 36, No. 1, (1992), pp. 11–25, ISSN 00264881.

Cherayil, G.D.& Scaria, K.S. (1970). Thin-layer chromatography of tissue lipids without extraction. J. Lipid Res., Vol. 11, No. 4, (07 1970), pp. 378-381, ISSN 00222275.

Chicoine, M.R., & Silbergeld, D.L. (1995). Invading C6 glioma cells maintaining tumorigenicity. *J Neurosurg* Vol. 83, No. 4, (10 1995), pp. 665-671, ISSN 0022-3085.

Ci, Y., Gao, T., Feng, J., & Guo Z. (1999). Fourier transform infrared spectroscopic characterization of human breast tissue: implications for breast cancer diagnosis. *Applied Spectrosc.*, Vol. 53, No. 3, (03.1999), pp. 312–315, ISSN 00037028

Cohenford, M.A. & Rigas B. (1998). Cytologically normal cells from neoplastic cervical samples display extensive structural abnormalities on IR spectroscopy: implications for tumor biology. *Proc. Natl. Acad. Sci. U S A.,* Vol. 95, No. 26, (12.1998), pp. 15327–15332, ISSN 00278424

Diem, M., Boydston-White, S. & Chiriboga, L. (1999). Infrared spectroscopy of cells and tissues: shining light onto a novel subject. *Appl. Spectrosc.*, Vol. 53, No. 4, pp. 148-161, ISSN 00037028

Eckel, R., Huo, H., Guan, H.W., Hu, X., Che X., & Huang W.D. (2001).Characteristic infrared spectroscopic patterns in the protein bands of human breast cancer tissue. *Vib. Spectrosc.,* Vol. 27, No. 2,, (12 2001), pp. 165–173, ISSN 09242031.

Fernandez, D.C., Bhargava, R., Hewitt, S.M.& Levin, I.W. (2005). Infrared spectroscopic imaging for histopathologic recognition. *Nature Biotechnology,* Vol. 23, No. 4, pp. 469-474, ISSN 10870156

Franck, P., Nabet, P. & Dousset, B. (1998). Applications of infrared spectroscopy to medical biology. *Cell. Mol. Biol.*, Vol. 44, No. 2, (03.1998), pp. 273- 275, ISSN 01455680

Gazi, E., Dwyer, J., Lockyer, N.P., Miyan, J., Gardner, P., Hart, C.A., Brown M.D., & Clarke N.W. (2005). A study of cytokinetic and motile prostate cancer cells using synchrotron-based FTIR microspectroscopic imaging. *Vib. Spectrosc.*, Vol. 38, No. 1-2, (07 2005), pp. 193–201, ISSN 09242031.

Goodacre, R., Vaidyanathan, S., Dunn, W.B., Harrigan, G.G. & Kell, D.B. (2004). Metabolomics by numbers: Acquiring and understanding global metabolite data, *Trends Biotechnol.*, Vol. 22, No. 5, (05.2004), pp. 245–252, ISSN 01677799.

Grobben, B, De Deyn, P.P., & Slegers, H. (2002). Rat C6 glioma as experimental model system for the study of glioblastoma growth and invasion. *Cell Tissue Res.*, Vol. 310, No. 3,, pp. 257-270, ISSN 0302766X.

Jackson, M., Ramjiawan, B., Hewko, M., & Mantsch, H.H.. (1998). Infrared microscopic functional group mapping and spectral clustering analysis of hypercholesteramic rabbit liver. *Cell. Mol. Biol.*, Vol. 44, No. 1, (02.1998), pp. 89–98, ISSN 01455680

Kleihues, P., Burger, P.C ., Scheithauer, B.W. (1993). The new WHO classification of brain tumours. *Brain Pathol.*, Vol. 3, No. 3, (03 1993), pp. 255-68, ISSN 10156305.

Kneipp, J., Lasch, P., Baldauf, E., Beekes, M., & D. Naumann, (2000). Detection of pathological molecular alterations in scrapie-infected hamster brain by Fourier transform infrared spectroscopy. *Biochim. Biophys. Acta*, Vol. 1501, No. 2-3, (06 2000), pp. 189-199, ISSN 09254439.

Koljenovic, S., Choo-Smith, L.P., Bakker Schut, T.C., Kros, J.M., van den Berge, H.J., & Puppels, G. (2002). Discriminating vital tumor from necrotic tissue in human glioblastoma tissue samples by Raman spectroscopy. *Lab. Invest.*, Vol. 82, No. 100, (10 2002) pp. 1265-1277, ISSN 00236837.

Koljenovic, S., Schut, T.B., Vincent, A., Kros, J.M., Puppels, G.J. (2005). Detection of meningioma in dura mater by Raman spectroscopy. *Anal. Chem.*, Vol. 77, No. 24, (12 2005), pp. 7958-7965, ISSN 00032700.

Krafft, C., Neudert, L., Simat, T., & Salzer, R. (2005). Near infrared Raman spectra of human brain lipids. *Spectrochim. Acta (Part A),* Vol. 61, No. 7, (05 2005), pp. 1529-1535, ISSN 13861425.

Krafft, C., Sobottka, S.B., Geiger, K.D., Schackert, G., & Salzer, R. (2007). Classification of malignant gliomas by infrared spectroscopic imaging and linear discriminant analysis. *Anal. Bioanal. Chem.*, Vol. 387, No. 5, (03 2007), pp. 1669-1677, ISSN 16182642

Krafft, C., Sobottka, S.B., Schackert, G., & Salzer R. (2006). Raman and infrared spectroscopic mapping of human primary intracranial tumors: A comparative study. *J. Raman Spectrosc.*, Vol. 37, No. 1-3, (01 2006), pp. 367-375, ISSN 03770486.

Krafft, C., Thummler, K., Sobottka, S.B., Schackert, G., & Salzer, R. (2006). Classification of malignant gliomas by infrared spectroscopy and linear discriminant analysis. *Biopolymers*, Vol. 82, No. 4, (07 2006), pp. 301-305, ISSN 00063525

Lasch, P & Naumann D. (1998). FT-IR microspectroscopic imaging of human carcinoma thin sections based on pattern recognition techniques. *Cell. Mol. Biol.*, Vol. 44, No. 1, (02.1998), pp. 189–202, ISSN 01455680.

Lasch, P., Haensch, W., Naumann, D., & Diem, M. (2004). Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis. *Biochim Biophys Acta., Vol.* 1688, No. 2, (03.2004), pp. 176–186, ISSN 09254439.

Laws, E.R. Jr & Shaffrey, M.E. (1999). The inherent invasiveness of cerebral gliomas: Implications for clinical management. *Int. J. Dev. Neurosci.,* Vol. 17, No. 5-6, (08 1999), pp. 413-420, ISSN 07365748

Levin, I.W. & Bhargava, R. (2005). Fourier transform infrared vibrational spectroscopic imaging: Integrating microscopy and molecular recognition. *Annu. Rev. Phys. Chem.*, Vol. 56, pp. 429-474, ISSN 0066426X

Li, Q.B., Sun, X.J., Xu, Y.Z., Yang, L.M., Zhang, Y.F., Weng, S.F., Shi, J.S & Wu, J.G (2005). Diagnosis of gastric inflammation and malignancy in endoscopic biopsies based on Fourier transform infrared spectroscopy. *Clin. Chem.*, vol. 51, No. 2, (02.2005), pp 346–350, ISSN 00099147

Liu, C., Zhang, Y., Yan, X., Zhang, X., Li, C., Yang W., & Shi, D. (2006). SR-FTIR spectroscopic preliminary findings of non-cancerous. J. Lumin., Vol. 119–120, SPEC, (07 2006), pp. 132–136, ISSN 00222313.

Mansfield, J.R., McIntosh, K., Crowson, A.N., Mantsch, H.H. & Jackson, M. (1999). LDA-guided search engine for the non-subjective analysis of infrared microspectroscopic maps. *Appl. Spectrosc., Vol.* 53, No. 11, (11.1999), pp. 1323–1330, ISSN 00037028.

Mansfield, J.R., Sowa, M.G., Scarth, G.B., Somorjai, R.L. & Mantsch, H.H. (1997). Analysis of spectroscopic imaging data by fuzzy C-means clustering. *Anal. Chem.*, Vol. 69, No. 16, (08.1997), pp. 3370–3374, ISSN 00032700.

Mohlenhoff, B., Romeo, M., Diem, M., & Wood, B.R. (2005). Mie-type scattering and non-Beer-Lambert absorption behaviour of human cells in infrared microspectroscopy. *Biophys J.,* Vol. 88, No. 5, (05.2005), pp. 3635–3640, ISSN 00063495.

Mourant, J. R., Yamada, Y.R. Carpenter, S., Dominique, L.R. & Freyer, J.P. (2003). FTIR spectroscopy demonstrates biochemical differences in mammalian cell cultures at different growth stages. *Biophys. J.*, Vol. 85, No. 3, (09.2003) 1938–1947, ISSN 00063495.

Nagano, N., Sasaki, H., Aoyagi, M., & Hirakawa, K. (1993). Invasion of experimental rat brain tumor: early morphological changes following microinjection of C6 glioma cells. *Acta Neuropathol*., Vol. 86, No. 2, (1993), pp. 117-125, ISSN 00016322.

Parker, F. S., (1971). *Application of infrared spectroscopy in biochemistry, biology and medicine* Plenum, New York.

Petibois, C. & Déléris, G. (2006). Chemical mapping of tumor progression by FT-IR imaging: towards molecular histopathology. *Trends in Biotechnology*, Vol. 24, No. 10, (08.2006), pp. 455-462, ISSN 01677799

Rigas, B., Morgello, S., Goldman, I., & Wong, P. (1990). Human colorectal cancers display abnormal Fourier-transform infrared spectra. *Proc. Natl. Acad. Sci. USA*, Vol. 87, No. 20, (10.1990), pp. 8140–8144, ISSN 00278424

San-Galli, F, Vrignaud, P., Robert, J., Coindre, J.M., Cohadon, F. (1989). Assessment of the experimental model of transplanted C6 glioblastoma in Wistar rats. *J. Neurooncol*., Vol. 7, No. 3, (9 1989), pp. 299-304, ISSN 0167594X.

Savitzky, A. & Golay, M.J.E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem*., Vol. 36, No. 8, (08 1964), pp. 1627-1639

Steiner, A., Shaw, G., Choo-Smith, L.P., Schackert, G., Steller, W., Abuid, H.M., & Salzer, R. (2003). Distinguishing and grading human gliomas by infrared spectroscopy. *Biopolymers*, Vol. 72, No. 6, (2003), pp. 464-471, ISSN 00063525.

Wang, J.S., Shi, J.S., Xu, Y.Z., Duan, X.Y., Zhang, L., Wang, J., Yang, L.M., Weng, S.F., & Wu, J.G. (2003). FT-IR spectroscopic analysis of normal and cancerous tissues of esophagus. *World J. Gastroenterol*., Vol. 9, No. 9, (09 2003), pp. 1897-1899, ISSN 10079327.

Wong, P.T., Wong, R.K., Caputo, T.A., Godwin, T.A. & Rigas, B. (1991). Infrared spectroscopy of exfoliated human cervical cells: Evidence of extensive structural changes during carcinogenesis. *Proc. Natl. Acad. Sci. USA*, Vol. 88, No. 24, (12.1991), pp. 10988–10992, ISSN 00278424

Yano, K., Ohoshima, S., Gotou, Y., Kumaido, K., Moriguchi, T.& Katayama H. (2000). Direct measurement of human lung cancerous and noncancerous tissues by Fourier transform infrared microscopy: can an infrared microscope be used as a clinical tool? *Anal. Biochem.* Vol. 287, No. 2, (12.2000), pp. 218–225, ISSN 00032697

Zhang, L., Small, G.W., Haka, A.S., Kidder, L.H., & Lewis, E.N. (2003). Classification of Fourier transform infrared microspectroscopic imaging data of human breast cells by cluster analysis and artificial neural networks. *Appl. Spectrosc*., Vol. 57, No. 1, (01.2003), pp. 14–22, ISSN 00037028.

8

Statistical Analysis for Automatic Identification of Ovarian Cancer Protein-Biomarkers Based on Fast Fourier Transform Infrared Spectroscopy

Marcano A.[1,2,3], D. Pokrajac[1,4], A. Lazarevic[1],
M. Smith[3], Y. Markushin[1,2] and N. Melikechi[1,2,3]
[1]*Center for Research and Education in Optical Sciences and Applications,*
[2]*Center for Applied Optics for Space Science,*
[3]*Department of Physics and Pre-Engineering*
[4]*Department of Computer and Information Sciences,*
Delaware State University, 1200 North Dupont Highway, Dover, DE 19901
United States of America

1. Introduction

Fast Fourier transform infrared (FTIR) spectroscopy has been used widely for the study of vibrations of protein molecules (Arrondo et al. 1993; Goormaghtigh et al. 1994; Haris & Chapman, 1994; Siebert, 1995; Barth & Zscherp, 2002; Petibois et al., 2006). Valuable information can be obtained of the secondary structure of the protein since peak positions and their relative amplitude are affected by the number of hydrogen bridges that sustain this secondary structure (Byler & Sussi, 1989; Fabian et al., 2001; Fabian et al., 2002). However, the spectral lines of proteins are usually broadened due to different molecular interactions thus making the identification of the structure difficult. Furthermore, identification of a particular protein within a complex matrix like a blood or a serum sample based on FTIR spectra is particularly challenging. Namely, direct application of automatic classification techniques is not a simple task, due to large numbers of attributes (measurements at different wavenumbers). Recently, principal component analysis (PCA) has been used as a statistical method for the feature extraction in the analysis of spectroscopic data aimed at detection of several complex organic samples (Hybl et al., 2003; Melikechi et al.; 2008, Lazarevic et al., 2009). In these methods, the spectroscopic data can be represented in a three-dimensional (or arbitrary dimension) space of eigenvector projections of the matrices corresponding to a series of experimental data measured for different selected wavelengths (Massart et al., 2003). In this regard, each point of this space represents a full set of spectroscopic measurements corresponding to one sample. Differences between the spectra can be then visualized graphically as different points in the space of eigenvectors. Linear discriminant analysis (LDA) or support vector machines (SVM), an advanced machine learning technique, can be subsequently used for automatic observing these differences between spectra. LDA generates linear models that separate classes based on the assumption that class-wise distributions are multivariate Gaussian with the same

covariance matrix (independent of the class label). SVM are classification algorithms that automatically assign a class label to a vector of data with theoretically best generalization (ability to predict the class outside the training data), independently of the data distribution. SVM generate a hyperplane in the transformed feature space (a non-linear transformation applied to the original data) such that the separation plane is as far from the data closest to it as possible. By using non-linear transformation, the likelihood that the training data can be separated by a hyperplane increases. By maximizing the distance between the data and the hyperplane, we achieve the smallest complexity of the classifier and hence, according to computational learning theory, maximize the generalization capability of the classification model. In this study, we propose to use the output of PCA analysis as input of LDA and SVM and to perform an automatic identification of protein molecules based on their FTIR spectra.

We use the proposed methodology to distinguish among the fast Fourier transform infrared (FTIR) spectra of proteins reported as possible biomarkers of ovarian cancer: monoclonal antibodies (MAB) and antigens (AG) of ovarian cancer marker CA125, Osteopontin (OPN), Leptin and insulin-like growth factor II (IGF2) (Mor et al., 2005; Schorge et al., 2004; Sutphen et al., 2004). We also complete a similar study on the common protein Bovine Serum Albumin (BSA) and human plasma samples for comparison purposes. We show that despite the presence of broadening mechanisms and evident similarities in the FTIR spectra of these proteins, the proposed method provides an automatic and effective identification of the proteins with almost perfect accuracy. This statistical procedure can also be applied to other spectroscopic methods such as fluorescence, NIR-VI absorbance spectroscopy and laser-induced breakdown spectroscopy.

As an important application we also perform deuteration of proteins and study the differences in the FTIR spectra introduced by this process using the PCA and LDA methods. FTIR spectra of deuterated versions of the proteins have been used extensively for the study of the secondary structure (Baenziger & Methot, 1995; Dave et al., 2002; Nie et al., 2005). Deuteration occurs by simple dilution of proteins in heavy water that contains the deuterium isotope of hydrogen (^2H). We have studied in details the changes induced by deuteration in the FTIR spectra of BSA and ovarian cancer biomarkers referred above. We have also explored the use of temperature and ultrasound to increase the changes. We use PCA and LDA methods to differentiate undeuterated and deuterated versions of the same protein. We propose that these methods can be used for identification of proteins within a matrix containing a large variety of proteins like a blood or serum sample. Furthermore, we propose a FTIR based immunoassay that uses the developed data analysis method and deuterated versions of the corresponding monoclonal antibodies for detection of protein biomarkers contained in a complex matrix like blood, plasma or serum samples.

2. Experimental method

For measuring the FTIR spectra we use an attenuated total reflection (ATR) FTIR spectrophotometer NICOLET 6700 (Thermo Industries, Inc). Drops of the samples are deposited over an aperture on the top of the device. This aperture connects to the surface of a diamond prism where the total reflection occurs. Samples under study are distilled and deionized water, heavy water (99.8% purity Deuterium oxide from Alfa Easer) and high

purity proteins (Sigma): BSA, 15 mM saline solution of MAB to AG CA125, AG CA125, Leptin, OPN and IGF2. Usually water masks most of the contribution from the proteins. To eliminate water peaks the samples are dried through simple evaporation of the solvent before collecting data. A drop of 5 μL of the solution is deposited over the aperture of the spectrophotometer. The samples are then left to dry at room temperature during 30 minutes. The drying process is monitored by taking spectra every 5 minutes until solvent (water of heavy water) contribution is depleted. When the drying process is complete the spectra do not show further changes. The dried protein sample forms a film over the aperture of several tens of micrometers good enough for total reflection spectroscopy. The spectra are collected with a resolution of 4 cm^{-1}. One hundred scans are averaged for each spectrum. The spectra show high reproducibility and a signal to noise value usually larger than 100. To collect data for the data analysis we repeat the spectroscopy experiment 40 times for each specimen. The deuteration of the proteins is performed using the dilution method at different concentrations and different dilution times. For solid samples like BSA we prepare directly a heavy water solution of the protein. For the solution samples we mix equal volumes of D_2O and the original protein solution. Deuteration can be improved by adding additional drops to the previously dried sample. Deuteration can also be improved by changing the temperature or using ultrasound. For this purpose we use an ultrasound cleaner with temperature control (Fisher Scientific FS20). The temperature is monitored with an independent thermocouple.

3. Classification methodology

We propose a statistical framework for automatic classification of the FTIR spectra of different proteins. The framework is illustrated in figure 1.

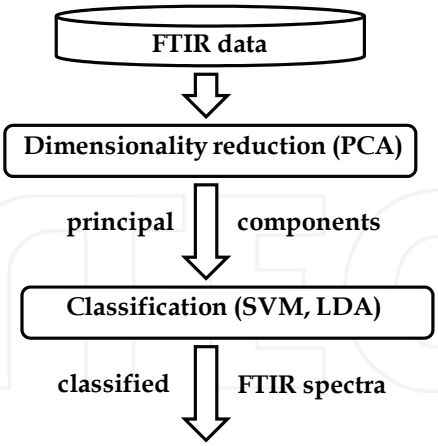


Fig. 1. Statistical framework for automatic FTIR spectra classification

The first step in the framework corresponds to dimensionality reduction, in which we reduce the number of frequencies in FTIR spectra using PCA. Principal components obtained through PCA are them used as an input to the classification module, which provides final classification of particular FTIR spectra.

3.1 Principal component analysis (PCA)

PCA is a powerful technique for dimensionality reduction in machine learning and data mining (Jolliffe, 2002). The central idea of PCA is to reduce the dimensionality of a data set consisting of a large number of interrelated variables (with non-diagonal covariance matrix), while retaining as much as possible of the variation present in the data set. This reduction is achieved by transforming to a new set of variables, which are called the principal components (PC). The PC are uncorrelated, and ordered in such a way that the first few retain most of the variation present in all of the original variables.

Suppose that **X** is a N-dimensional matrix of k-dimensional random variables $[x_1; x_2; x_N]$, and that the variances of the k random variables and the structure of the covariances or correlations between the k variables are of interest. Assume that we intend to approximate the vector x_i as a linear combination of m<k predetermined variables. In other words, assume that we would like to determine $\hat{\mathbf{x}}_i = \sum_{j=1}^{m} a_{ij}\mathbf{v}_j$, such that the mean square error $E\left(\left|\hat{\mathbf{x}}_i - \mathbf{x}_i\right|^2\right)$ is minimized. It can be proven that the mean square error is minimized when $\mathbf{v}_i$, i = 1,…,m are eigenvectors corresponding to m largest eigenvalues of the covariance matrix of **X**, and when a_{ij} are principal values projections of vector $\mathbf{x}_i$ with respect to its mean and first m eigenvectors. The vector $\hat{\mathbf{x}}_i$ contains m variables and thus it is typically stated that m "most significant" features are extracted out of k original coordinates.

The covariance matrix C of **x** can be estimated as:

$$C = \frac{1}{N-1} X^T X \ , \tag{1}$$

Its eigenvectors (column vectors) and corresponding eigenvalues satisfy the following condition:

$$C\,\mathbf{v}_j = \lambda_j \mathbf{v}_j \ , \tag{2}$$

We can formally define eigenvector matrix and the diagonal matrix of eigenvalues respectively as:

$$\mathbf{V} = [\mathbf{v}_1\ \mathbf{v}_2 \ldots\ \mathbf{v}_k] \ , \tag{3}$$

$$\Lambda = \mathrm{diag}\left(\lambda_1, ..., \lambda_k\right) . \tag{4}$$

Therefore we can compute the coefficients a_{ij} as $a_{ij} = \mathbf{x}_i\ \mathbf{v}_j$. Computation of PCA for high-dimensional data by definition may be very cumbersome, since it has $O(k^3)$ complexity, where k is the number of dimensions. The reason for such computational complexity is the requirement to compute eigenvalues and eigenvectors of a k*k matrix $C = \frac{1}{N-1} X^T X$. To make computation feasible, we will follow recently proposed approach (Bishop, 2006) that can extract up to N-1 principal components with the largest eigenvalues, when N < k.

Define $C^* = \frac{1}{N-1} X X^T$, and let U and Λ^* be respectively matrices of eigenvectors and eigenvalues of C^* such that:

$$C^* U = U \Lambda^* \ , \tag{5}$$

If we assume $V = X^T U$ and consider the estimated covariance matrix C, we can easily obtain that:

$$CV = \frac{1}{N-1} X^T X X^T U = X^T C^* U = X^T U \Lambda^* = V \Lambda^* \quad , \tag{6}$$

In other words, V is the eigenvector matrix of C and Λ^* is corresponding diagonal matrix of eigenvalues. Since C^* is of size N*N, using this technique would allow huge computational savings when N<<k. Note that vectors in V are not necessary normalized (to have a unit norm). Hence, to achieve orthonormal eigenvectors, an additional normalization step is required.

3.2 Linear Discriminant Analysis (LDA)

LDA (Krzanowski, 1988; Seber, 1984) is a statistical technique that classifies objects by computing the logarithm of the likelihood function (likelihood is the probability of the class given the observed data). Here, data from each class is assumed to belong to a multivariate Gaussian distribution. The Gaussian distributions corresponding to different classes are assumed to have the different means but the same covariance matrix, leading to the linear discrimination.

Formally, given the estimates of the prior probabilities p_j, and means $\boldsymbol{\mu}_j$ for each class j, and the estimate of the covariance matrix *C*, the logarithmic likelihood for a sample specified by a vector $\mathbf{y}_i$ can be computed as

$$l(j) = -\frac{1}{2}\ln|C| + \ln p_j - \frac{1}{2}\left(y_i - \boldsymbol{\mu}_j\right)^T C^{-1}\left(y_i - \boldsymbol{\mu}_j\right) \quad , \qquad j = 1,...,c \ , \tag{7}$$

where c is the total number of classes.

Using eq. (7), the classification of an example from a test set, specified by vector $\mathbf{y}_{new}$, is performed according to:

$$c_{new} = \arg\max_j l(j) = \arg\max_j \left(\ln p_j - \frac{1}{2}\boldsymbol{\mu}_j C^{-1} \boldsymbol{\mu}_j^T + \boldsymbol{\mu}_j C^{-1} \mathbf{y}_{new}^T \right) . \tag{8}$$

Hence, the separation plane between classes i and j can be described as a hyperplane:

$$f_{ij}(\mathbf{y}) = \left(\boldsymbol{\mu}_i C^{-1} - \boldsymbol{\mu}_j C^{-1}\right) y^T + \left(\ln p_i - \ln p_j - \frac{1}{2}\boldsymbol{\mu}_i C^{-1} \boldsymbol{\mu}_i^T + \frac{1}{2}\boldsymbol{\mu}_j C^{-1} \boldsymbol{\mu}_j^T \right) = 0 . \tag{9}$$

For each class, we can define a decision margin as a minimal distance between a sample from a class and the separation planes.

Let $\mathbf{y}_{i,j}$, i =1,…,n_j be row feature vectors from the training set belonging to class j and let n_j be the number of vectors in class j. We estimate the class priors, means, and the covariance matrix as:

$$p_j = \frac{n_j}{\sum_{j'=1,...,c} n_{j'}}$$

$$\boldsymbol{\mu}_j = \frac{1}{n_j}\sum_{i=1,\ldots,n_j} \mathbf{y}_{i,j} \tag{10}$$

$$C = \frac{1}{\sum_{j'=1,\ldots,c} n_{j'} - 1}\sum_{j'=1,\ldots,c} \frac{C_{j'}}{n_{j'} - 1}$$

where:

$$C_j = \frac{1}{n_j - 1}\sum_{i=1,\ldots,n_j} \left(\mathbf{y}_{i,j} - \boldsymbol{\mu}_j\right)^T \left(y_{i,j} - \boldsymbol{\mu}_j\right) \ . \tag{11}$$

3.3 Support Vector Machines (SVM)

LDA provides optimal classification with linear decision boundaries if its assumption of class-specific Gaussian distributions with identical covariances is satisfied. However, if this assumption is not satisfied, the optimal decision boundaries could be obtained by using SVM (Vapnik, 2000). The main idea of SVM is to construct a separation hyperplane, which optimally separates data examples belonging to two classes, such that the minimal distance between points and the separation hyperplane is maximized. Such constructed hyperplane provides the best generalization of unknown examples. SVM use structural risk minimization principle and aim to achieve zero training error while minimizing the complexity of the model. However, if linear separation is not possible, SVM work towards minimization of the number of misclassified examples on the training set by introducing the slack variables and regularization. Formally, SVM learning can be represented as the following quadratic programming problem (Bishop, 2006):

$$\begin{aligned} &\min_{\mathbf{w},\boldsymbol{\xi}_i,d_0} \left(\frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\boldsymbol{\xi}_i\right) \text{s.t.} \\ &\left(\mathbf{w}^T\mathbf{y}_i + d_0\right)\cdot c_i \geq 1 - \boldsymbol{\xi}_i, i = 1,\ldots,N \\ &\boldsymbol{\xi}_i \geq 0, i = 1,\ldots,N. \end{aligned} \tag{12}$$

where, $\mathbf{w}$ is vector defining the separation hyperplane, d_0 is the intercept of the separation hyperplane, $c_i \in \{-1,1\}$ is a class label of the i^{th} example determined by attribute vector $\mathbf{x}_i$, ξ_i is the slack variable corresponding to the i^{th} example, C is a preset regularization constant, and $\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i)$ is a vector representing a non-linear function of $\mathbf{x}_i$. In general, $\mathbf{x}_i$ and $\mathbf{w}$ can be infinitely dimensional.

Using the Karush-Kuhn-Tucker (KKT) theorem (Karush, 1939), SVM learning can be represented as the optimization in dual space of Lagrangian multipliers λ_i. In this case, the learning phase reduces to the following optimization problem:

$$\begin{aligned} &\max_{\lambda}\left(\sum_{i=1}^{N}\lambda_i - \sum_{l=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j c_i c_j \mathbf{y}_i^T\mathbf{y}_j\right) \text{ s.t.} \\ &\sum_{i=1}^{N}\lambda_i c_i = 0 \\ &\lambda_i \geq 0, i = 1,\ldots,N \\ &\lambda_i \leq C, i = 1,\ldots,N. \end{aligned} \tag{13}$$

An example $\mathbf{x}_{new}$ from the test set is subsequently classified according to the following equation:

$$c_{new} = \text{sign}\left(\mathbf{w}^T \mathbf{y}_{new} + d_0\right) \quad , \tag{14}$$

where $\mathbf{y}_{new} = f(x_{new})$, which can be expressed using the Lagrangian multipliers as:

$$c_{new} = \text{sign}\left(\sum_{i:\lambda_i>0} \lambda_i c_i \mathbf{y}_i{}^T \mathbf{y}_{new} + \frac{1}{N_s}\left(\frac{1}{c_i} - \sum_{j:\lambda_j>0} \lambda_j c_{ji}{}^T \mathbf{y}_i{}^T \mathbf{y}_j\right)\right), \tag{15}$$

where N_s denotes the number of support vectors—points closest to the separation hyperplane (i.e., number of non-zero Lagrangian multipliers).
Similar as in the case of LDA, with SVM we can explicitly calculate the separation planes in the transformed space specified by:

$$w^T y + d_o = 0 \quad , \tag{16}$$

where

$$w = \sum_{i:\lambda_i>0} \lambda_i y_i c_i \quad , \tag{17}$$

$$d_0 = \frac{1}{N_s} \sum_{i:\lambda_i>0} \left(\frac{1}{c_i} - w^T \mathbf{y}_i\right).$$

If we select features in the transformed space to be proportional to eigenvalues and eigenfunctions of a symmetric non-negative definite kernel, then, due to the Mercer's theorem, we can write $\mathbf{y}_i^T \mathbf{y}_j = K(\mathbf{x}_i, \mathbf{x}_j)$ where K is symmetric non-negative-definite function of two vectors (Bishop, 2006). Then, due to the Mercer's spectral theorem for non-negative definite symmetric kernels, SVM learning and classification can be stated directly using original (non-transformed) feature vectors as:

$$\max_{\lambda}\left(\sum_{i=1}^{N}\lambda_i - \sum_{l=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j c_i c_j K(\mathbf{x}_i, \mathbf{x}_j)\right) \text{ s.t}$$

$$\sum_{i=1}^{N}\lambda_i c_i = 0$$
$$\lambda_i \geq 0, i = 1,...,N \tag{18}$$
$$\lambda_i \leq C, i = 1,...,N.$$

$$c_{new} = \text{sign}\left(\sum_{i:\lambda_i>0} \lambda_i c_i K(\mathbf{x}_i, \mathbf{x}_{new}) + \frac{1}{N_s}\left(\frac{1}{c_i} - \sum_{j:\lambda_j>0} \lambda_j c_{ji}{}^T K(\mathbf{x}_i, \mathbf{x}_j)\right)\right), \tag{19}$$

This makes possible using implicit and infinitely dimensional transformation **f**. Popular choices of kernel function include:

- Linear kernel: $K(\mathbf{u}, \mathbf{v}) = \mathbf{u}^T \mathbf{v}$;

- Polynomial kernel (*p* is a prespecified parameter): $K(\mathbf{u},\mathbf{v}) = (1+\mathbf{u}^T\mathbf{v})^P$;
- Exponential kernel (σ is a pre-specified parameter): $K(\mathbf{u},\mathbf{v}) = e^{-\frac{1}{2\sigma^2}\|\mathbf{u}-\mathbf{v}\|^2}$.

The original SVM technique is designed for a two-class problem. For a multiclass problem (i.e., c>2) we use Directed Acyclic Graph SVM (DAG-SVM) method (Platt et al., 2000) which for a c-class problem trains c(c-1)/2 two-class support machines and the class decision is performed based on successive elimination of classes as a result of a two-class comparison. In comparison to one-to-rest classifiers, the application of DAG-SVM is more practical, since it does not result in imbalanced training sets (Hsu & Lin, 2002; Jiang et al. 2005).

3.4 Classification accuracy evaluation

To validate the accuracy of the classification model on the data unseen during the learning process, we use a four-fold cross validation, that can be described as follows (Bishop, 2006): 1) split randomly the dataset into four subsets; 2) set aside one of the subsets as the test set while the other three subsets are chosen to form the training set; 3) Utilize the training set to learn the classification model and employ the test set to evaluate the accuracy of classification on data unseen during the learning process; 3) repeat the process four times so that each of the four subsets has a chance to be a test set; 4) use averaged results from the four classification experiments as an overall measure of the model performance.

As a measure of performance, we utilize overall classification accuracy, the ratio of correctly classified samples for all classes versus the number of all classified samples in the test set (Bramer, 2007), defined as:

$$\text{Overall Accuracy} = \frac{\text{correctly classified samples from all classes}}{\text{total number of samples}} \times 100[\%]. \qquad (20)$$

4. Classification of proteins using PCA and SVM analysis of their FTIR spectra

Figure 2 depicts the FTIR spectra from dried MAB to CA125, BSA, human plasma, MAB to ILGF2, MAB to Leptin and MAB to OPN. All spectra exhibit a similar structure. The origin of the peaks has been well documented in the literature (reviewed by Barth & Zscherp, 2002). The spectra have several distinctive regions. The first region corresponds to the interval 2800-3500 cm^{-1}. A NH_2 region around 3200 cm^{-1} is strongly overlapped with OH stretching band. The region 1800-2700 cm^{-1} is relatively free of peaks. Amide bands are characteristics in the region 1200-1700 cm^{-1}. Those arise from the amide bonds that link the amino acids. The amide I centered about 1740 cm^{-1} corresponds to the stretching mode of the C=O bond of the amide. It may have some contribution from CN stretching and CCN deformations. The amide peak II centered around 1550 cm^{-1} corresponds to the bending mode of the NH bond of the amide with contributions from C=O in plane bending and NC stretching. Amide III mode is the in-phase combination of NH in-plane bending and CN stretching. Other smaller peaks corresponding to CC stretching and CO bending are observed in this region. The characteristics of these peaks provide information about the

secondary structure of the proteins since the hydrogen bonds that establish this structure, are mostly associated to the CO and NH bonds. The wide peak in the region 400-800 cm^{-1} corresponds to librations with contribution from other rotational and low energy vibrational lines. In figure 3 we show the results of the use of the first two PCA variables to represent the data presented in figure 2. Despite the evident similarities between the spectra the data are perfectly separable even with the use of only the first two PCA variables. For MAB to Leptin, ILGF2 and OPN the separation is larger.

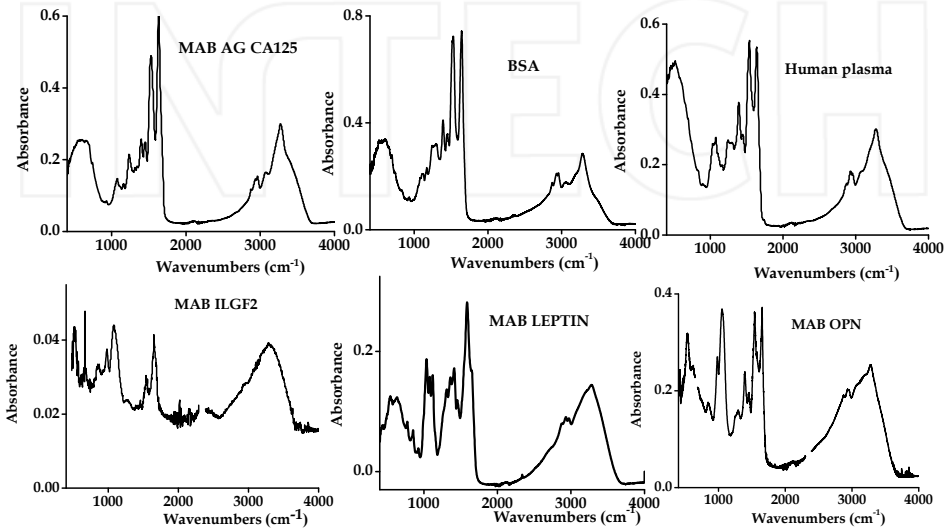


Fig. 2. FTIR spectra of MAB to AG CA125, BSA, human plasma, MAB to ILGF2, MAB to Leptin and MAB OPN

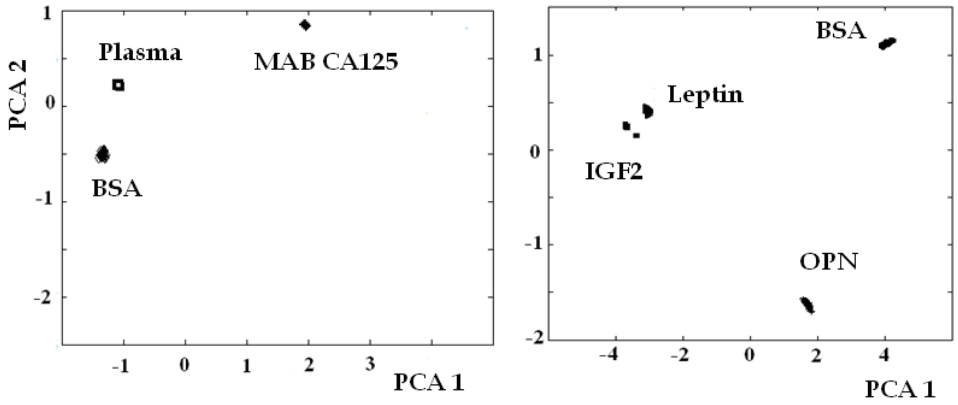


Fig. 3. Two-dimensional PCA of the data presented in figure 2

From figure 4 we can see that even the first two principal components are sufficient to achieve perfect separation of protein classes. Therefore, even the application of linear SVM is able to provide perfect accuracy on both training and test sets (100% accuracy). The result does not depend on the number k of principal components used (k>1). The average number of support vectors per class is relatively small (Figure 4) and practically does not depend on the number of principal components used, which indicates good generalization and stability of the proposed technique. The results demonstrate the possibility of automatic classification of proteins using PCA and linear SVM with accuracy of nearly 100%. Hence, this justifies the application of the conceptually simpler LDA technique. Namely, LDA is also capable of achieving 100% accuracy using as little as 2 principal components. Hence, below we discuss the use of LDA to separate the FTIR spectra of proteins and their deuterated versions aimed at the development of a FTIR based immunoassay.

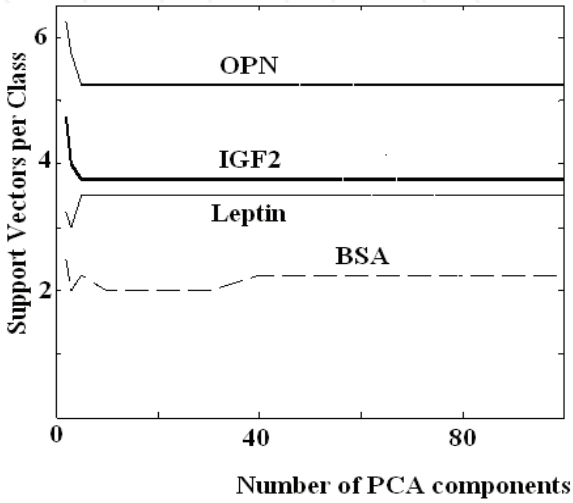


Fig. 4. Support vector per class as a function of the number of PCA components after Lazarevic et al. (2009). Reproduction authorized by the International Society for Optical Engineering SPIE

5. FTIR of deuterated proteins

Deuterium is a stable isotope and can be used as a labeling agent. Deuteration occurs by simple dilution of proteins in heavy water that contains the deuterium isotope of hydrogen (^2H). Hydrogen atoms on the surface of the protein are exposed to a fast exchange with deuterium atoms while hydrogen atoms deeply buried within the protein molecule exchange at a low pace. As an effect by substituting hydrogen atoms by deuterium atoms vibration modes of OH (hydroxyl peaks), NH_2 (amide peaks), NH and/or CH can be affected. Deuteration also induces the appearance of a strong peak in the region around 2400 cm^{-1}. This region is usually free of peaks for most of the proteins. Besides its evident advantages and extensive use for the study of the secondary structure deuteration of proteins can have another important application still not considered in the

literature. Indeed, the FTIR spectrum of a deuterated protein is different from the non-deuterated one, and this can be used for their identification within a matrix containing a large variety of proteins, e.g., a blood, plasma or serum sample. Furthermore, as we demonstrate, the use of PCA and LDA can identify these differences automatically with high accuracy.

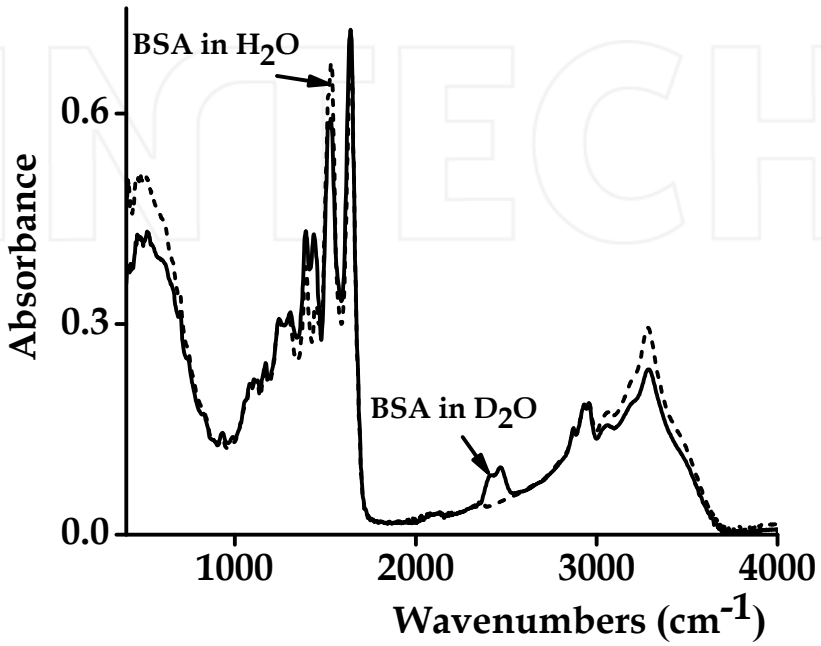


Fig. 5. FTIR spectra of non-deuterated BSA and partially deuterated BSA

As it is well known, the spectrum of water exhibits the stretching symmetric and antisymmetric modes at 3253 cm^{-1} and 3315 cm^{-1}, respectively, the bending mode at 1647 cm^{-1} and libration peaks observed below 700 cm^{-1}. The spectrum also exhibits a small peak at 2094 cm^{-1} that corresponds to the interaction between bending and libration modes. The heavy water spectrum exhibits a similar structure of peaks but all the peaks are shifted by a factor of 1.37 in close correspondence to the factor of 1.41 calculated from the differences of masses between hydrogen and deuterium atoms. The spectral widths of the D_2O lines are also reduced by a factor of 1.35 comparing to those of water. Of special interest are the peaks of the stretching vibration of deuterium oxide molecule which are centered at 2401cm^{-1} and 2471 cm^{-1}. Proteins FTIR spectra are usually free of peaks in this area (see figure 2). Deuterated proteins can have peaks in this region, a feature that can be used for protein identification and calibration. As a consequence, shifts of spectral lines, changes in relative amplitudes and changes in the spectral widths are expected for deuterated proteins (Marcano et al., 2008).

In figure 5 we show the FTIR spectra of the dried BSA sample from solutions of distilled double dionized water (dash line) and D_2O (solid line) prepared at concentration of 42 μg/mL. The amide I peak (1644 cm^{-1}) remains almost unaltered while the amplitude of the amide II peak (1530 cm^{-1}) decreases. This peak corresponds to NH bending vibrations which are strongly affected by substitution of hydrogen by deuterium atoms. Amplitude increase is observed for other peaks in the region 1200-1300 cm^{-1}. Remarkable is the presence of the peaks in the region 2400-2500 cm^{-1} which is free of peaks not only for BSA but also for a number of antibody proteins (see figure 2). These peaks correspond to stretching OD vibration. Correspondingly, the hydroxyl peaks in the region 3000-3500 cm^{-1} are reduced.

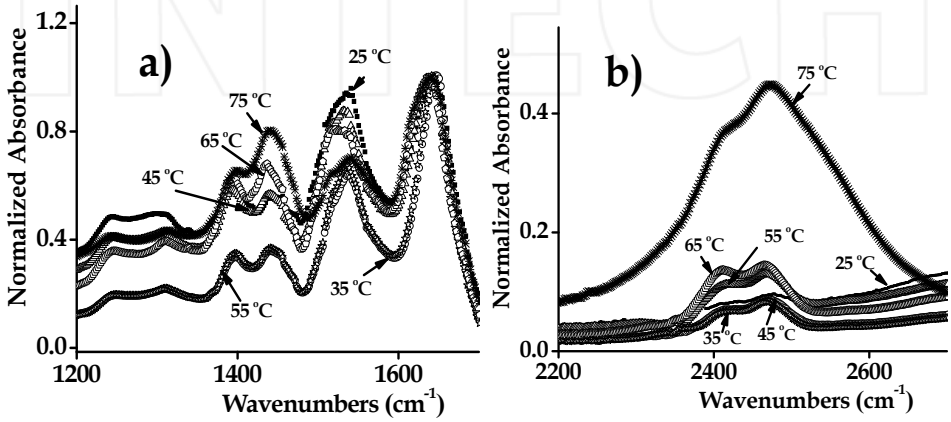


Fig. 6. Details of the FTIR spectra of BSA as a function of temperature in the regions 1200-1700 cm^{-1} and 2200-2600 cm^{-1}

Increase in temperature deepens the changes observed. In figure 6 we show details of the FTIR spectra of BSA in the regions 1200-1700 cm^{-1} (a) and 2200-2600 cm^{-1} (b), respectively, after heating the solution from 25 °C up to 75 °C. High temperature breaks the hydrogen bonds opening the protein molecule and exposing it to wide deuteration. The effect is small up to a certain temperature (60 °C in our case) but when the thermal energy is enough to break the hydrogen bonds the effect increases substantially. For the sample heated up to 75 °C the peaks at 2400 cm^{-1} are more than 5 times larger than the one for 55 °C (see figure 6b). The amide II peak is depleted as well as other peaks at 1200 cm^{-1}(see figure 6a). The depletion is also remarkable for the hydroxyl peaks at 3500 cm^{-1}. Deuteration can be significantly increased by the use of ultrasound. Ultrasound shakes the molecule exposing its hydrogen bonds to deuteration. In figure 7 we show the results of deuteration of BSA in D_2O at concentration of 500 mg/mL by using ultrasound during 120 minutes at room temperature. After exposing to ultrasound the solution is put to rest for long term dilution (1 week) at 6 °C. The changes in the absorbance FTIR are remarkable. The peak at 2466 cm^{-1} dominates the center of the spectrum. The amide II peaks is almost depleted and the peak at 1434 cm^{-1} triples its amplitude value.

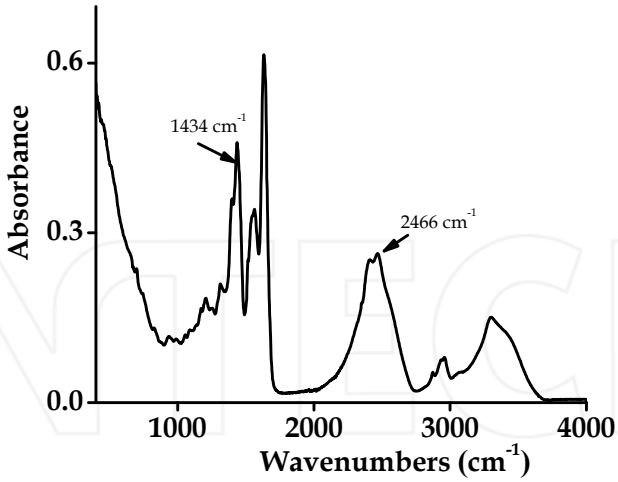


Fig. 7. FTIR spectrum of highly deuterated BSA obtained using ultrasound

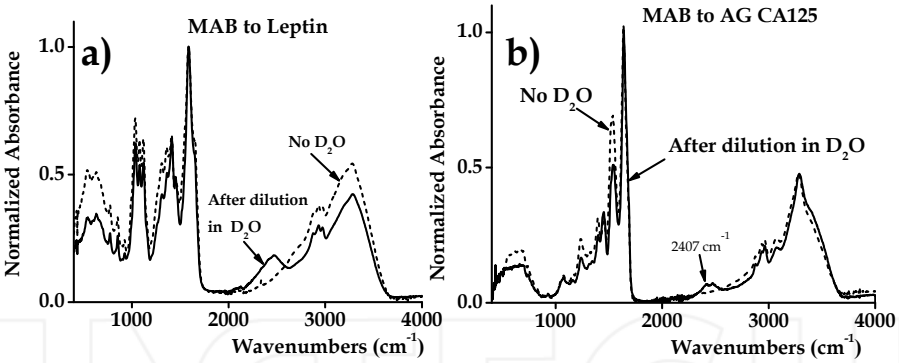


Fig. 8. FTIR spectra of deuterated MAB to Leptin and MAB to AG CA125

We have studied the effects of deuteration on MAB to Leptin, MAB to AG CA125, AG CA125, Leptin, OPN and IGF2. All these proteins are originally diluted in saline solution by the supplier of chemicals. For deuteration we use 5 μL of the sample and diluted it into 5 μL of heavy water. One drop of this diluted solution is then deposited to dry over the spectrometer. In figure 8 we show the FTIR spectra of MAB to Leptin (8a) and MAB to AG CA125 (8b) from the original saline solution (dot lines) and heavy water dilution (solid lines). The spectra are normalized with respect to the amplitude of the amide I peak. Again we observe the surge of the DO peak in the region around 2400 cm^{-1} and also changes in the relative amplitudes of hydroxyl and amide II peaks. Reduction in the spectral width of the peak is also observed.

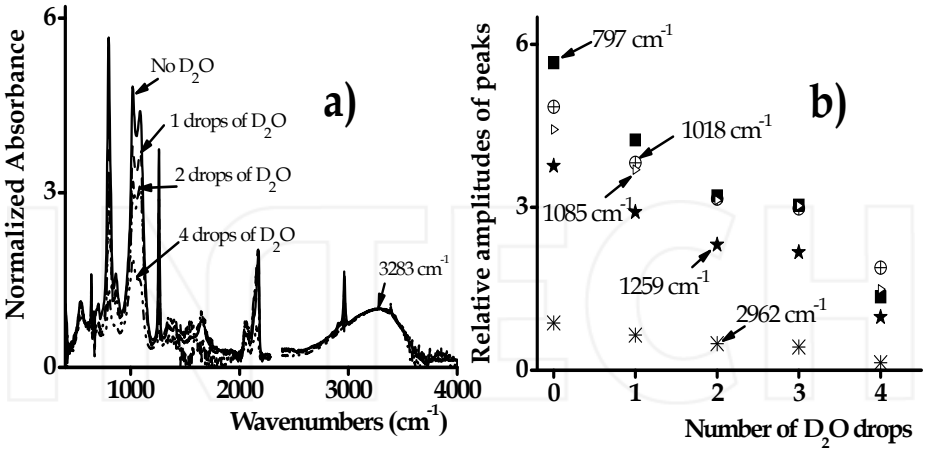


Fig. 9. FTIR spectra of AG CA125 exposed to different amount of D_2O

We have also studied the effect of deuteration of ovarian cancer AG CA125, OPN, IGF2 and leptin. In figure 9 we show the effect of deuteration over the ovarian cancer AG CA125 after adding drops of D_2O (figure 9a) subsequently. The sample is left to dry after addition of each drop and before recording the spectra. We observe decrease in the relative amplitude of several peaks all over the spectrum. In figure 9b we plot the relative amplitudes of five of these peaks as a function of the amount of D_2O used. Similar results are obtained for leptin, OPN and IGF2.

Figures 5-9 demonstrate that deuteration of proteins is relatively easy to achieve by simple dilution in heavy water for both the monoclonal antibody proteins and their corresponding antigens. If required the impact of deuteration can be increased by increasing the temperature or by applying ultrasound.

6. Use of PCA method for detection of deuterated proteins

As expected, using PCA to perform feature extraction can lead to distinguish with high efficiency between deuterated and undeuterated versions of the same protein. In figure 10 we plot the tridimensional principal projection of the FTIR data from AG Leptin exposed to 1, 2 and 3 drops of D_2O. A good separation between the data is obtained. The absolute distance between the data increases with the number of D_2O drops as correspond to larger deuteration effect as suggested by figure 10. However, the effect depends on the type of protein. In figure 11 we show the result for BSA. In this case, the distance between the unexposed sample PCA data and the exposed ones does not change monotonically with the numbers of D_2O drops. This may be related to parasitic D-H exchange with water of the surrounding the sample atmosphere over the time of the experiment. Nevertheless, a very clear separation between BSA samples with different numbers of D_2O drops is achieved with 100% classification accuracy using LDA with four-fold cross-validation.

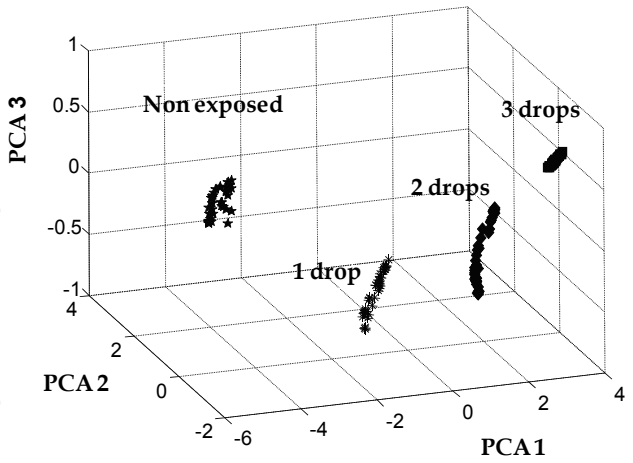


Fig. 10. Tridimensional PCA representation of the FTIR data from AG Leptin exposed to successive drops of D_2O. Non-exposed to D_2O data are included for comparison

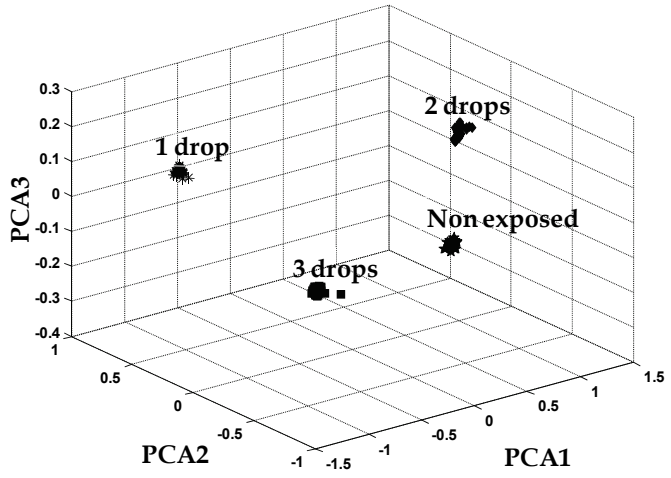


Fig. 11. Tridimensional PCA of FTIR spectra of BSA exposed to 1, 2 and 3 drops of D_2O. The non-exposed to D_2O samples is included for comparison

Different regions of the spectra contribute differently to the separation. In figure 12 we show the plot of the PCA loadings (absolute value of components for eigenvectors $\mathbf{v}_1$-$\mathbf{v}_3$ from Eq. 2 as functions of the wave-numbers. The larger the absolute value of the PCA loading is the larger is the importance of the corresponding spectral region for a more efficient separation.

In the figure we see the importance of the low wavenumbers region (400-600 cm^{-1}), the amide peaks region (1300-1600 cm^{-1}) and the DO peak region (2400-2500 cm^{-1}).

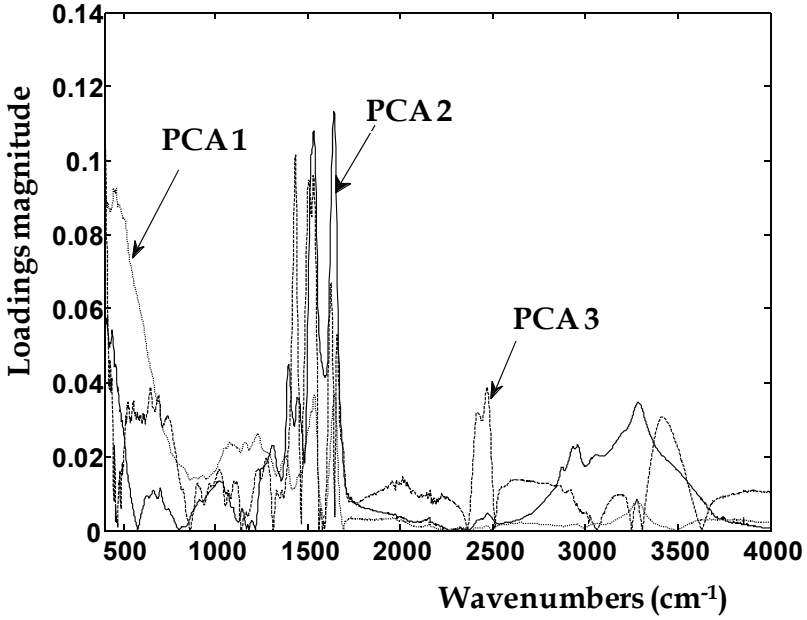


Fig. 12. PCA variables as functions of the wavenumbers corresponding to the data of figure 11

7. Detecting a protein antibody in a complex matrix

The proposed methods can be used to detect the presence of a particular protein in a complex matrix, such as blood, plasma, serum or other biomedical samples. The efficiency of separation can be increased by using deuterated versions of the proteins. In figure 13a we show the FTIR spectra of pure human plasma (stars), plasma with added non-deuterated (diamonds) and deuterated (squares) MAB to CA125 at a concentration of 50 μg/ml normalized by the amplitude of the peak at 3275 cm^{-1}. The differences are more remarkable for the deuterated samples comparing to the non deuterated ones. In figure 13b we show the tridimensional PCA plot corresponding to these data. Despite the similarities of spectra, the data are separable in the PCA coordinates space. The separation is larger for the deuterated version of the protein. The absolute distance between the centers of data cluster in the PCA coordinates space corresponding to plasma and the cluster corresponding to deuterated MAB to CA125 is more than twice larger than the distance of the center of the plasma cluster to the non-deuterated protein data cluster. Although the concentration used is relatively high, the result demonstrates possibilities of detection of proteins embedded in a complex matrix and the increase in sensitivity when using deuterated versions of the proteins.

The spectral changes in deuterated proteins and the statistical and data mining methods used for their analysis can be applied to develop new kinds of immunoassays for detection of antigen proteins. The immunoassays are aimed at detection of a particular antigen protein embedded in a complex matrix, such as blood, serum or a plasma sample. Figure 14 describes two such immunoassays. Figure 14a depicts an immunoassay where non-deuterated protein antibodies are deposited over a glass substrate (step 1). The plate is then exposed to the sample. The antibody proteins on the plate capture their corresponding antigens from the sample (step 2). Finally, the system is exposed to the presence of the deuterated versions of the antibody proteins. These deuterated antibodies are then attached to the trapped antigens forming a sandwich structure which is wash away to remove non captured proteins (step 3). This sandwiched structure can then be analyzed using an FTIR spectrophotometer. In a second type of immunoassay the first step is the same (see figure 14b). Then, the plate is exposed to the sample which has been previously diluted in heavy water. Deuterated antigens can then be captured by their antibody protein deposited on a plate. The rest of the sample can be washed away. Finally, the presence of the antigens can be detected by performing the FTIR experiment over the treated plate.

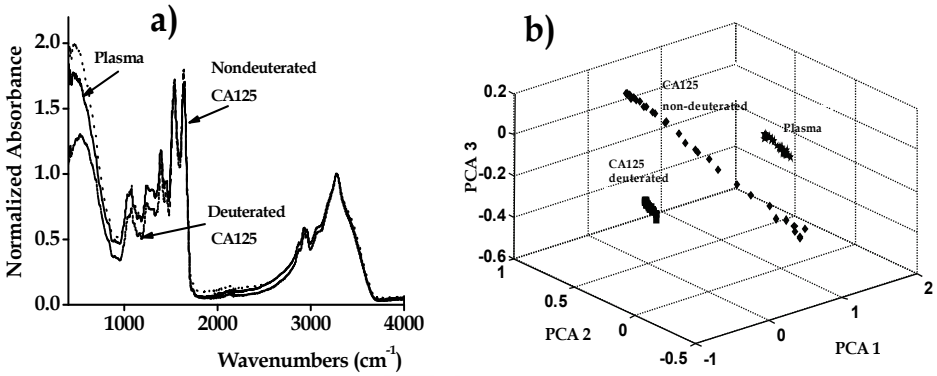


Fig. 13. a) Normalized FTIR spectra of plasma and plasma containing deuterated and non-deuterated MAG AG CA125, b)Tridimensional PCA plot of FTIR data from plasma and deuterated and non-deuterated versions of the protein MAB to CA125

Several steps need to be completed before developing a FTIR-based immunoassay of practical use. Deuteration can affect the bioactivity of the proteins. In this regard, the affinity constant between the antibody and antigen can depend on the level of deuteration. The level of deuteration can be also affected by parasitic D-H exchanges that can mask the real results. Practical comparison with well established immunoassays such as ELISA must be completed. However, we show that the use of deuteration techniques in combination with statistical methods, such as PCA, LDA and SVM, will play a crucial role in the developing of this new kind of FTIR-based immunoassays aimed at detection of a targeted protein in a complex biosample.

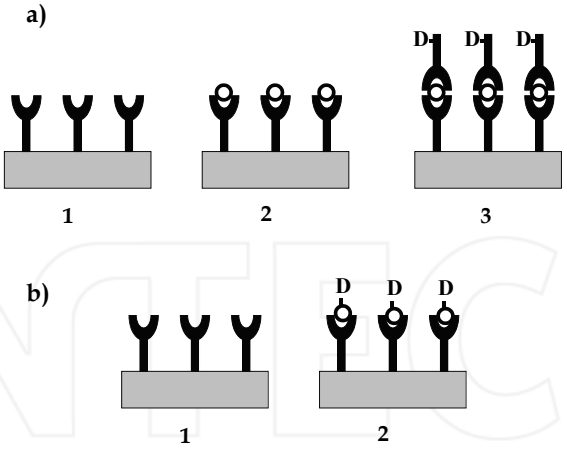


Fig. 14. Proposed FTIR based immunoassays for detection of a given antigen

8. Conclusions

In this study, we demonstrate that using combination of PCA analysis and statistical and data mining classification techniques (LDA and SVM) it is possible to automatically determine a class of the FTIR sample of an unknown protein using a small number of principal components. In such a case, using conceptually simpler LDA analysis (that relies on stronger assumptions about the data than SVM) is justified. The full advantage of non-linear SVM could, however, be expected in case of more complex and noisy spectroscopic data. The proposed data analysis technique is computationally fast and can in principle be applied in on-line learning classification framework. Work in progress includes testing the proposed technique on larger datasets to exclude the small variability of samples (the sample bias) as a potential reason for extremely high classification accuracy. We show that the techniques distinguish between different proteins with similar FTIR spectra and between deuterated and non-deuterated versions of the same protein. Furthermore, we demonstrate the use of the method for separation and identification of proteins embedded in a complex matrix of proteins such as plasma. We show that deuteration increases the sensitivity of the method. Finally, we propose an immunoassay that is aimed to utilize the demonstrated sensitivity of the methodology to detect a particular antigen protein in a complex biosample.

9. Acknowledgements

This research has been possible thanks to the support of the National Science Foundation (NSF-CREST grant Nº 0630388 and NSF-MRI grant Nº 0922587), National Institute of Health (NIH Grant Nº P20 RR016472), Department of Defense (DoD/DoA Grants Nº 45395-MA-ISP, and Nº 54412-CI-ISP) and of the National Aeronautics and Space Administration (NASA URC 5 grant Nº NNX09AU90A). We also would like to thank Blood Bank of Delmarva, Delaware, for providing the human plasma samples.

10. References

Arrondo, J. L. R.; Muga, A.; Castresana, J. & Goñi, F. M. (1993). Quantitative studies of the structure of proteins in solution by Fourier-transform infrared spectroscopy. *Prog. Biophys. Molec. Biol*., Vol. 59, No. 1, 23–56, ISSN 0079-6107.

Baenziger, J. E. & Methot, N. (1995). Fourier transform infrared and hydrogen/deuterium exchange reveal an exchange-resistant core of alpha-helical peptide hydrogens in the nicotinic acetylcholine receptor. *J. Biol. Chem*., Vol. 270, No. 49, 29129-29137, ISSN 0021-9258.

Barth, A. & Zscherp, C. (2002). What vibrations tell us about proteins. *Quarterly Reviews of Biophysics,* Vol. 35, No. 4, 369–430. Cambridge University Press, DOI: 10.1017/S0033583502003815. ISSN 0033-5835.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer + Business Media LLC, ISBN-10 0-38731073-8, Singapore.

Bramer, C. M. (2007), *Principles of Data Mining (Undergraduate Topics in Computer Science)*, Springer-Verlag, ISBN-10: 1-84628-765-0, London.

Byler, D. M. & Susi, H. (1989). Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers,* Vol. 25, 469-487, ISSN 1097-0282.

Dave, N.; Lórenz-Fonfrı□a, V. A.; Villaverde J.; Lemonnier R.; Leblanc G. & Padrós, E. (2002). Study of Amide-proton Exchange of *Escherichia coli* Melibiose Permease by Attenuated Total Reflection-Fourier Transform Infrared Spectroscopy: Evidence of structure modulation by substrate binding. *J. Biol. Chem*. Vol. 277, 3380-3387, ISSN 0021-9258.

Fabian H. & Shultz, C. P. (2001). *Encyclopedia of Analytical Chemistry,* R. A. Meyers (Ed.), Wiley, 5779-5803. ISBN: 978-0-471-97670-7, Chichester.

Fabian H. & Mantele W. (2002). *Handbook of Vibrational Spectroscopy*, J. M. Chalmers, P. R. Griffiths (Eds.) Wiley, 3399-3425, ISBN: 978-0-471-98847-2, Chichester

Goormaghtigh, E.; Cabiaux, V. & Ruysschaert, J. M. (1994). Determination of soluble and membrane protein structure by Fourier transform infrared spectroscopy I. Assignments and mode compounds. *Subcell. Biochem.* Vol. 23, 329–362, ISSN -0306-0225.

Haris, P. I. & Chapman, D. (1994). Analysis of polypeptide and protein structures using Fourier transform infrared spectroscopy. *Methods in Molecular Biology, Microscopy, Optical Spectroscopy, and Macroscopic Techniques*, vol. 22 (eds. C. Jones, B. Mulloy & A.H. Thomas), 183–202. Humana Press Inc. ISBN 0-89603-232-9, Totowa, NJ.

Hsu, C. W & Lin, C. J. (2002). A comparison of methods for multi-class support vector machines. *IEEE Trans. Neural Networks*, vol. 13, No. 2, 415-425, ISSN 1045-9227.

Hybl, J. D.; Lithgow, G. A. & Buckley, S. G. (2003). Laser-induced breakdown spectroscopy detection and classification of biological aerosols. Appl. *Spectros*. Vol. 57, 1207-1215.

Jiang, Z. Q.; Fu, H. G. & Li, L. J. (2005). Support vector machine for mechanical faults classification. *Journal of Zhejiang University Science* (Online) 6A(5). http://www.zju.edu.cn/jzus/2005/A0505/A050513.pdf.

Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer Verlag, ISBN 0-387-95442-2, New York.

Karush, W. (1939). *Minima of Functions of Several Variables with Inequalities as Side Constraints*. M.Sc. Dissertation. Dept. of Mathematics, Univ. of Chicago, Chicago, Illinois. Available from http://wwwlib.umi.com/dxweb/details?doc_no=7371591.

Krzanowski, W. J., 1988). *Principles of Multivariable Analysis: A User's Perspective*. Oxford University Press, ISBN 0 198 8507089, New York.

Lazarevic, A.; Pokrajac, D.; Marcano, A. & Melikechi, N. (2009). *Support vector machine based classification of fast Fourier transform spectroscopy of proteins*, Proc. SPIE, Vol. 7169, 71690C; DOI: 10.1117/12.809964, ISBN 9780819474155, San Jose, California, USA, January 2009, SPIE, Bellingham.

Massart, D. L.; Vandenginste, B. G.; Deming, S. N.; Michotte, Y. & Kaufman, L. (2003). *Chemometrics: A Textbook*. Elsevier, ISBN 0-444-42660-4, Amsterdam.

Marcano A.; Markushin, Y.; Melikechi, N. & D. Connolly (2008). Fourier Transform Spectroscopy of Deuterated Proteins. *Linear and Nonlinear Optics of Organic Materials VII by Rachel Jakubiak, Proc. SPIE*, Vol. 7049, pp. 70490z-1-8. ISBN 9780819472694, San Diego, California, USA, August 2008, SPIE Bellingham.

Melikechi, N.; Ding, H.; Rock, S.; Marcano, A. & Connolly, D. (2008). Laser-induced breakdown spectroscopy of whole blood and other Liquid organic compounds. *Optical Diagnostic and Sensing VIII, Editors G. Cote and A. V. Priezzhev, Proc. SPIE*, Vol. 6863, 68630O1-7, DOI:10.1117/12.761901, ISBN 9780819470386, San Jose, California, USA, January 2008, SPIE, Bellingham.

Mor, G.; Visintin, I.; Lai, Y., Zhao, H.; Schwartz, P. Rutherford, T.; Yue, L.; Bray-Ward, P. & Ward D. C. (2005). Serum protein markers for early detection of ovarian cancer. *PNAS*, Vol. 102, No. 21, 7677-7682, on line ISSN 1091-6490.

Nie, B.; Stutzman, J. & Xie, A. (2005). A vibration spectra maker probing the hydrogen-bonding status of protonated Asp and Glu residues. *Biophys. J.,* Vol. 88, No 4, 2833-2847, ISSN 0006-3495.

Petibois, C.; Gionnet, K.; Goncalves, M.; Perromat, A.; Moenner, M. & Deleris, G. (2006). Analytical performances of FT-IR spectrometry and imaging for concentration measurement within biological fluids, cells, and tissues. *Analysis*, Vol. 131, 640-647, ISSN 0003-2638.

Platt, J.; Cristianini, N. & Shawe-Taylor, J. (2000). Large Margin DAGs for Multiclass Classification. *Advances in Neural Information Processing Systems,* Vol. 12, pp. 547-553, ISBN-13 978-0-262-56145-7.

Schorge, J.O.; Drake, R. D.; Lee, H.; Skates, S. J.; Rajanbabu, R.; Miller, D. S.; Kim, J. H.; Cramer, D. W.; Berkowitz, R. S. & Mok, S. C. (2004). Osteopontin as an adjunct to CA125 in detecting recurring ovarian cancer. *Clin. Cancer Res.*, Vol. 10, No. 10, 3474-3478, ISSN 1557-3265.

Siebert, F. (1995). Infrared spectroscopy applied to biochemical and biological problems. *Methods Enzymol.*, Vol. 246, 501–526, ISBN 978-0-12-182147-0.

Sutphen, R.; Xu, Y.; Wilbanks, G. D.; Fiorica, J.; Grendys Jr., E. C.; LaPolla, J. P.; Arango, H.; Hoffman, M.S.; Martino, M.; Wakeley, K.; Griffin, D.; Blanco, R. W.; Cantor, A. B.; Xiao, Y. J. & Krischer, J. P. (2004). Lysophospholipids are potential biomarkers of ovarian cancer. *Cancer Epidemiol. Biomarkers Prev*. Vol. 13, No. 7, 1185-1191, ISSN 1055-9965.

Vapnik, V. (2000). *The Nature of Statistical Learning Theory*, Springer Verlag, ISBN 0-397-98780-0, New York.

9

The Analysis of Heart Sounds and a Pocket Computer Application via Discrete Fourier Transform

Gür Emre Güraksın[1], Uçman Ergün[2] and Ömer Deperlioğlu[2]
[1]*Afyon Kocatepe University, Engineering Faculty*
Computer Engineering Department
[2]*Afyon Kocatepe University, Engineering Faculty*
Biomedical Engineering Department
Afyonkarahisar
Turkey

1. Introduction

The heart is one of the two crucial centers for human life. Thus, any disorder concerning the heart which may be able to occur is of prime importance to human health. The mortality rate stemming from heart diseases took the second lead after brain embolism in the world from 1985 to 2006 (Jiang & Choi, 2006). Being such significant to humans, the heart organ is consisted of two cycles. At the moment of the closure of mitral and tricuspid valves, ventricular contraction comes out, alias one cycle begins with systole and ends with diastole (Sharif et al., 2000). Auscultation with stethoscope is a preferential method that the doctors use in order to differentiate normal cardiac systems from the abnormal ones that come out (Sinha et al., 2007). The listened heart sounds are formed through the flow of blood entering and exiting the heart and with the movements of cardiac valves connected to this flow. The sounds comprised of this blood flow are listened by the physicians via stethoscope. The heart sounds are interpreted and it is determined whether the patient has any disorder about the heart. On the other hand, auscultation method has some limitations. Auscultation depends on the physician's interpretation of different heart sounds, hearing skill, experience and expertize (Kandaswamy et al., 2004). The required experience and expertize are achieved as a result of long examinations and diagnoses made by the physicians. Even though one physician has the necessary training so as to conduct auscultation and to diagnose cardiac disorders, he is still in need of clinical experience. Lacking experience and expertize are especially a difficulty for the newly graduates and medical interns. However, unsuitable conditions of the environment and incompatibility of the patient can also lead to deficiency in the diagnosis process. Because of these difficulties which may be experienced, auscultation method that is listening with a stethoscope has been insufficient in the exploration of heart abnormalities.

As the auscultation method fail to meet the needs of the physicians, they also make use of Electrocardiography (ECG) data along with stethoscope. ECG method, generated through the advancement of the technology, is a procedure helping physicians in the diagnosis of the

heart disorders. ECG is the wave form taking the record of the electrical activity of the heart via electrodes attached to the skin. A simple and cheaply method, ECG is constantly used by the physicians. ECG records and their analysis that are used to detect defects in the heart are relatively a good method. On the other hand, if the likely generating system of heart is utterly acting with small heart defects, it may look very difficult to diagnose these abnormalities with the analysis of ECG records as nearly no change in the ECG will be discovered. However, it was determined that this situation has almost led to changes among the sounds produced by the heart (Sinha et al., 2007). As it is not possible to detect some heart abnormalities with ECG, listening to heart sounds with auscultation method has earned more importance. Owing to the reasons explained above, there is a need for faster and more accurate diagnosis in the record and analysis of these heart sounds.

The very first phase of developing the system that will help the physicians interpret the heart sounds accurately is the signal processing methods. When the studies based on these methods are searched, it can be seen that Fourier analysis has been used in many of the studies. Thanks to the Fourier analysis, it is possible to examine sound signals in the frequency space. In studies especially done on the classification of heart sounds, the frequency analysis of these digitized heart sounds are processed and then passed to classification stage with various artificial intelligence methods. If we survey some of the studies in literature that used heart sounds and Fourier transform together, in one study Segaier et al. developed a digital algorithm in order to detect the first heart sound (S1) and the second heart sound (S2) and to characterize systolic murmurs. The study done on the pediatry patients, Short-Time Fourier Transform (STFT) was used so as to carry out the analysis of the heart sounds taken from the patients (El-Sagaier et al., 2005). In another study, Folland et al., in the moment of auscultation, applied Fast Fourier Transform (FFT) and Levinson-Durbin algorithm to heart sounds to analyse the abnormalities in the heart sounds; and applied the data to the Multilayer Perceptron (MLP) and Radial Basis Function (RBF) artificial neural networks to classify abnormal sounds. Eventually, it was determined that in the classification of the heart sounds, the sensitivity levels that the MLP and RBF neural networks have gained were 84% and 88% consecutively (Folland et al., 2002). In another study, Debbal ve Breksi-Reguig have done the time-frequency analysis of S1 and S2 heart sounds, and applied Wigner distribution and wavelet transform techniques to the heart sounds. They made a comparison between FFT and STFT and with these techniques they implemented (Debbal & Breksi-Reguing 2007). In one of the study by Abdel-Alim et al., they made a classification of heart sounds that belonged to different cardiac valve disorders with the use of a feed forward artificial neural network. In the course of analysis of these sounds, discrete wavelet transform, FFT, and linear prediction coding methods were employed. In the end, the classification success was found out to be %95.7 (Abdel-Alim et al., 2002).

In this study, the heart sounds achieved through a stethoscope were initially computed, and they were subjected to Discrete Fourier Transform (DFT) and next the graphics and the frequency spectrum in the time domain that belonged to the heart sounds were drawn on the pocket computer. Thus, frequency spectrum of normal and abnormal heart sounds was obtained via DFT, so it was aimed to prepare the physicians with some more data in the course of auscultation.

2. Structure of the heart and heart sounds

Heart is an empty muscle that pumps the blood to the blood vessels in the whole body (Sharif et al., 2000). The most significant and primary duty of the heart is to dispatch and pump the blood to the circulation system like a pump (Ahlström, 2006).

As it can be seen in figure 1, the heart is composed of two parts: the right heart and the left heart. The right heart pumps the blood into lungs. This cycle is called as pulmonary cycle. On the other side, the left heart is the part that provides all the organs and the whole body with the oxygen and nutrients (Sharif et al., 2000). In addition, there are four chambers in the heart that are known as the right and the left atriums and ventricles. These two atriums are the places where the blood entering the heart is stored. The ventricles, on the other hand, convey the blood to the whole body like a pump. As the heart contracts, the blood makes a pressure towards the valve and moves from the atrium to the ventricle (Barschdorff et al., 1990).

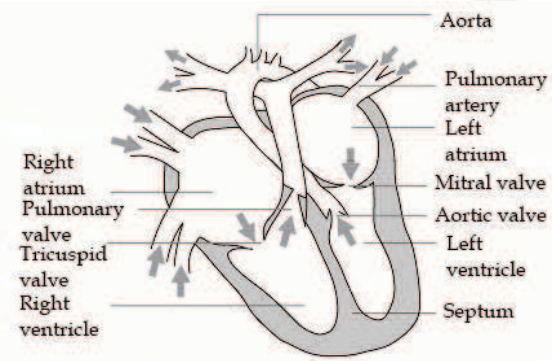


Fig. 1. Structure of the Heart

The relation among the volume, pressure and flow of the blood in the heart determines the opening and the closing of the valves. Normal heart sounds occur in the course they close. Besides, the sounds occurring in the heart and in the vessels with the flow are constituent of the heart sounds, too. But, as a matter of fact, how they occur is still a matter of discussion (Ahlström, 2006).

The abnormalities in the heart structure mostly reflect to the sounds that the heart produce (Leung et al., 2000). The formation of heart sounds and murmurs are generally developed through the actions of myocardial walls, the opening and closing of cardiac valves, and the blood flow into and out of the ventricles (Kemaloğlu & Kara, 2002).

While the sinoatrial node contracts itself and the contraction spreads to the atria, the pressure of left atrium surpasses the pressure of left ventricle. This contraction is then extended to the ventricles. This moment is the time when ventricle starts to depolarize. The contraction spreading into the ventricle contracts the muscles of the ventricle and leads to a contraction in the ventricle. The pressure in the left ventricle starts to rise and as soon as it reaches to the pressure of the left valve, the atrium and the valve ventricles close. At that moment, S1 comes out. Mitral valve normally closes earlier than the triscuspid valve. For that reason, S1 has two elements namely mitral and triscuspid. The frequency band is 20–45 Hz and the period is 50–100 ms.

The ventricle pressure goes on rising, and as the pressure goes over the pressure of aorta, firstly aorta valve and next pulmonary valve open. Soon after starts the period of sending the blood out of ventricles. As the ventricle muscles relax, the ventricle pressure starts to decrease. At the moment that the internal pressure in the ventricle goes below the aorta pressure, the aorta valve closes. The pulmonary valve closes respectively. The closure of these two valves forms the S2 sound. The frequency band is 50–70 Hz and the period is 25–50 ms.

When the internal pressure of the ventricle drops below the pressure of atrium, mitral and triscuspid valves open and the ventricles are filled with blood. As the ventricles are filled with blood, the vibrations of the ventricle muscles form the third heart sound (S3). It could be heard in the young while it is an indicator of disorder of myocard function in the overaged. This sound arises almost after 150 ms after the closure sound of aorta.
In the last parts of blood fulfillment of the ventricles, the flow of blood that was again quickened with depolarization of the atrium revibrates the ventricle walls and in some pathologic cases, this causes the fourth heart sound (S4). S4 is normally not heard in the adults but could be taken in children.
Four sounds seen in figure 2 are known as the simple heart sounds. Apart from these sounds, some sounds as murmur may occur in some heart disorders. These sounds are in the frequency band of 100–600 Hz and are long-time compared to the simple heart sounds (Kemaloğlu & Kara 2002).

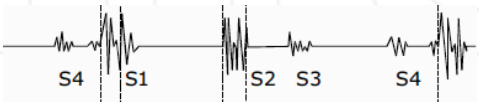


Fig. 2. The First, Second, Third and Fourth Heart Sounds

The sounds that are named as murmur and caused by the blood passing through the cardiovascular system loudly are the significant examples of abnormal sounds. The timing of the murmur and the level of height have remarkable significance concerning the situation of the heart. For instance; in the course of diastole a murmur marks erroneous functionality of the heart valve. On the other hand, in the course of systole, the murmurs may be related with the healthy and pathological heart depending on the acoustic character of the murmur (Ölmez & Dokur, 2003).
The murmurs are formed through irregular blood flow and as a result of narrowing and leaking valves or the existence of the abnormal passages in the heart**.** Eventually, the blood flow that comes out leads to steady and irregular vibrations that are transmitted from the cardiac and tissues belonging to the chest to surface of the chest (Ahlström, 2006).
While defining the heart sounds, we need to draw attention to the frequency, density and the quality of the sound. For that reason, it is a must to listen to the heart sounds carefully that are used as reference S1 and S2 and to determine the place accurately in the course of listening. These sounds experience the systole and diastole phases as the heart operates, and the sound differentiations in these places provide related information in connection with the disorders in the heart (Say, 2002).

3. Fourier transform

When the signals in the real world are searched, it could be said that the signals that are encountered practically are the time domain signals and the size measured is the function of the time. For that reason, the signal needs to be transferred to a different domain through an application of mathematical transform and some information from the constituents that represent the signal in the domain can be obtained about the signal. For example, the frequency spectrum (frequency constituents) of the signal is gained with Fourier transform. The information hidden in the time domain is brought out in the frequency domain (Say, 2002).

The Fourier display of the signals plays an extremely substantial role in signal processing in both continuous time and discrete time. Thus, it supplies a method to match the signals to a domain and makes it possible to study on them. Fourier transform allows for a distinctive way to interpret the signal and systems (Hayes, 1999).

Fourier transform is a method of analysis that was developed by Jean B. Joseph Fourier, a French physicist and a mathematician, in 1987 when he was studying on the Fourier's research about heat and dispersion and it plays an important part in the signal processing (Swanson, 2000). In the study that Fourier conducted in 1807, he said that it could be achieved through selecting and gathering the sinus and cosine waves among the continuous and periodical signals (Smith 1997).

The Fourier transform that is used to determine the frequency constituent of the raw signal in the time domain can be defined with the two equities as follows:

$$X(\omega) = \int_{-\infty}^{\infty} x(t).e^{-j\omega t}.dt \tag{1}$$

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega).e^{j\omega t}.d\omega \quad \omega = 2\pi f \tag{2}$$

With Fourier transform, the signal is extricated into complex exponential functions that have various frequencies. In the equations t, specifies time; ω, specifies angular frequency and f specifies the frequency. x, specifies the signal in the time domain, X specifies the signal in the frequency domain. In the equation 1 given above, the Fourier transform of $x(t)$ and in the equation 2, inverse Fourier transform are displayed.

When the equation 1 is viewed, it is seen that $x(t)$ signal is multiplied with an exponential term in a certain f frequency and the integral of this multiplication is taken from minus infinity to plus infinity all over the time. It should be taken into account that the effect of the f frequency constituent to the integration will be the same no matter when it may come out among these times. The integration result will not change whether the f frequency constituent comes out in the course of $t1$ or $t2$. Fourier transform only indicates whether a certain frequency constituent exists or not. Just a spectral constituent of a signal can be gained through Fourier transform (Say, 2002).

Today, given that, the signal processing is held by computer algorithm and the computers may work in limited length and with discrete signals, it is noticed that the Fourier transform to be used will be discrete time Fourier , in other words, DFT (Smith, 1997).

3.1 Discrete Fourier Transform

On the contrary to some arrays defined theoretically, the Fourier transform of real arrays may not be calculated. Thus, it is not convenient for the digital signals to use Fourier transform. That the frequency is displayed analogically and it requires infinite number samples are some of the basic reasons of this ineptitude.

Due to these difficulties, if the importance of Fourier transform in signal processing is taken into account, a more practical transform needs to be defined. The new transform defined as normal spaced N frequency point (Ωk) around a unit disc and N sample of $x(n)$ arrays is called as DFT (Kayran & Ekşioğlu, 2004).

DFT is in essence a kind of transform such as Fourier arrays transform and Fourier integral transform. The transform feature is very powerful for the time arrays, which also allows for inverse transform. As can be noticed from its name, it owns utterly similar characteristics with Fourier integral transform. It especially defines the spectrum of a time array (Cochran et al., 1967).

This transform that also allows inverse transform includes distinctive qualities. The principal quality is the equivalent of multiplication of two DFT in the time domain is the total convolution of arrays. In addition, many spectrum analysis methods are based on DFT (Kayran & Ekşioğlu, 2004).

DFT is defined through the equation 3 given below:

$$A_r = \sum_{k=0}^{N-1} X_k \exp(-2\pi jrk / N) \quad r = 0,...,N-1 \tag{3}$$

Here, A_r, symbolizes r^{th} coefficient of the DFT and X_k symbolizes k^{th} sample of a time arrays composed of N sample. X_k's may be complex numbers while A_r's are always complex numbers. The formation of the formula that is consistent with the notation given in the equation 3 is mostly shown with the formula given in the equation 4:

$$A_r = \sum_{k=0}^{N-1} (X_k) W^{rk} \quad r = 0,...,N-1 \tag{4}$$

$$W = \exp(-2\pi j / N) \tag{5}$$

Since X_k's are commonly the values of discrete time points of a function, r arrays are sometimes referred as the frequency of DFT. DFT, too, is termed as "discrete time, finite range Fourier transform".

If we take a look at the inverse transform of equation 4:

$$X_l = (1/N) \sum_{r=0}^{N-1} A_r W^{-rl} \quad l = 0,...,N-1 \tag{6}$$

This relation (equation 6) is termed as the inverse of DFT (Cochran et al., 1967).

In the figures between 3 and 7 below, the flow diagram of DFT algorithm that was taken as a basis in the programming of the pocket computer was given. The main module seen in figure 3 calls for four modules, which are used in the calculation process.

The next module, shown in figure 4, sets up the twiddle factor arrays, that is the computation and storage of the sample values of cosines and sinus over one cycle, with argument starting from zero with increment of *2Π/N*. This module is named as *twid_fac*. *i* variable, here, is used as loop counter. On the condition that *i* variable reaches *N* that is the data size, the loop is terminated. The *tfc* and *tfs* arrays whose sizes are equal to *N* are used to hold the sample values of sinus and cosines respectively.

The next module seen in figure 5 reads the real and imaginary parts of the complex input data, and in turn transfers them into *N* dimensioned *xr* and *xi* arrays. It is assumed that the real parts of the input data are stored before the imaginary parts in the input file. If the data is real, the data in the *xi* array is initialized to *0* and the data in the *xr* array is read. This module is called *in-put.*

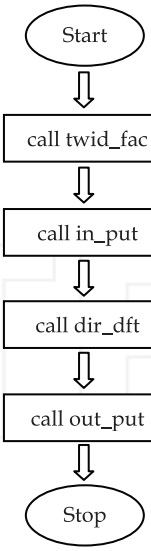


Fig. 3. The Main Module for Direct Implementation of the DFT

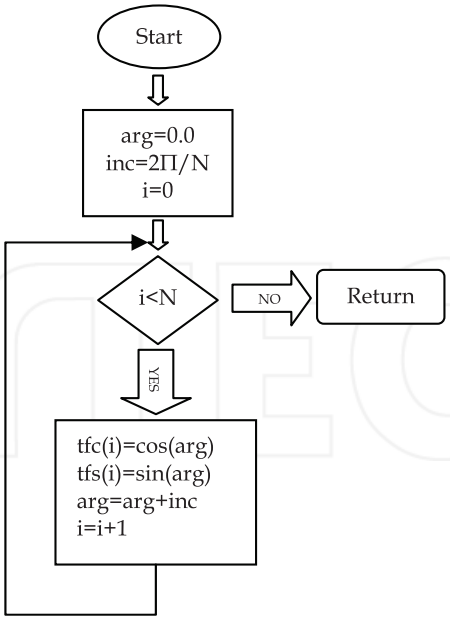


Fig. 4. The Twiddle Factor Module

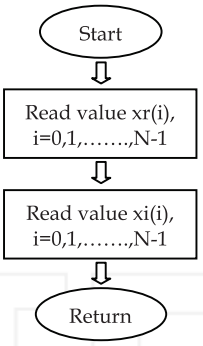


Fig. 5. The Input Module

The next module that is *dir_dft* module is given in figure 6. *dir_dft* module is used in the calculation of DFT coefficients. The calculation has been achieved via two nested loops. While the outer loop controls the frequency index, the inner loop controls the access of the data values. In each outer loop's iteration only one coefficient is calculated. The real and imaginary parts of the coefficients are stored, respectively, in arrays *XR* and *XI*, each of size *N*. The access of correct twiddle factor values is carried out using the *mod* function. Inside the inner loop, each coefficient is computed according to the DFT definition.

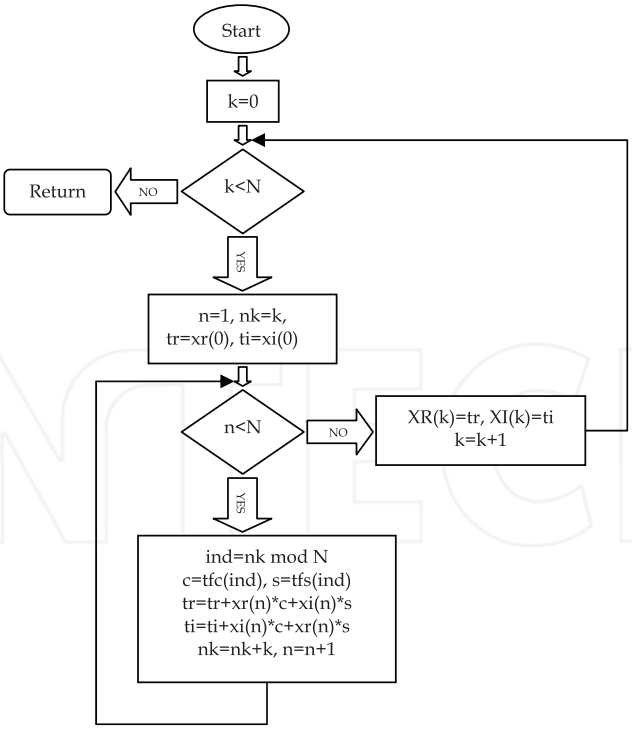


Fig. 6. The DFT Module

The last module, shown in Figure 7, prints the real and imaginary parts of the coefficients, respectively, from the arrays *XR* and *XI,* one coefficient in each iteration. This module is called *out-put (Sundararajan, 2001).*

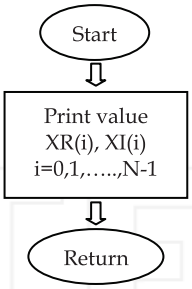


Fig. 7. The Output Module

DFT calculations are used in many scientific and engineering applications (Winograd, 1976). For the effective and advantageous use of DFT, some of its basic properties need to be learnt. It is beneficial to talk about its features shortly. The first characteristic of DFT is the linearity feature. DFT is a linear transform. Another feature of its is the symmetry feature. DFT values that are made of real values and equal to a periodical array are complex and periodical. One more feature is the principle of similarity concerning the selectivity of time and frequency. The indefinity principle of DFT is connected with the terms in the time and frequency domain of DFT. It is equal to the well-known indefiniteness principle in physics. This term is not the result of physical properties, but it is the result of a chief mathematical formulation. The final feature is the conditions of link-equivalence between DFT and Fourier Transform. Since DFT is the approximate of the continuous Fourier transform, researchers are concerned with it. The validity of this approximation is certainly based on the related wave form (Kayran & Ekşioğlu, 2004).

4. DFT application on pocket computer

In this study, the frequency analysis of the heart sounds taken from the patient via electronic stethoscope was actualized, and on the pocket computer it was displayed both in the time and frequency domain. Just as the system was completely developed mobile, it was aimed that the physicians may use it in the course of clinical examination. The physician using the system in the course of clinical examination will be able to make a more accurate and fast diagnose through listening to the heart sounds and following the heart sounds both in the time and frequency domain on the device.

4.1 Digitization of the sound

In order to play and store the heart sounds on the computer media, we need to have some kind of file formats. Because of these formats, multimedia file can be listened and stored on the computer. The *wav* (Waveform Audio Format) format to be used in our study is a kind of sound file. It is a commonly used sound file format. Unlike the other sound formats, the sound in the *wav* file format cannot be stored by compressing; it can only be stored through digitizing. As this file format is not compressed, it holds a lot of space. On the other hand, the sound quality is quite good. In this study, the *wav* sound format has been selected as it saves the digitized sound with its original state before compressing.

One *wav* file is in general consisted of two parts. In the first part, there is general information concerning the data. The second part is the part where the original data starts.
In figure 8, the structure of a *wav* file format in general has been given (a. Güraksın, 2009).

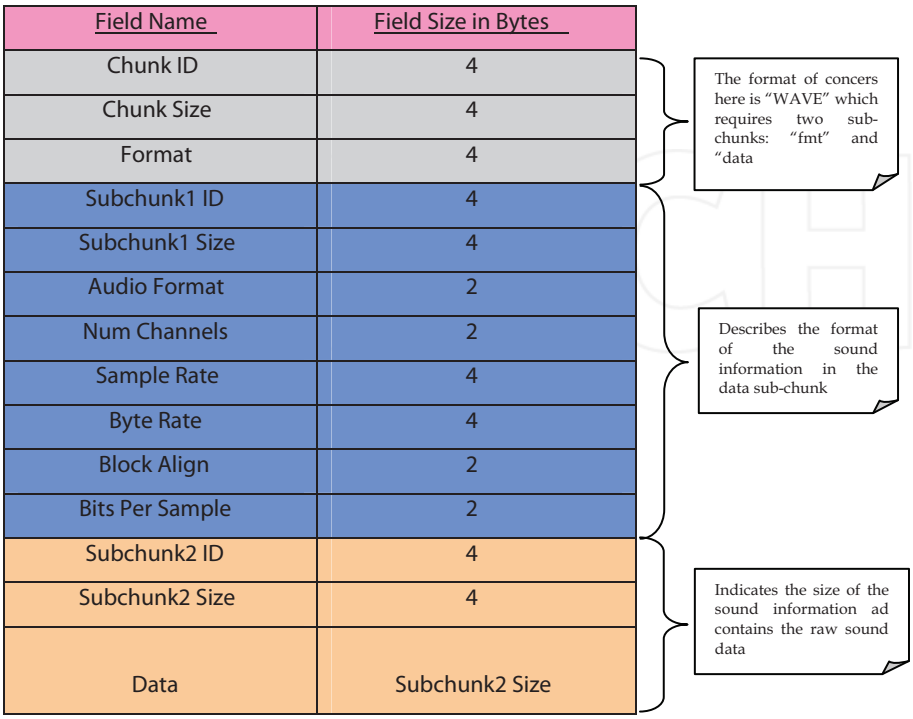


Fig. 8. The Canonical *wav* File Format

The heart sounds taken from the patient are digitized through a piece of code written accordingly to the mentioned format. During the digitization process, the digital data that are in the hexadecimal basis are obtained by reading *wav* format and are then converted into decimal basis. Thanks to the obtained digital format of the sound, it is made possible to draw a graphic of the sound on the pocket computer and DFT can be applied to the sound. With the help of a software that operates on the pocket computer, the digital data obtained from the wav sound format are turned into sound graphic in another code, and again with the help of another programming code it is subjected to DFT.

4.2 Raw data obtainment

The sounds used in this study were taken by using Littmann 4100 model electronic stethoscope from Afyon Kocatepe University with the 07.AFMYO.01 numbered scientific research project (a. Güraksın, 2009). With model 4100 Littmann electronic stethoscope, it is possible to record 6 different heart sounds. Thanks to this, the sounds taken continuously from six patients are stored within the stethoscope itself. In addition, the saved sounds will be able to be transferred to pocket computer with the help of infra-red technology within the stethoscope. The sounds recorded with Littmann 4100 model electronic stethoscope are

stored in *e4k* format. This format was converted into *wav* format on the pocket computer via a program given by Littmann. After converting the heart sounds into *wav* format, we pass to the next phase of signal processing. By and large, there are some stethoscopes in the market that are able to convert the stored sounds directly into wav format. By using such a kind of stethoscope, there will be no need to use a second transform. In order for the sounds taken by stethoscope to be processed, HP iPAQ hx2000 pocket computer that allows for infrared transaction was used.

4.3 Drawing DFT graphics of the data of heart sounds on pocket computer

The aim of this study is to get the frequency domain graphic that was obtained by DFT and the time domain graphic of the heart sounds taken from the patients drawn on the pocket computer. The software developed for this purpose was prepared through the use of programming language of C# in Microsoft Visual Studio 2005 media.

As you can see in figure 9, firstly the sound was taken from the patient by using the Littmann 4100 model electronic stethoscope. Next, the sound to be analysed in a program written in Visual Studio 2005 media was digitized, and it was subjected to DFT method. In the final stage, the graphics of the processed heart sounds both in the time domain and frequency domain on the pocket computer were drawn. The user interfaces of the software developed on the pocket computer were shown below in figures between 10 and 13.

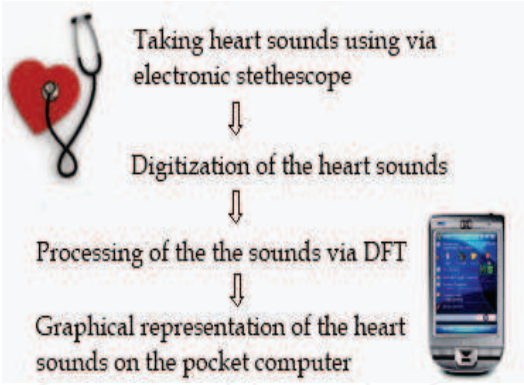


Fig. 9. Flow Diagram of the System

On the introduction screen of the software produced with C# programming language in Microsoft Visual Studio 2005 medium, there appears a selection screen shown in figure 10 on which you could select one of the heart sounds that was at first taken via Littmann 4100 model electronic stethoscope and converted into *wav* format on the pocket computer.

On this screen, the heart sound that belongs to a patient is selected, and the selected heart sound is prepared with the algorithm adapted to the pocket computer from *wav* format to the signal processing phase. After that, it moves to graphic interface. On the graphic interface, after selecting the button named as "Graph", the software on the computer subjects the sound data that was digitized beforehand to DFT. After this process is completed, the data obtained from the DFT is sent into another array. As the signal processing phase is finished, the graphic of the processed heart sound in the time domain is drawn on the upper screen of the pocket computer (the graphic plane writes as "SND" on

Fig. 10. The selection of patient's heart sound data to be examined

the left), and the frequency spectrum obtained as a result of DFT is drawn on the lower screen of the pocket computer (the graphic plane writes as "DFT" on the left). This process almost takes between 5 and 10 seconds. The screen display obtained by drawing on the pocket computer of the time domain graphic and frequency spectrum that belongs to a normal heart sound has been given in figure 11.

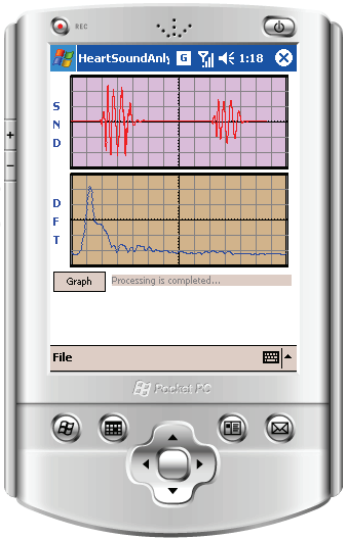


Fig. 11. The screen display with drawing of time domain graphic and frequency spectrum of heart sounds taken from a normal heart

In figure 12 and 13, the screenshots of time domain graphics and frequency spectrum that belong to the heart sounds in which in turn Pulmonary Stenosis and Mitral Stenosis disorders were detected were shown on the pocket computer.

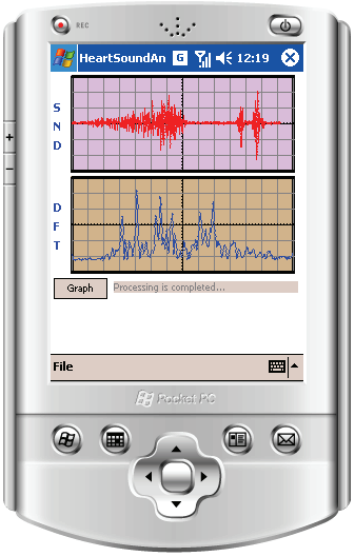


Fig. 12. The screenshot with drawing of a time domain graphic and frequency spectrum of a heart sound that owns Pulmonary Stenosis Heart Disorder

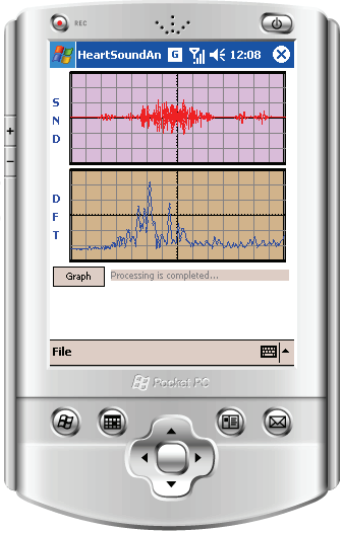


Fig. 13. The screenshot with drawing of a time domain graphic and frequency spectrum of a heart sound that owns Mitral Stenosis Heart Disorder

When there is a comparison between the frequency spectrums of normal heart sound and the heart sounds that have mitral and pulmonary stenosis heart disorders, it is noticed that the heart sounds taken from the normal heart possess less frequency constituents. (b. Güraksın et al., 2009). Because a healthy heart produces a periodical sound while a heart that has any kind of disorder produces some different sounds other than S1 and S2 sounds. Thus, these sounds include noise. The irregular and turbulent blood flows that cause all these sounds lead to inclosure of high frequency constituents in heart sounds.

5. Conclusion

In this actualized study, the heart sounds gained with the use of electronic stethoscope were digitized and then subjected to DFT. Finally, the graphic in the time domain and frequency spectrum that belong to the heart sounds was obtained on the pocket computer. Thus, frequency spectrum of normal and abnormal heart sounds was gained via DFT. As a result, the physicians were prepared with more data in the course of auscultation. By providing the physicians with alternative methods other than listening, it was aimed to help them make a more accurate and faster diagnosis.

The basic purpose in this study is not to diagnose the heart sounds directly. Instead, it was aimed to form a substructure for the prediction methods that may be used in the diagnosis phase. This formed structure sets a substructure for the artificial intelligence programs such as neural network, support vector machines, neuro-fuzzy, which are used for classification. (c. Güraksın et al., 2009). The obtained DFT data can simply be made use of in the classification mechanisms.

Acknowledgement The authors wish to acknowledge the Afyon Kocatepe University Scientific Research Council with the Project number "07.AFMYO.01" for the collection of the heart sounds used in this study.

6. References

Abdel-Alim, O. ; Hamdy, N. & El-Hanjouri, M. A. (2002), Heart Diseases Diagnosis Using Heart Sounds, *Proceedings of Radio Science Nineteenth National Conference*, pp. 634–640, Nov., 2002.

Ahlström, C. (2006), *Processing of the Phonocardiographic Signal-Methods for the Intelligent Stethoscope*, Thesis, Linköping University, Institute of Technology, Linköping, Sweden.

Barschdorff, D. ; Ester, S. ; Dorsel, T. & Most, E. (1990), Neural network based multi sensor heart sound analyses, *Proceedings of Computers in Cardiology*, pp.303–306, USA, Sep., 1990, Chicago.

Cochran, W. T.; Cooley, J. W.; Favin, D. L.; Helms, H. D.; Kaenel, R. A.; Lang, W. W.; Maling, G. C.; Nelson, D. E. ; Rader, C. M. & Welch, P. D. (1967), What is the Fast Fourier Transform?, *Transactions on Audio and Electroacoustics*, Vol. 15, Oct., 1967, pp.45–55.

Debbal, S. M. & Bereksi-Reguig, F. (2007), Time-Frequency Analysis of the First and the Second Heartbeat Sounds, *Applied Mathematics and Computation*, Vol. 184, Jan, 2007, pp. 1041–1052.

El-Segaier, M. ; Lilja, O. ; Lukkarinen, S. ; Sörnmo, L. ; Sepponen, R. & Pesonen, E. (2005), Computer-Based Detection and Analyses of Heart Sound and Murmur, *Annals of Biomedical Engineering*, Vol. 33, No. 7, Jul., 2005, pp. 937–942.

Folland, R. ; Hines, E. L. ; Boilot, P. & Morgan, D. (2002), Classifying Coronary Dysfunction Using Neural Networks Through Cardiovascular Auscultation, *Med. Biol. Eng. Comput.*, Vol. 40, 2002, pp. 339–343.

Güraksın G.E. (2009), *Classification of the Heart Sounds via Artificial Neural Network*, Ms Thesis, Institute of Natural and Applied Sciences, Afyon Kocatepe University, Afyonkarahisar, Turkey, 2009.

Güraksın G.E.; Ergün U. & Deperlioğlu Ö. (2009), Performing Discrete Fourier Transform Of The Heart Sounds On The Pocket Computer, 14th National Biomedical Engineering Meeting, İzmir, Turkey, 2009, pp. 124-127.

Güraksın G.E.; Ergün U. & Deperlioğlu Ö. (2009), Classification of the Heart Sounds via Artificial Neural Network, International Symposium on Innovations in Intelligent Systems and Applications, Trabzon, Turkey, 2009, pp. 507-511.

Hayes, M. H. (1999), *Digital Signal Processing*, The McGraw-Hill Companies Inc., USA.

Jiang, Z. & Choi, S. (2006), A cardiac sound characteristic waveform method for in home heart disorder monitoring with electric stethoscope, *Expert Systems with Applications*, Vol. 31, 2006, Aug., pp. 286–298.

Kandaswamy, A. ; Kumar, C. ; Ramanathan, R. ; Jayaraman, S. & Malmurugan, G. (2004), Neural classification of lung sounds using wavelet coefficients, *Computers in Biology and Medicine*, Vol. 34, Jul., 2004, pp. 523–537.

Kayran A. H. & Ekşioğlu E. M. (2004), *Bilgisayar Uygulamalarıyla Sayısal İşaret İşleme*, Birsen Publishing House, 1. Edition, İstanbul, Turkey.

Kemaloğlu, S. & Kara, S. (2002), Instrumentation For Synchronous Acquiring Of ECG Signals Together Heart Sounds, *Journal of Erciyes University Institute of Science and Technology*, Vol. 18, (1-2), pp. 28–33.

Leung, T. S.; White, P. R. ; Collis, W. B. ; Brown, E. & Salmon, A. P. (2000), Classification of Heart Sounds Using Time-Frequency Method and Artificial Neural Network, *22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 2, pp. 988–991.

Ölmez, T. & Dokur, Z. (2003), Classification of Heart Sounds Using an Artificial Neural Network, *Pattern Recognition Letters*, Vol. 24, pp. 617–629.

Say, Ö. (2002), *Analysis of Heart Sounds and Classification by Using Artificial Neural Networks*, Ms Thesis, İstanbul Technical University, Institute of Natural and Applied Sciences, İstanbul, Turkey, 2002.

Sharif, Z. ; Zainal, M. S. ; Sha'ameri, A. Z. & Salleh, S. H. S. (2000), Analysis and Classification of Heart Sounds and Murmurs Based on the Instantaneous Energy and Frequency Estimations, *Proceedings of TENCON*, Vol. 2, pp. 130–134, 2000.

Sinha, R. K. ; Aggarwal, Y. & Das, B. N. (2007), Backpropagation Artificial Neural Network Classifier to Detect Changes in Heart Sound due to Mitral Valve Regurgitation, *Journal of Medical Systems*, Vol. 31, Jun., 2007, pp. 205–209.

Smith, S.W. (1997), *The Scientist and Engineer's Guide to Digital Signal Processing*, California Technical Publishing, 1. Edition, San Diego, USA.

Sundararajan, D. (2001), *The discrete fourier transform: theory, algorithms and applications*, World Scientific Publishing Co. Pte. Ltd., USA.

Swanson, D. (2000), *Signal Processing for Intelligent Sensor System*, Marcel Deker, 1.Edition, NY, USA.

Winograd, S. (1976), On Computing the Discrete Fourier Transform, *Proceedings of the National Academy of Sciences*, Vol. 73, No. 4, Apr., 1976, pp. 1005-1006.

10

Towards a Reconfigurable FFT : Application to Digital Communication Systems

Florent Camarda,Jean-Christophe Prevolet and Fabienne Nouvel
IETR/INSA Rennes
France

1. Introduction

Since the beginning of humanity, people have always wanted to communicate with each other. Today, this objective remains the same although a huge technological step has been taken since the first communication link. More recently, with the development of electronics systems, the number of interconnected devices has considerably increased and the quest of performances becomes more and more a reality.

In this context, digital communication systems have provided an efficient alternative to classical analog devices. For example, in the telecommunication domain, as analog television was a great technological progress less than a century ago, 3D HD television sets are currently being commercialized and will progressively replace the old devices.

The proliferation and variety of such digital devices have also led to the elaboration of multiple standards destined to ease the exchange of information between the different devices. Unfortunately, there are still too many standards to improve the portability and interconnections between different systems. The existence of such an amount of various standards is related to economical and technological constraints but also to political and strategical reasons which makes the interoperability between standards very difficult. This generally imposes the engineers to design as many circuits as standards which seems very inefficient and may be seen as a waste of time and money. This solution is even worse as a particular standard does not differ completely from another one. There are generally a lot of similarities between standards in terms of proposed algorithms and functions. It is sometimes sufficient to modify some parameters of a specific algorithm to change from one standard to another.

For example, in current communication systems, many applications referring to several standards make use of Orthogonal Frequency Division Multiplexing (OFDM) modulations (Weinstein & Ebert, 1971) such as Digital Terrestrial Television Broadcasting (DTTB), Digital Audio Broadcasting (DAB), Wireless Local or Personal Area Network (WLAN-WPAN), HomePlug 1.0 and HomePlug AV, HyperLAN, 802.11 standards, Digital Subscriber Line standards (xDSL), 3GPP, Long Term Evolution (LTE), etc.

The OFDM principle is mainly based on the use of the Fast Fourier Transform (FFT) and its Inverse (IFFT) (P. Duhamel & M. Vetterli, 1990). However, the sizes of FFT employed in all

these standards are various. DVB standards use 2048, 4096, 8192, 16384 or 32768 points FFT, Homeplug standards require 128 or 3072 points FFT, the chinese DTV standard employs a 3780 points FFT. Moreover, some of these communication systems need to implement several FFT sizes within a particular standard. It is the case for SC-FDMA or for all Time Domain Synchronous OFDM (TDS-OFDM) systems (J. Wanget al., 2003), in which a second FFT size is often needed for synchronization.

In order to develop a multi-standards receiver, the solution which is usually retained by manufacturers consists in implementing various Intellectual Properties (IPs) blocks, each associated to a unique function and a unique standard. This solution is often under-optimized in the sense that it does not allow the reuse of components within the chip and thus constitutes a waste of hardware resources.

Another approach is to design a generic and unique system capable of dealing with the specifications of different standards. This circuit has to be designed in such a way that it must be able to adapt itself to a particular standard or a dedicated application.

This concept, known as dynamic reconfigurability, provides not only a high flexibility but also permits to considerably reduce the time to market and the design effort.

Dynamic reconfigurability is a powerful concept since it enables a circuit to change its functionality on the fly according to the users' needs or the environmental conditions. The dynamic reconfiguration may be performed at different granularities depending on the technology. For example, some circuits allow to be reconfigured at a gate level by modifying the contents of a logic equation. This is typically the case of FPGAs circuits (Field Programmable Gate Array) which are widespread circuits that constitute a good alternative to ASICs (Application Specific Integrated Circuits). Other circuits may be reconfigured at a higher level of granularity, for example at a functional level. In this case, a specific operator (or function) is modified to change the behavior of an application.

In general, the level of performances depends on the granularity of reconfiguration. If the granularity is very fine, the flexibility is high but the timing performances are very low since a lot of resources have to be reconfigured. As a consequence, the time which is necessary to reconfigure the system constitutes a real issue and limits the use of such circuits in real time embeddded systems where timing constraints are generally tight.

Nevertheless, some circuits make intensive use of reconfiguration at a high level of granularity and permit to obtain not only high performances but also flexibility. This is generally performed by reconfiguring specific operators or group of operators in order to change the design functionality.

In this article, we present a new system capable of dealing with multi communication standards. The basis of this work has consisted in developing a unique FFT operator which is capable of being reconfigured to adapt to the different standards' specification. In our case, the FFT operator should be able to compute FFTs of size 2K, 4K, 8K and 3780 points.

In section 2, the nature of the FFT operator will be presented as well as the different possibilities for its implementation. Section 3 will describe the proposed reconfigurable architecture and obtained performances after implementation. An optimization is presented in section 4 with its performances results.

2. FFT Algorithms

The Fourier transform is a very useful operator for image or signal processing. Thus it has been extensively studied and the litterature about this subject is very rich. The Discrete Fourier Transform (DFT) is used for the digital signal processing and its expression is given in (1)

$$X(k) = \sum_{n=-\infty}^{\infty} x(n) e^{j2\pi \frac{nk}{N}} \tag{1}$$

It appears obvious that this expression can not be computed in a finite time due to the infinite bounds. From that, the usully computed expression is the N-points Fast Fourier Transform given in (2)

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{j2\pi \frac{nk}{N}} \tag{2}$$

The expression of the FFT is bounded and computable with a finite algorithmic complexity. This complexity is expressed as an order of multiplications and additions. Computing a N-points FFT without any simplification requires an algorithmic complexity of $\mathcal{O}(N^2)$ multiplications and $\mathcal{O}(N^2)$ where $\mathcal{O}$ denotes the "order of" multiplications and additions. Note that the real number of additions is $N(N-1)$ which is $\mathcal{O}(N^2)$. This reduction of complexity is however not sufficient for the large FFT sizes that are used in many digital communications standards.
A great reduction of this complexity can be achieved by using an efficient algorithm. It exists many FFT algorithms through the litterature like the Radix algorithm (J. W. Cooley & J. W. Tukey, 1965), the Prime Factor Algorithm (PFA) also called Good Thomas Algorithm (I. J. Good, 1958). We also can cite the algorithms of Rader-Brenner (C. M. Rader, 1968), Bruun's (G. Bruun, 1978) or the Winograd Fourier Transform Algorithm (WFTA) (S. Winograd, 1978). The most known and used of them is unmistakably the Radix algorithm. However other algorithms like the WFTA can be very convenient for specific FFT sizes. In this section, we focus on the development of the Radix algorithm. The Mixed-Radix algorithm will then be introduced. We also present two advantages of the WFTA and propose a way to combine the Radix and WFTA algorithms.

2.1 The radix algorithm

In 1965 Cooley and Tukey proposed a new way to compute the FFT for sizes that are power of 2, this is the birth of the Radix algorithms. The basic principle is based on the decomposition of a FFT of size N in two FFTs of size $\frac{N}{2}$.

2.1.1 Decimation in time / Decimation in frequency

This decomposition can be achieved by two means which are called Decimation In Time (DIT) and Decimation In Frequency (DIF) depending on the regrouped terms. Equation (3) present the decomposition for the DIT while equations (4) and (5) are for the DIF. Thanks to this the algorithmic complexity becomes reduced to $\mathcal{O}(\frac{N}{2}^2)$ since $N = 2$.

$$\begin{aligned} X(k) &= \sum_{n=0}^{N-1} x(n) e^{j2\pi \frac{nk}{N}} \\ &= DFT_{\frac{N}{2}}\left[x(2n')\right] + e^{j2\pi \frac{k}{N}} DFT_{N/2}\left[x(2n'+1)\right] \quad \text{with } n' \in \left\{0, ..., \frac{N}{2}-1\right\} \end{aligned} \tag{3}$$

This decomposition is called DIT because the N time samples are reordered in two groups of N/2 samples. One group contains the even indexed samples, while the second one contains the odd indexed samples. FFT of size N/2 are then computed for each group. The frequency samples are computed in order.

$$\begin{aligned} X(2k') &= \sum_{n=0}^{N-1} x(n) e^{j2\pi \frac{n2k'}{N}} \\ &= DFT_{\frac{N}{2}}\left[x(n) + x(n+\frac{N}{2})\right] \quad \text{with } k' \in \left\{0, ..., \frac{N}{2}-1\right\} \end{aligned} \tag{4}$$

$$\begin{aligned} X(2k'+1) &= \sum_{n=0}^{N-1} x(n) e^{j2\pi \frac{n(2k'+1)}{N}} \\ &= DFT_{\frac{N}{2}}\left[\left(x(n) - x(n+\frac{N}{2})\right) e^{j2\pi \frac{n}{N}}\right] \quad \text{with } k' \in \left\{0, ..., \frac{N}{2}-1\right\} \end{aligned}$$

This decomposition is called DIF because the even indexed frequency samples are computed by the first N/2-points FFT while the odd indexed samples are computed by the second one. The time samples are not reordered before computation. In order to simplify these equations all terms in $e^{j2\pi \frac{x}{N}}$ are usually expressed as W_N^x.

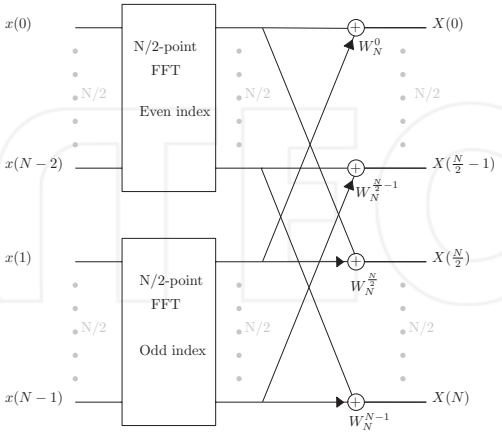


Fig. 1. Radix-2 Decimation In Time

Figure (1) represents the computation of the previous DIT. With this structure it is still necessary to perform N multiplications by the twiddle factors W_N^k to compute the N FFT outputs. However it is possible to use a property of the twiddle factors expressed in (5), to remove half of these multiplications, by replacing some additions by substractions.

$$W_N^{x+\frac{N}{2}} = -W_N^x \tag{5}$$

Then the opimized DIT decomposition will be computed as presented in figure (2)

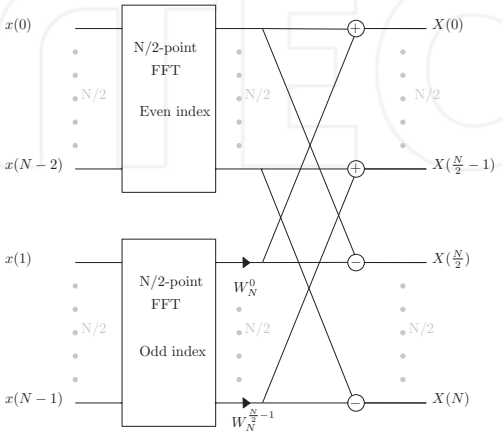


Fig. 2. Simplified Radix-2 Decimation In Time

Figure (3) presents the computation scheme for the DIF decomposition. In this last one, it appears that the N time samples are precomputed in order while the frequency samples are reordered in an even indexed group and an odd indexed group. This decomposition also requires $\frac{N}{2}$ multiplications by the twiddle factors and use the same Radix-2 butterfly in pre-computation that was used for the post computation of the DIT computation.
For these last two schemes, the algorithmic complexity is again reduced of $\frac{N}{2}$ multiplications but is still $\mathcal{O}(\frac{N}{2}^2)$. However, if the FFT size is power of 2, this decomposition can be performed until requiring $\mathcal{O}(\frac{N}{2}log_2(N))$ 2-points FFT butterflies. Because the computation of a 2-points FFT, represented figure (4), does not involve any multiplication, such decomposition requires only multiplications by the twiddle factors. This leads to an algorithmic complexity of $\mathcal{O}(\frac{N}{2}log_2(N))$.
Figure 5 shows the result of such a decomposition for a 8-points FFT with both decimations. The first thing that we can observe on this picture is that many of $\frac{N}{2}log_2(N)$ twiddle factors are $W_N^0 = 1$ which allows a reduction of the algorithmic complexity. A second important point is the high regularity in the data flow mapping. The data reordering is also very regular and can be determined by the "bit reversed" algorithm.
While all this study has been led for the Radix-2 algorithm, it is possible to generalize it to other Radix, and to the Radix-4 in particular.

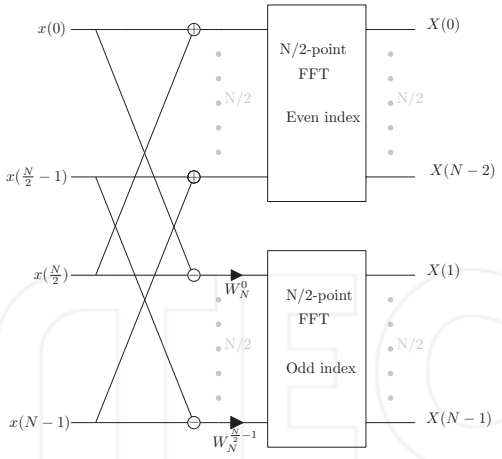


Fig. 3. Radix-2 Decimation In Frequency

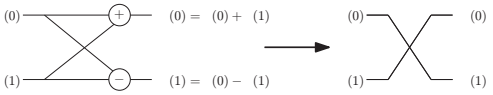
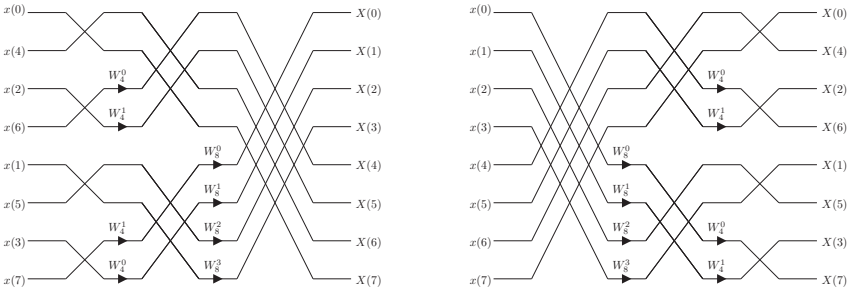


Fig. 4. Radix-2 Butterfly



(a) DIT (b) DIF

Fig. 5. 8-points FFT by using Radix-2 algorithm with DIT or DIF

2.1.2 Radix-R algorithm

As the Radix-2 butterfly computes a FFT of size 2, the Radix-4 butterfly does the same for FFTs of size 4. The four outputs of this butterfly only need trivial multiplications by $\{-1, 1, -j, j\}$ as presented in the equation (6).

$$\begin{aligned} X(0) &= x(0) + x(1) + x(2) + x(3) \\ X(1) &= x(0) + jx(1) - x(2) - jx(3) \\ X(2) &= x(0) - x(1) + x(2) - x(3) \\ X(3) &= x(0) - jx(1) - x(2) + jx(3) \end{aligned} \tag{6}$$

This butterfly however presents a gain regarding the algorithmic complexity when it is used instead of the Radix-2 by reducing the necessary number of stages in the decomposition. The major drawback of using this algorithm is that it is suitable for FFT sizes that are power of 4, which reduces considerably the field of applications.
As with the Radix-2 or the Radix-4 algorithm, it is possible to use the Radix-R algorithm for other R values. Nevertheless, the Radix-R butterfly will require no trivial multiplications into the butterfly and thus increase its algorithmic complexity. The equation 7 shows that Radix-3 butterfly computation requires 4 complex multiplications by coefficients around the unit circle. Pay attention not to confuse these coefficients with the twiddle factors needed between the stages.

$$\begin{aligned} X(0) &= x(0) + x(1) + x(2) \\ X(1) &= x(0) + W_3^1.x(1) + W_3^2.x(2) \\ X(2) &= x(0) + W_3^2.x(1) + W_3^1.x(2) \end{aligned} \tag{7}$$

Except for R multiple of 2, where it may be decomposed by using $R = 2$ or $R = 4$, a Radix-R butterfly requires $(R-1)^2$ muliplications and $R(R-1)$ additions. Thus the complexity becomes $\mathcal{O}(R^2 log_R(N))$. This is the reason why sizes which are power of two are promoted in most of digital communication standards.

2.2 Mixed-radix algorithm

The Radix-2 algorithm proposed by Cooley and Tukey can be derived to provide the mixed-Radix algorithm (R. C. Singleton, 1969) (G. L. DeMuth, 1989). This last one employs the combination of different Radix-R algorithms allowing the computation of more FFT sizes. For example, to compute a 12-points FFT, the size N can be decomposed in $N = 2 \times 2 \times 3$. In this configuration, the FFT will be computed by a first and a second stage of six Radix-2 butterflies and a third stage of four Radix-3 butterflies. Nevertheless this FFT can also be computed by other decompositions such as $N = 2 \times 3 \times 2$, $N = 3 \times 2 \times 2$, $N = 3 \times 4$ or $N = 4 \times 3$. Due to the reduced number of computation stages and to the optimal complexity of the Radix-4 butterfly, the last two decompositions will require the lower complexity.
To summarize, the complexity of a Radix or Mixed-Radix algorithms will depend on the complexity of the Radix-R butterflies that are used and the number of twiddle factors necessary between the stages.

2.3 Advantages of the Winograd fourier transform algorithm

In 1978, S. Winograd presented his works on the computation of the DFT by using cyclic convolution and introduced the Winograd Fourier Transform Algorithm which reduces the algorithmic complexity of the FFT. This reduction is achieved thanks to two points.
The first gain comes from the computation of butterfly of size R, which may be computed with a complexity of $\mathcal{O}(R)$ instead of $\mathcal{O}(R^2)$. The algorithms for $R \in \{2, 3, 4, 5, 7, 8, 9\}$ are available in (S. Winograd, 1978).
The second optimization is performed thanks to an ingenious mapping of the data flow. This mapping does not implement any twiddle factors between stages. However, this mapping can only be realized for FFT sizes which can be decomposed in several terms R_i that must be mutually primes. This is the major constraint of this type of optimization.

For a FFT of size N that can be decomposed in $N = R_1 \times R_2$ which are mutually primes, the mapping will be determined by the following rules.
First the input data vector $[x(0)...x(N-1)]$ have to be reordered in a two dimensionnal matrix of size (R_1, R_2). The position (r_n, c_n) of the sample $x(n)$ in this matrix is given by the relation (8).

$$\begin{aligned} r_n &= n \mod R_1 \\ c_n &= n \mod R_2 \end{aligned} \tag{8}$$

Then the computation of the Winograd butterfly R_1 will be computed over the R_2 groups of data which will be taken column-wide and similarly for the R_2 butterfly.
If the FFT size has to be decomposed in $I > 2$ terms (still mutually primes), this mapping can be applied with a two dimentionnal matrix of size $(R_1, \prod_{i=2}^{I} R_i)$. Thus the computation of the $\prod_{i=2}^{I} R_i = \frac{N}{R_1}$ will be performed using the same algorithm with the dimensions $(R_2, \prod_{i=3}^{I} R_i)$ for the R_1 groups generated by the first grouping.

2.4 Combined radix and WFTA algorithms

As the decomposition of the FFT in small-N FFTs was explained previously, it is possible to regroup and compute these small-N FFT as desired. So it is possible to compute a part of the FFT using the WFTA algorithm and the other part using the Mixed-Radix algorithm. However the decomposition has to be properly performed in order to take advantage of the properties of these two algorithms.
As the WFTA has better performances than the Radix algorithm, it is preferable to first regroup the mutally primes factors and compute a small FFT with the WFTA. Then the other terms will require a Mixed-Radix mapping. However the first advantage of the WFTA which reduce the complexity of the butterflies have to be use for Radix-R butterflies when $R \notin \{2, 4\}$[1].
Another optimization can be achieved by an efficient organization of the Radix/WFTA FFT. The number of *different* twiddle factors Nb_W that will have to be stored in memory for the FFT computation can be reduced if the different stages are well organized. This number of *different* coefficients is $\mathcal{O}(\prod_{i=1}^{i_0}(R_i)$ for the stage i_0, $i_0 \neq 1$ with the DIT computation, or for the $I - i_0$, $i_0 \neq 1$ for the DIF. For $i_0 = 1$ all the coefficients are $W_N^0 = 0$.

3. Proposed architecture

3.1 Presentation

In order to design the most flexible architecture capable of implementing the FFT, we decomposed the algorithm in several WFTA or Radix blocks depending on the considered N size. For example, a 64-points FFT may be decomposed in $N = 4 \times 4 \times 4$ computation stages, each stage being configured using the *Radix-4* butterfly. In the case of a 3780-points FFT, it may be decomposed in $N = 3 \times 3 \times 3 \times 4 \times 5 \times 7$ which leads to implement the *WFTA-3*, *Radix-4*, *WFTA-5* and *WFTA-7* butterflies (Z.-X. et al., 2002).
Such decompositions make it possible to consider a reconfigurable WFTA/Radix pipelined architecture that allows to compute several sizes of FFT by reconfiguring each stage's butterfly.

[1] in fact the WFTA butterfly complexity is the same as the Radix butterfly for $R = 2$ or $R = 4$

In this section, we focus on the presentation of an architecure dedicated to a multiple Digital Terrestrial Television Broadcasting (DTTB) demodulation. The concerned sandards are DVB-T (ETSI, 2004), ISDB-T (ARIB, 2001) and DTMB (Chinese National Standard, 2006) for which the possible FFT sizes are 2048 (2K), 4096 (4K), 8192 (8K) and 3780 points. Nevertheless, it is possible to adapt the architecture to other FFT sizes. Table (1) presents a way to decompose these sizes by implenting the different Radix or WFTA butterflies. The presence of a Radix-8 butterfly is necessary to compute the 8192-points FFT in six stages, but another stage of computation can be added in order to split this butterfly in two by using Radix-4 and Radix-2 butterflies.
As shown in section 2.4, the number of different twiddle factors coefficients that will have to be computed depends on the placement of the different butterflies. To provide the output samples in order, a DIT computation is considered. With such a decomposition, the quantity of required twiddle factors will be minimized by placing the higher radix butterflies at the last stages of computation. Moreover, as the second advantage of the WFTA can be employed in the 3780 decomposition for the $3 \times 4 \times 5 \times 7 = 420$ part, all these 420-points FFT computations have to be processed in the last stages.

	Butterflies Configuration					
	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6
3780	WFTA-3	WFTA-3	WFTA-3	Radix-4	WFTA-5	WFTA-7
2048	Radix-2	Radix-4	Radix-4	Radix-4	Radix-4	Radix-4
4096	Radix-4	Radix-4	Radix-4	Radix-4	Radix-4	Radix-4
8192	Radix-4	Radix-4	Radix-4	Radix-4	Radix-4	Radix-8

Table 1. A possible decomposition of the concerned FFT sizes

Thus, such a decomposition in six stages of computation may be performed by the 6-stages reconfigurable architecture presented in figure (6).
Each stage is built on a reconfigurable WFTA/Radix module implementing the corresponding butterfly. The module may be reconfigured depending on the chosen FFT size. The butterfly input and output data are stored in the symbol memory. The address decoders select the memory address to read the inputs and write the results. A twiddle block is also implemented at each stage of the computation (except for the last one). This block is located between the butterfly module and the symbol memory and performs the multiplications by the appropriate twiddle factors. The input and output buffers are used to respectively store the time domain and frequency domain OFDM samples. Finally, the "memory data and address" multiplexers switch the memories between stages of computation.

3.2 Operating description

The N complex samples corresponding to the S_n OFDM time domain symbol are stored in the input buffer. During this storage, the WFTA/Radix module of the first stage processes its computation according to the required configuration (Radix-2, WFTA3 or Radix-4). This process is performed on the N previously stored samples that constitute the S_{n-1} OFDM symbol.
Each subsequent module performs the same operation in parallel until the WFTA/Radix module in the stage 6, that operates on the S_{n-6} OFDM symbol. While these computations are performed, the frequency domain samples corresponding to the FFT of the S_{n-7} are available at the output buffer.

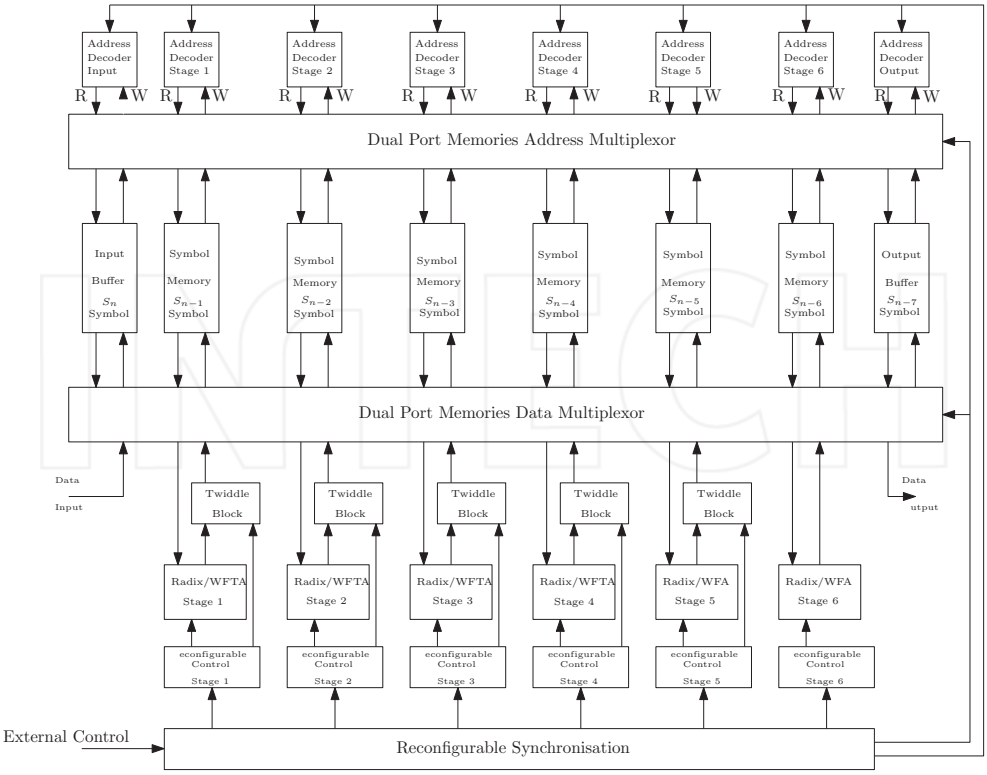


Fig. 6. 6 stages reconfigurable architecture for FFT

As soon as the S_n symbol is completely stored, the different computation stages must be completed. Memories are then circularly switched i.e. the previously input buffer becomes the symbol memory 1, the symbol memory i becomes the symbol memory $i+1$ and the output buffer becomes the new input buffer. This permutation is performed by the two memory multiplexers, one for the data bus, the other for the address bus.
When the algorithm requires the twiddle factors multiplications, these operations are computed by the twiddle block at the output of the butterfly, before being stored in the symbol memory.

3.3 Presentation of the Processing Blocks

3.3.1 WFTA/radix module

Each WFTA/Radix module is composed of three stages. A first stage aims to compute additions, a second performs the multiplications by coefficients and the last adder stage generates the output results. The figure (7) presents the architecture of such a module.
First, the set of butterfly inputs are stored in the butterfly input register. As soon as possible, the adder/substractor starts to process two values. These values are either the time domain samples stored in the register or a previous result which comes from a reconfigurable delay block. The second stage performs a multiplication of the results coming from the first stage by the coefficients which are stored in the coefficient memory. As the coefficients are either real or

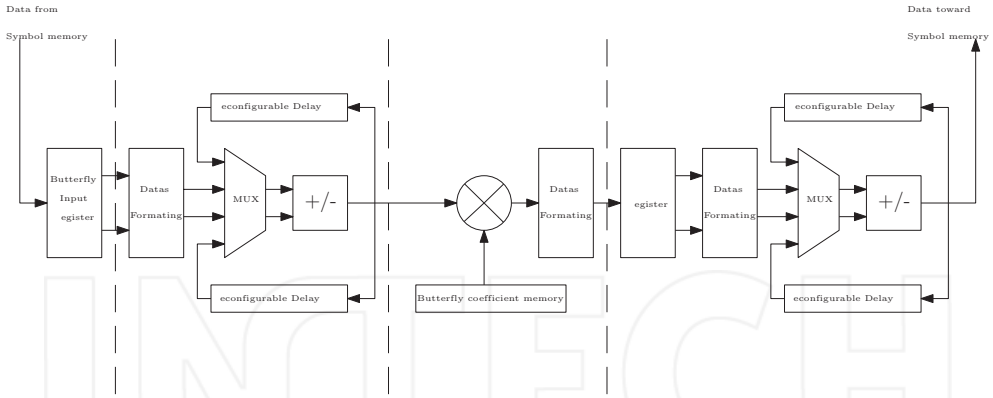


Fig. 7. Architecture of a WFTA/Radix module

pure imaginary, only two real multipliers are necessary to implement the complex multiplier. Since there are only few coefficients, their resource utilization is negligible compared to the resources required by symbol memories. Finally, the output adder/substractor part operates similarly to the first one and provides the butterfly outputs.

3.3.2 Twiddle blocks

As no twiddle factors are needed at the output of the last stage, the twiddle block does not appear on the figure 6. The architecture of the twiddle blocks is depicted in figure (8). This block just performs the complex multiplications by the twiddle factors stored in the associated memory. When multiplications are not necessary, a multiplication by $W_N^0 = 1$ is computed to maintain a constant propagation delay through the architecture.
Such a multiplication can be implemented by either four real multipliers and four real adders or three real multipliers and five real adders.

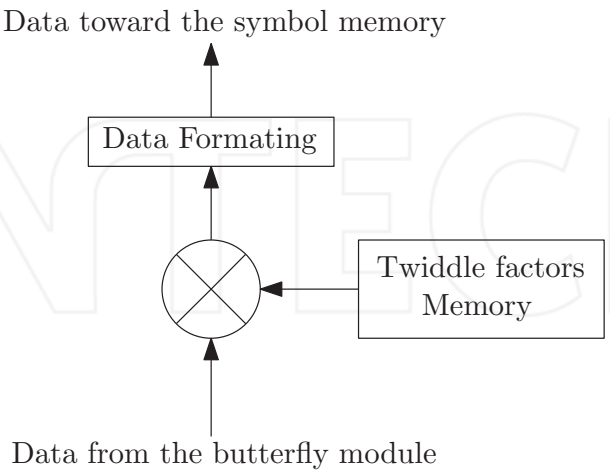


Fig. 8. Twiddle block

3.3.3 Symbol memories, address decoders

The symbol memories have to be connected to the right stage depending on the computation step. Then, during the computation stage, the memory will send data to the butterfly while recording data from the twiddle block. The time domain samples are overwritten by the frequency domain samples. Nevertheless, the read and write addresses are always different at the same time, due to the butterfly latency. For that purpose, a Simple Dual Port Random Acces Memory (SDP-RAM) have been implemented. The size of this Random Access Memory (RAM) has been defined according to the maximum number of points to be computed in the design (8192 samples). Each complex sample is coded into 32 bits (16 bits for the real part and 16 bits for the imaginary part). Thus, the size of the symbol memory has been set to $2 \times 16 \times 8192$ bits which corresponds to $32kB$.
The address decoders compute the right sequence of memory address that have to be read and written. Thus an address decoder block is dedicated to a stage of computation. As the size of the FFT can be reconfigured, these blocks also have to be reconfigured.

3.4 Performances of the 6-Stages architecture

The complete architecture has been simulated and implemented on an Altera stratix EP3SE50F484C2. This FPGA has been envisaged as a prototyping circuit for future implementation in an Application Specific Integrated Circuit (ASIC). The considered FPGA exhibits a lot of logical resources as well as dedicated blocks that are optimized for MAC (Multiplications/Accumulations) operations and memory. Thus, it is perfectly suited for the considered types of applications.

3.4.1 Resources usage

Implementation results for the 6-stages architecture are provided in Table (2). In this table, the implemented architecture is compared to commercial IPs provided by Altera (Altera, 2009) in terms of resource usage.

	Logic Cells	Registers	Memory bits	DSP blocks
Available on Stratix III	38,000	38,000	5,455,872	384
6-stages architecture	4,077	2,544	3,769,080	44
Altera 2K IP	4,138	6,943	208,329	40
Altera 4K IP	4,557	7,530	425,563	40
Altera 8K IP	5,270	8,670	884,785	48

Table 2. Resources Utilization for the 1-Stage and the 6-Stages Architecture After Implementation on Stratix III

It seems important to notice that the proposed architecture requires less logic cells, registers and DSP block than its counterparts. This is especially true if we consider that this architecture is also able to compute different sizes of FFT (2K, 3780, 4K, 8K).
Another interesting point is that the memory resource is the most consuming in this architecture. This is mainly due to the parallelization of several stages in the architecture. Thus the equivalent of six OFDM symbols must be stored in memory.
Globally, implementation results demonstrate the feasibility of implementing such architectures on reconfigurable circuits without consuming too many resources. This makes

it possible to envisage the implementation of this architecture onto ASICs which generally present higher performances and exhibit more logical resources.

3.4.2 Time performances

The time that is required to store a complete symbol into the input buffer RAM is the OFDM symbol duration T_{OFDM}. This duration depends on the sampling frequency, the size of the IFFT and the guard interval duration, used for the transmission. The minimun T_{OFDM} and associated maximum OFDM throughput for each standard and FFT sizes are provided in Table (3).

		DVB	ISDB	TDMB
3780	Duration (μs)	-	-	555
	Throughput ($symb/s$)	-	-	1802
2048	Duration (μs)	231	259	-
	Throughput ($symb/s$)	4329	3861	-
4096	Duration (μs)	462	519	-
	Throughput ($symb/s$)	2165	1927	-
8192	Duration (μs)	924	1039	-
	Throughput ($symb/s$)	1082	962	-

Table 3. Minimum OFDM symbol duration with Guard Interval and corresponding throughput

The architecture must, at least, respects these performances to be suitable in the OFDM receiver. Table (4) describes the number of clock cycles required to compute a complete stage of the FFT. According to this table, it may be seen that the WFTA-7 stage demands the most important computation time per sample.

FFT size	3780	2K	4K	8K
WFTA7	11351	-	-	-
WFTA5	7570	-	-	-
WFTA3	3787	-	-	-
Radix-8	-	-	-	16398
Radix-4	3787	2055	4103	8199
Radix-2	-	2055	-	-

Table 4. Module Execution Time in Number of Clock Cycles.

In order to provide a comparison between the proposed architecture and existing solutions, several implementation tests have been led. The obtained results are presented in Table (5) which summarizes the symbol rate as well as the computation latencies for the proposed architecture according to different configurations. The comparison has been performed with different IPs of Altera implementing the corresponding sizes of DFT. The maximum rate that may be achieved is conditioned by the WFTA/Radix Module 6 which exhibits the most important complexity.
Regarding Table (5), for each FFT size, the 6-stages architecture delivers a similar symbol rate than the commercial IPs. Note that a significant difference remains for the 8192 points FFT. This is due to the internal structure of our proposed architecture that only consists of 6 stages

		Proposed 6-stages Architecture	Altera IP
Symbol rate @ 100MHz (OFDM symbols / s)	2K	48,661	48,828
	4K	24,372	24,414
	8K	6,098	12,207
	3780	8,809	NC
Latency (clock cycles)	2K	10,282	6,144
	4K	20,522	12,288
	8K	49,201	24,576
	3780	30,289	NC

Table 5. Symbol Rate and Latency According to Different Configurations

of computations. In order to compute a 8K FFT, it is then necessary to use a Radix-8 Module which requires much more computing cycles compared to Radix-4 or Radix-2 Modules.
Concerning the latency, an increase has to be deplored in the proposed architecture. This is mainly due to the time that is necessary to store a complete symbol into an intermediate buffer between two consecutive stages of computation. Nevertheless, the drawback of latency may be compensated by the high symbol rate which is available at the output of the design.

4. Optimization for the studied applications

The previous architecture is optimized for a high symbol rate. However, it is possible to modify it by reducing the number of stages from six to one as described in figure (9).

4.1 Operating description

According to the figure (9), during the storage of a current symbol in the input buffer, the unique WFTA/Radix Module will be reconfigured to compute the different stages of the whole FFT. The connected symbol memory is employed for the computation of all stages. At the end of the computation, the input buffer becomes the new symbol memory, the symbol memory becomes the output buffer with the FFT transform results and the previous output buffer becomes the new input buffer which will store the new incoming OFDM symbol.
This operating mode allows a flexible number of computation stages. Thus the 8192-points FFT may be computed by using only *Radix-4* and *Radix-2* algorithms in seven stages of computation ($8192 = 4 \times 4 \times 4 \times 4 \times 4 \times 4 \times 2$).
The used module can be implemented in 6 possible configurations: *Radix-8*, *WFTA7*, *WFTA5*, *Radix-4*, *WFTA3* and *Radix-2*. All combinations of these possibilities can be envisaged to compute several FFT sizes. The number of stages is related to the number of terms in the FFT size decomposition. Nevertheless, an increase of this number implies a reduction of the symbol rate.
Furthermore, for particular sizes of FFT which are neither power-of-2 nor adapted to the WFTA mapping algorithm, new Twiddle factors have to be calculated and stored in the twiddle block memory. In order to take into account all possible configurations, it is possible to implement an external memory which contains all these coefficients and load them in the Twiddle memory block according to the chosen FFT size.

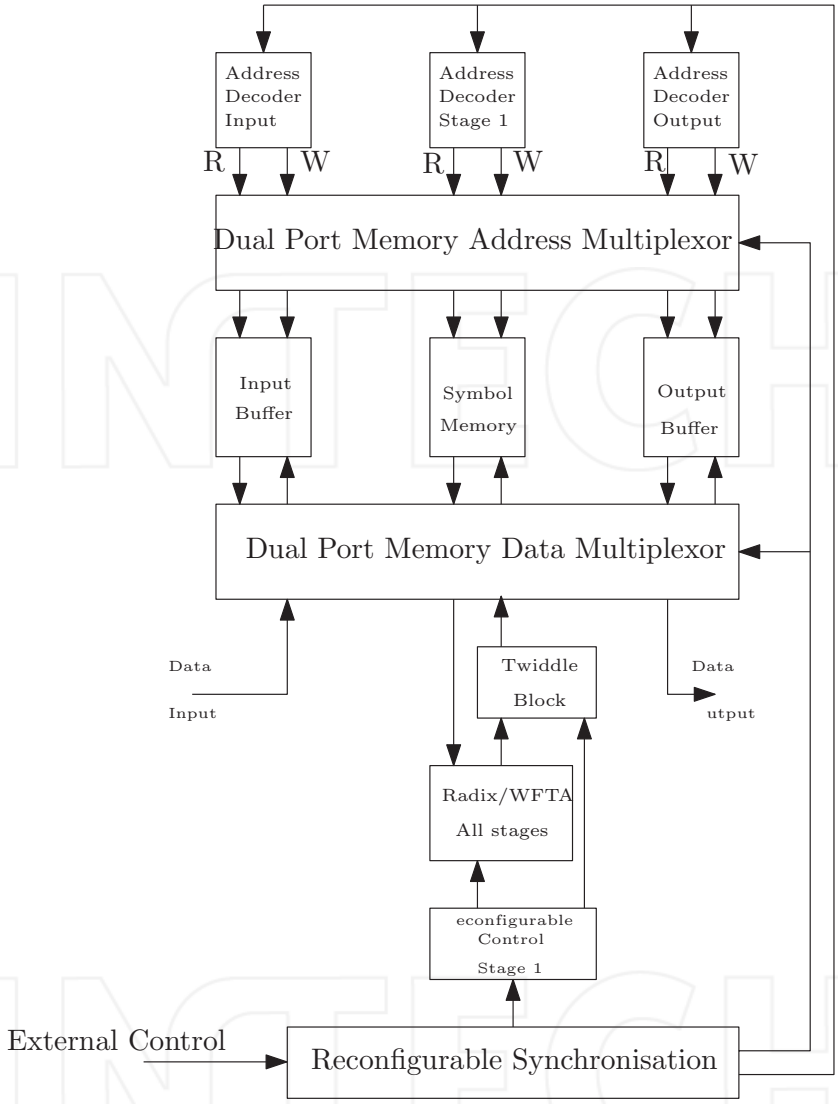


Fig. 9. 1-stage reconfigurable architecture for FFT

4.2 Reconfiguration of the architecture

As the reconfiguration of the WFTA/Radix Modules for the 6-stages architecture occurs only when the FFT size is modified, the reconfiguration is completely static. On the other hand, the Butterfly module of the 1-stage architecture must be reconfigured during runtime. This reconfiguration is possible almost instantaneously. In fact, only the control of the resources has to be reconfigured. Thus it is possible to reconfigure it by using a multiplexed control.

4.3 Resources Usage for the 1-Stage architecture

The 1-stage architecture has been implemented into the same FPGA target. The corresponding resources usage is summarized in Table (6). In this configuration, the obtained results are more competitive than Altera IPs, especially for memory usage. This may easily be seen when comparing the number of memory bits in the 8K mode. Moreover, it is clear that logical resources are reduced significantly.

	Logic Cells	Registers	Memory bits	DSP blocks
Available on Stratix III	38,000	38,000	5,455,872	384
1-stage architecture	811	548	622,896	8
Altera 2K IP	4,138	6,943	208,329	40
Altera 4K IP	4,557	7,530	425,563	40
Altera 8K IP	5,270	8,670	884,785	48

Table 6. Resources Utilization for the 1-Stage and the 6-Stages Architecture After Implementation on Stratix III

4.4 Timing considerations for the 1-Stage architecture

The execution times and latency that are depicted in Tables (4) and (5) are given for the 6-stages architecture but are still available for the 1-stage architecture. On the other hand, a comparison between the symbol rate performances achieved by the 1-stage architecture and the required symbol rate imposed by the DTTB standards are provided in Table (7).

FFT Size	1-stage architecture @ 100MHz	DVB-T/H	ISDB-T	TDMB	FLO
2048	8110	4329	3861	-	-
4096	4062	2164	1926	-	1200
8192	1742	1082	962	-	-
3780	2935	-	-	1801	-

Table 7. OFDM Symbol Rate Comparison Between Studied Standards and Proposed Architecture

According to these results, we may conclude that the optimized structure is still suitable for the FFT computation of the considered standards. However a compomize is always possible by combining the two methods seen in sections 3 and 4

5. Conclusion

In this chapter, we have presented two algorithms that may be combined to compute a FFT. Depending on the size of this transform, some advantages can be exploited by taking into account a meticulous organization in the algorithms combination. Thus, an optimal reduction of the algorithmic complexity will imply an optimal use of the hardware resources. Moreover, an architecture has been proposed for the computation of these algorithms. This architecture is able to deal with a large amount of FFT sizes, decomposable in product terms that are 2,3,4,5,7 or 8. A growth either of the largest FFT size or of the number of reconfigured sizes imply the use of more memory resources which is the most delicate point of the architecture.

A major advantage of this architecture is its possibility to be dynamically reconfigured from one to another FFT size. This allows to reuse the same circuit to compute several FFT sizes within the same modulation/demodulation standard, as it is necessary with the DTMB standard or the SC-FDMA modulation. As explained in section 4, this architecture provides a high flexibility that permits to achieve the best compromize between hardware resources and computation throughput. For prototyping purposes, this architecture has been successfully simulated and implemented in a FPGA. Nevertheless, the architecture targets an implementation in ASIC circuits whose technology exhibits higher performances than FPGA. Furthermore, the field of application is wide for this architecture. As mainly expressed, the convergence of multiple standards in a same device may be performed without implementing several modulation or demodulation circuits. Another example may be to envisage an implementation of OFDM/MIMO (Multiple Input Multiple Output) systems since the number of FFTs directly varies with the number of antennas in this type of applications. Finally, these concepts may also be used in a software radio context since it deals with reconfiguration of resources according to the channel or to environmental changes.

6. References

Weinstein, S. B. & Ebert, P. M. (1971). Data transmission by frequency-division multiplexing using the discrete fourier transform, In : *IEEE Transactions on Communication Technology, vol 19, pp. 628-634*

J. Wang, Z. Yang, C. Pan, M. Han & L. Yang (2003). A combined code acquisition and symbol timing recovery method for TDS-OFDM , In : *Broadcasting , IEEE Transactions on, vol. 49, pp. 304-308*

I. J. Good (1958). The interaction algorithm and practical Fourier analysis , In : *J. R. Statist. Soc., ser. B, vol.20, pp. 361-372*

P. Duhamel & M. Vetterli (1990). Fast Fourier transforms: a tutorial review and a state of the art, In : *Signal Processing, vol. 19, pp. 259-299*

J. W. Cooley & J. W. Tukey (1965). An algorithm for the machine calculation of complex fourier series , In : *Proc. Math. Comp., pp. 297-301*

S. Winograd (1978). On computing the Discrete Fourier Transform, In : *Proc. Math. Comp., pp. 175-199*

C. M. Rader (1968). Discrete Fourier transforms when the number of data samples is prime, In : *Proc. IEEE, vol. 56, pp. 1107-1108*

G. Bruun (1978). z-Transform DFT filters and FFTs, In : *IEEE Trans. on Acoustics, Speech and Signal Processing (ASSP), vol. 26, pp. 56-63*

R. C. Singleton (1969). An algorithm for computing the mixed-radix fast fourier transform, In : *IEEE Transactions on audio and electroacoustics, vol 17, NO.2, pp. 93-103*

G. L. Demuth (1989). Algorithms for defining mixed radix FFT flow graphs, In : *IEEE Transactions on acoustics, speech and signal processing, vol 37, NO.9, pp. 1349-1358*

ETSI (2004). Digital Video Broadcasting-Terrestrial (DVB-T); Framing structure, channel coding and modulation for DTV, In : *ETSI standard, Nov. 2004*

ARIB (2001). Integrated Services Digital Broadcasting-Terrestrial (ISDB-T); specification on channel coding, framing structure and modulation, In : *ARIB standard, May 2001*

Chinese National Standard (2006). Framing Structure, Channel Coding and Modulation for Digital Television Terrestrial Broadcasting System, In : *Chinese National Standard GB 20600-2006*

Z.-X. Yang, Y.-P. Hu, C.-Y. Pan & L. Yang (2002). Design of a 3780-point IFFT processor for TDS-OFDM, In : *Proc. IEEE, vol. 48, pp. 57-61*

Altera (2009). In : *FFT mega core function user guide*

11

Fourier Transform Based Transmission Systems for Broadband Wireless Communications

Mingqi Li, Yun Rui and Zhiyong Bu
Shanghai Institute of Microsystem and Information Technology, & Shanghai Advanced Research Institute, Chinese Academy of Sciences, CAS
China

1. Introduction

In recent years, Fourier Transform (FT), as an effective signal processing technology, is more and more popularly applied to wireless communications. By the FT technologies, it can not only reduce the implementation complexity of traditional transmission systems, but also bring in some new features, thus constructing new transmission systems. The current mainstream transmission schemes utilizing DFT technologies include Orthogonal Frequency-Division Multiplexing (OFDM) [1], Discrete Fourier Transform Spread Orthogonal Frequency Division Multiplexing (DFT-S-OFDM) [2] and Filter bank modulation [3]. For OFDM systems, by an IDFT at the transmitter, the whole frequency-selective wideband channel is divided into several flat narrow band sub-channels, which is benefit to overcome the effects of multi-path in wireless channels. For DFT-S-OFDM system, the uplink transmission scheme for 3GPP-Long Term Evolution (3GPP-LTE) standard, besides the IDFT served the same function as in the OFDM systems, additional DFT processing is performed to the transmitted constellation symbols before OFDM modulation. In this way, the whole modulation method can be viewed as a DFT-based interpolation processing, and the modulated signals can be regarded as single carrier signals with low Peak-to-Average Power Ratio (PAPR) property. For filter-bank systems, the FT can be used both to reduce implementation complexity and to construct the cyclic prefix (CP) based block transmission scheme, which has merits of both the filter-bank systems with robustness against to multiple access interference (MAI) and the CP based block transmission systems with simple frequency equalization.

The chapter is organized as follows. Firstly, the implementation structure of OFDM transmitter, time-frequency description of OFDM signals, and effects of timing- and frequency-offset and channel multi-path are discussed detailed. Then, we present DFT-S-OFDM system model, describe time-frequency property of DFT-S-OFDM signals, analyze the effects of carrier frequency-offset (CFO) quantitatively and compare the SIR and PAPR performances with that of OFDM systems. Next, a DFT spread Generalized Multi-Carrier (DFT-S-GMC) system is presented. The time-frequency properties of DFT-S-GMC signal, the DFT-based implementation method and the receive SINR are addressed. Finally, conclusions are collected.

2. OFDM transmission systems

OFDM plays a significant role in modern broadband communication systems. The wireline high-speed access technology, i.e. Asymmetric Digital Subscriber Line (ADSL), was the first widely used application for the FT-based OFDM system. Several other wireless standards, such as the IEEE 802.11a Wireless Local Area Network (WLAN) and IEEE 802.16 (WiMAX) series, have adopted OFDM as a key transmission technology. IEEE 802.20 working group on mobile broadband wireless access uses OFDM as the wireless high speed transmission technology. In the area of cellular mobile communications, OFDM was also adopted as a basic downlink transmission scheme of 3GPP-Long-Term Evolution (3GPP-LTE) standard and the incoming 3GPP-LTE-Advanced standard [6]. OFDM is also widely applied in the areas of audio and video broadcasting. Digital Audio Broadcasting (DAB), initiated as a European research project in 1980s, adopts coded OFDM as the transmission technology. DVB-T based on OFDM in an 8 MHz channel is now a popular technology for terrestrial video broadcast in the world. An additional new application area of OFDM is in Ultra-Wideband (UWB) personal area networks [4].

2.1 OFDM system model

Figure 1 illustrates the principle structure of OFDM transmitter. Assume that the user-specific K data symbols are $\{a_k\}$, $0 \le k \le K-1$. After the OFDM modulation, the transmit signals can be expressed as

$$s_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{K-1} a_k exp\left(j2\pi n\left(k_0 + k\right)/N\right), \quad n = -N_g, ..., -1, 0, 1, ..., N-1 \tag{1}$$

where N_g is the CP length. To simplify the system model, the localized sub-carrier mapping is applied, i.e., $b_{k'} = \begin{cases} a_k, & k' = k_0 + k \\ 0 & otherwise \end{cases}, k' = 0, ..., N-1$, k_0 is the user-specific sub-carrier allocation offset.

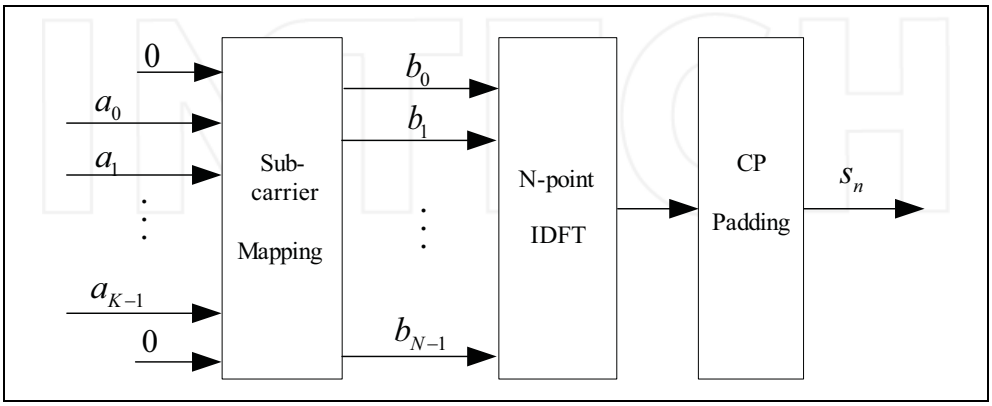


Fig. 1. OFDM transmitter

2.2 Time-frequency properties of OFDM signal

2.2.1 Time-frequency description of OFDM signal

Fig. 2 describes the time-frequency property of OFDM signal. According to the implementation principle, OFDM is a block-based transmission scheme. As a result, each OFDM symbol has a rectangle waveform in time-domain. The rectangle waveform can be expressed as

$$p(t)=\begin{cases}1 & ,|t|<T/2\\ 0 & ,\ otherwise\end{cases} \tag{2}$$

By the properties of the Fourier transform, the spectrum of each sub-carrier has a sinc-function shape in frequency-domain. Namely, the Fourier transform of $p(t)$ is $P(f)=\mathrm{sinc}(\pi Tf)/\pi f$. As it is shown, the spectrum function $P(f)$ has following properties

$$P(f)=\begin{cases}1 & ,f=0\\ 0 & ,f=n/T,(n=\pm1,2,...)\end{cases} \tag{3}$$

Equation (3) shows that $P(f),P(f-\Delta f),....,P(f-n\Delta f)$ are orthogonal to each other, and $\Delta f=1/T$. This means although the spectrum of sub-carriers are overlapped each other, the orthogonality among each sub-carriers can be maintained when the sub-carrier spacing is set to be Δf.

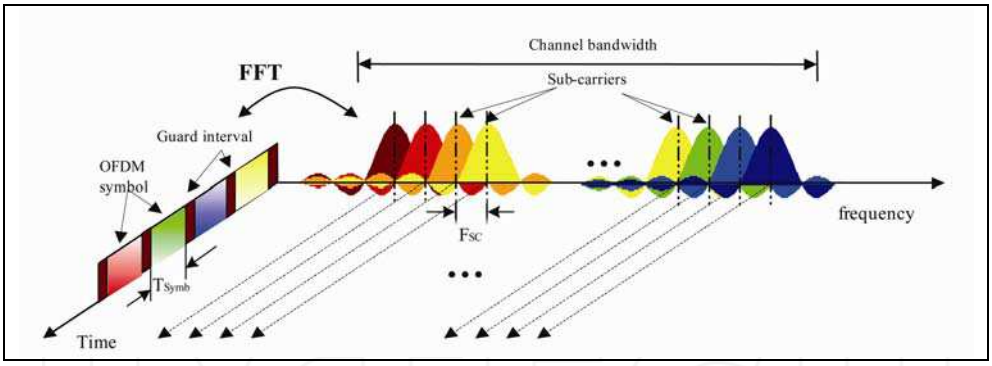


Fig. 2. The time-frequency property of OFDM signal

2.2.2 Cyclic prefix and frequency-domain equalization

When the transmitted signal in time domain is s_n and the discrete Channel Impulse Response (CIR) is $\{h_i\}_{i=0}^{K-1}$, the received signal can be given as

$$r_n=\sum_{i=0}^{K-1}h_i s_{n-i}+w_n \tag{4}$$

where w_n is the Additive White Gaussian Noise (AWGN).

According to the DFT theory, the frequency-domain multiplication is equivalent to the time-domain circular-convolution. When the signal $\tilde{s}_n$, $\tilde{s}_n=s_n, n=0,...,N-1$, is passed through

the channel, only linear-convolution with the CIR occurs. To mimic circular-convolution, we can copy a partial of samples in the tail of $\tilde{s}_n$, and pad to the head of $\tilde{s}_n$ to construct a CP. Usually the length of CP is kept greater or at least equal to the length of the channel delay spread, i.e., L. Consequently, by performing DFT to the time-domain signal r_n, we can get the frequency-domain received signal in the k-th sub-carrier

$$R_k = b_k H_k + W_k, k = 0,...,N-1 \tag{5}$$

where H_k is channel frequency response and $H_k = \sum_{i=0}^{L-1} h_i e^{j2\pi ik/N}$. W_k is the AWGN in frequency domain. b_k is the transmitted modulated symbol in the k-th sub-carrier as described in Equation (1).

To recover the transmitted symbol, we can apply single-point frequency-domain equalization to each sub-carrier. For example, using ZF equalization, the desired signal can be denoted as

$$\hat{b}_k = \frac{R_k}{H_k}, k = 0,...,N-1 \tag{6}$$

2.2.3 Effects of time and frequency offset on demodulated signal

a. Effect of timing offset

Symbol timing is one of the key factors in OFDM synchronization, which decides the accurate selection of FFT window starting position for OFDM demodulation. As shown in Fig. 3, the ideal synchronization position of symbol timing is the first sample of the transmitted signal removing the CP. If the FFT window is opened at the permitted zone, the received signal will produce no Inter-Symbol Interference (ISI), only inducing the common phase rotation on demodulated symbols. While the initial location is selected outside the permitted area, it inevitably will produce the ISI, thus resulting in the Inter-sub-carrier Interference (ICI).

Assuming over the AWGN channel, if the symbol timing position is captured in the permitted zone, then the output demodulation signal on the k'-th sub-carrier of the m-th OFDM symbol is expressed as

$$\hat{a}_{m,k'} = a_{m,k'} exp\left(-j2\pi\xi k'/N\right),\ k' = 0,\ldots,N-1 \tag{7}$$

where ξ is the timing offset. However, if the symbol timing position is outside the permitted zone, for example delay ξ samples, then the output demodulation signal on the k'-th sub-carrier of the m-th OFDM symbol can be denoted as

$$\begin{aligned}\hat{a}_{m,k'} = &\frac{N-\xi}{N} a_{m,k'} \exp\left(j2\pi\xi k'/N\right) \\ &+\frac{1}{N}\sum_{n=0}^{N-1-\xi}\sum_{k=0;k=k'}^{K-1} a_{m,k'} \exp\left(j2\pi\left(n+\xi\right)/N\right) \\ &+\frac{1}{N}\sum_{n=N-\xi}^{N-1}\sum_{k=0}^{K-1} a_{m+1,k} \exp\left(j2\pi\left(n-N-N_g+\xi\right)k/N\right)\exp\left(-j2\pi nk'/N\right)\end{aligned} \tag{8}$$

where the second item on the right side is for ICI, and the third part is for ISI. As shown in Fig. 4, the constellations of the demodulated signals are divergent besides the phase rotation.

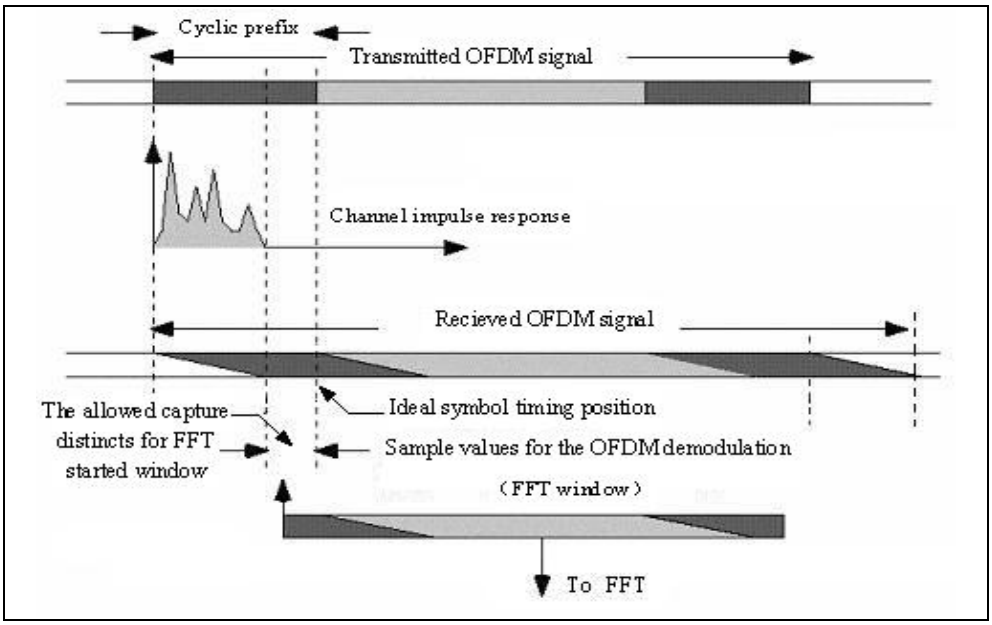


Fig. 3. OFDM symbol timing

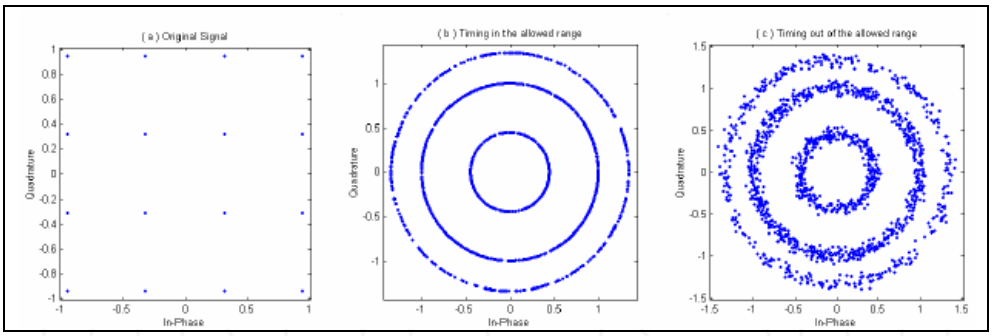


Fig. 4. The constellation of demodulated signal with timing offset

b. Effect of carrier frequency offset

In OFDM system, the existed CFO will lead to frequency shift of the received signal. If the offset of the sub-carrier frequency is integral multiple of sub-carrier spacing, the orthogonality among sub-carriers is still maintained, just with a shift relative to the sub-carrier for the data symbols. However, if the frequency offset is a fractional sub-carrier spacing, the ICI will be introduced. For OFDM systems composed of a large number of sub-carriers, sub-carrier bandwidth is relatively much smaller compared with the channel

bandwidth. Therefore, the small amount of frequency offset will result in substantial BER performance degradation.

Assuming the normalized fractional CFO is ε, and then the demodulated signal can be expressed as

$$a_{k'} = \alpha a_{k'} exp\left(j\pi\varepsilon\left(N-1\right)/N\right) + \sum_{n=0}^{N-1}\sum_{k=0;k\neq k'}^{K} a_{k'} exp\left(j2\pi n\left(k-k'+\varepsilon\right)/N\right) \tag{9}$$

where α is the attenuation experienced by all sub-carriers, and

$$\alpha = \frac{sinc\left(\varepsilon\right)}{sinc\left(\varepsilon/N\right)} \tag{10}$$

The second part on the right side of Equation (9) is an ICI item. Fig. 5 shows the demodulated signal constellation under the AWGN channel with ε = 0.05. As shown in the figure, the CFO on one hand results in the whole constellation phase rotation, and on the other hand, due to the impact of ICI, a divergent phenomenon is generated among constellation points.

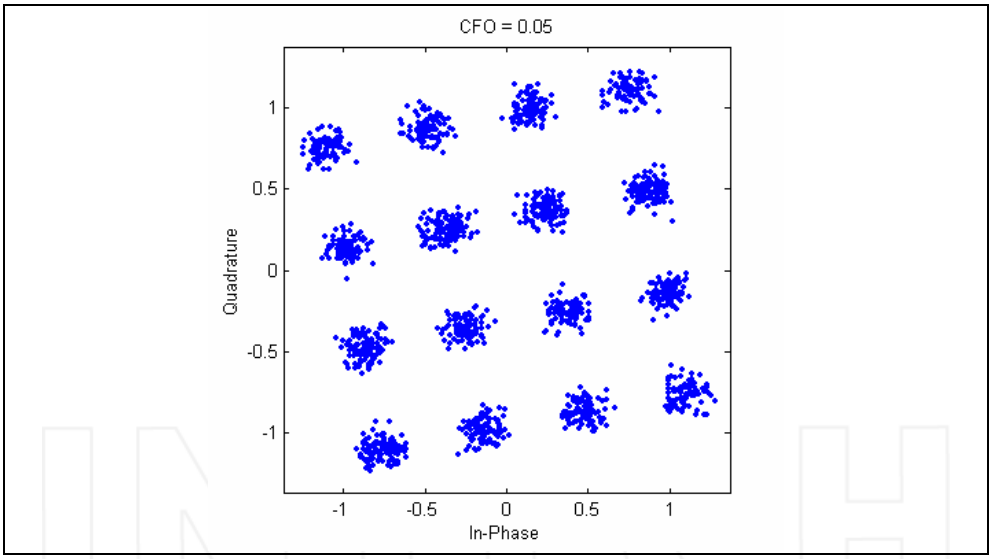


Fig. 5. The constellation of demodulated signal with CFO = 0.05

2.3 Performances analysis and numerical results

a. SIR effects caused by the carrier frequency-offset

According to Equation (9), the SIR of k'-th sub-carrier is given as

$$SIR_{k'}^{O} = \frac{\left|\alpha\right|^2}{E\left[\left|ICI_{k'}\right|^2\right]} \tag{11}$$

where $ICI_{k'}$ is the inter-sub-carrier interference on the k'-th sub-carrier, and its variance can be expressed as[15]

$$E\left[\left|ICI_{k'}\right|^2\right]=\sum_{\substack{k=0\\k\neq k'}}^{K-1}\left(\frac{\sin c\left(k-k'+\varepsilon\right)}{\sin c\left(\left(k-k'+\varepsilon\right)/N\right)}\right)^2 \tag{12}$$

b. Uncoded BER Performance of OFDM

In order to evaluate the error probability, without loss of generality, we focus on the signal received on the first sub-carrier. Moreover, let us now consider the QPSK modulation. By the conditional SINR, the approximate BER becomes [14]

$$BER\approx\frac{1}{2}Q\left(\sqrt{\frac{\frac{\sin^2\pi\varepsilon}{M^2\sin^2\frac{\pi\varepsilon}{M}}\cos^2\left(\frac{\pi}{4}+\frac{\pi\varepsilon(M-1)}{M}\right)}{\frac{1}{2}\left(1-\frac{\sin^2\pi\varepsilon}{M^2\sin^2\frac{\pi\varepsilon}{M}}\right)+\frac{1}{SNR}}}\right)+\frac{1}{2}Q\left(\sqrt{\frac{\frac{\sin^2\pi\varepsilon}{M^2\sin^2\frac{\pi\varepsilon}{M}}\sin^2\left(\frac{\pi}{4}+\frac{\pi\varepsilon(M-1)}{M}\right)}{\frac{1}{2}\left(1-\frac{\sin^2\pi\varepsilon}{M^2\sin^2\frac{\pi\varepsilon}{M}}\right)+\frac{1}{SNR}}}\right) \tag{13}$$

where M is the modulation order, and SNR is the average received signal to noise ratio. Fig. 6 shows simulation results of the impacts of the CFO on the BER performance of OFDM system with QPSK modulation over the AWGN channel.

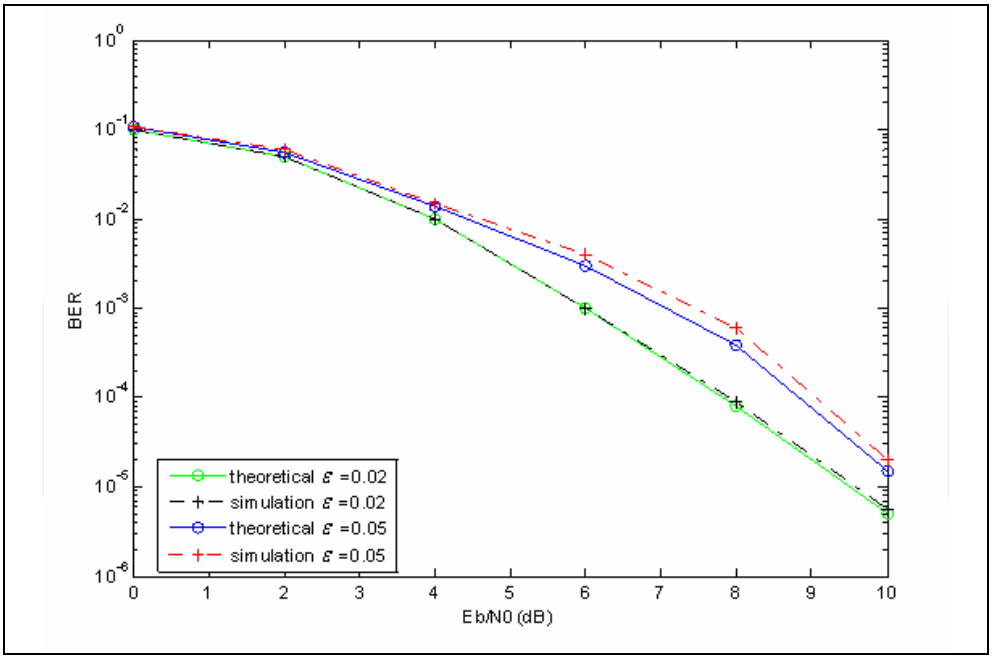


Fig. 6. BER of OFDM system with QPSK modulation under the impacts of the CFO over the AWGN channel

3. DFT-S-OFDM transmission systems

In order to meet the emerged new requirements, many powerful and implementation-efficient transmission schemes have been proposed for the standardization of the latest and future communication systems. For the downlink transmission, the OFDM scheme has been widely accepted due to its high spectral efficiency and flexible resource allocation. For the uplink transmission, however, the power efficiency is particularly critical for mobile terminals with the restriction on the transmission power and power consumption. Therefore, the PAPR performance becomes one of the most important criterions in selecting the transmission scheme for the uplink. From this point of view, the single-carrier based frequency division multiple access (SC-FDMA) scheme is favoured by future wideband wireless communications. In fact, one kind of SC-FDMA schemes, i.e., DFT-S-OFDM, is accepted as the uplink basic transmission scheme by the 3GPP-LTE standard and incoming 3GPP-LTE-Advanced standard [5][6]. In order to reduce the PAPR, the DFT-S-OFDM scheme utilizes the DFT spreading processing on the transmitted constellation symbols before OFDM modulation. By this way, the modulation method can be view as a DFT-based interpolation processing, and the modulated signals can be regarded as single carrier signals.

3.1 DFT-S-OFDM system model

The structure of DFT-S-OFDM transmitter is shown in Fig. 7. Assume the user-specific K data symbols are $\{a_m\}$, $0 \le m \le K-1$. After the DFT based spreading, the output signals can be expressed as

$$x_k = \frac{1}{\sqrt{K}} \sum_{m=0}^{K-1} a_m exp\left(-j2\pi mk / K\right), \quad 0 \le k \le K-1 \tag{14}$$

Then according to the localized allocation pattern, the signal is converted into a time-domain signal by N -point IDFT processing. N is greater than K. After the CP padding, the time-domain signal, i.e., the SC-FDMA symbol, can be written as

$$s_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{K-1} x_k exp\left(j2\pi n\left(k_0 + k\right) / N\right), \; n = -N_g, ..., -1, 0, 1, ..., N-1 \tag{15}$$

where k_0 is the user-specific sub-carrier offset and N_g is the CP length.

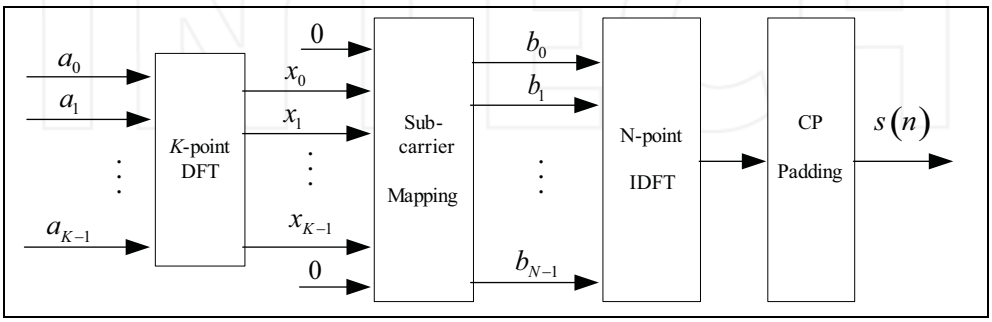


Fig. 7. DFT-S-OFDM transmitter

3.2 Time-frequency properties of DFT-S-OFDM signals

a. Time-frequency description of DFT-S-OFDM signal

Figure 8 describes the time-frequency property of DFT-S-OFDM signal. For DFT-S-OFDM signals, the time-domain waveform can be viewed as a DFT-based interpolation of transmitted constellation symbols. Therefore, the energy distribution within one DFT-S-OFDM symbol keeps the same transmission order of the constellation symbols in the time-domain. In the frequency-domain, due to the DFT based spread spectrum processing, the spectrum of each transmitted constellation symbol is distributed on all occupied sub-carriers, i.e., each sub-carrier contains only a part of spectrum component of the transmitted constellation symbol, which is substantially different form OFDM signals.

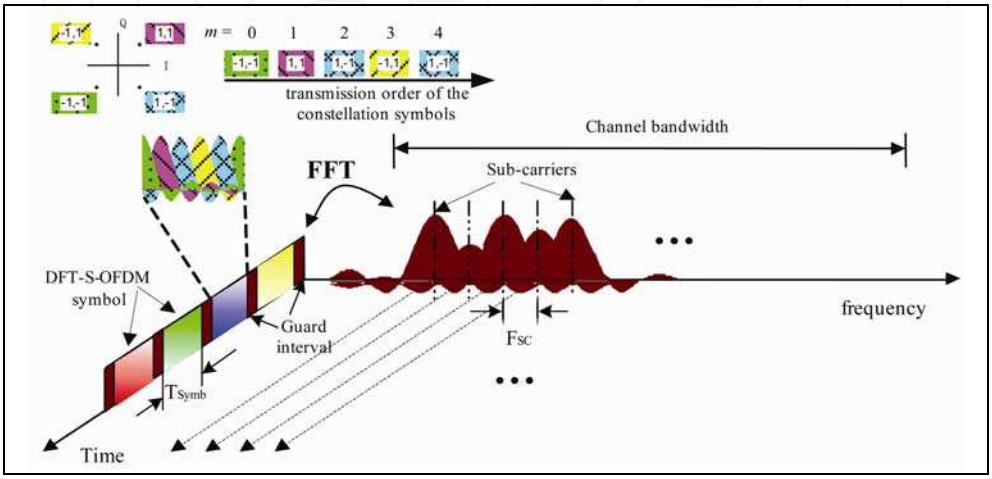


Fig. 8. The time-frequency property of DFT-S-OFDM signal

b. Effects of time and frequency offset on demodulated signal

DFT-S-OFDM systems are subjected to the CFO and TO as well. Although the DFT-S-OFDM scheme is based on the OFDM technology, the effects of CFO on two systems are very different, because the DFT-S-OFDM systems can be viewed as transmitting symbols in the time-domain, whereas the OFDM systems are usually regarded as transmitting symbols in the frequency domain.

1. Effect of time offset

First, if the timing point is allocated inside the permitted zone, similar as the analysis for OFDM system, the received signal in the frequency domain for DFT-S-OFDM can be denoted as

$$\hat{x}_k = x_k exp\left(-j2\pi\xi k / N\right), \quad k = 0,1,...,K-1 \tag{16}$$

While after the K-point IDFT despreading, then the demodulated symbol can be given as

$$\hat{a}_{m'} = \frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} \hat{x}_k exp\left(j2\pi m'k / K\right), \quad m' = 0,1,...,K-1 \tag{17}$$

From the above expression, we can see that although the timing point is allocated inside the permitted zone, it not only induce the common phase rotation, but also brings the ICI on the

demodulated symbols, which is shown in Fig. 9. This character is different from OFDM system. Furthermore, if the timing point is outside the permitted zone, it will cause both the ISI and ICI, which is similar with OFDM.

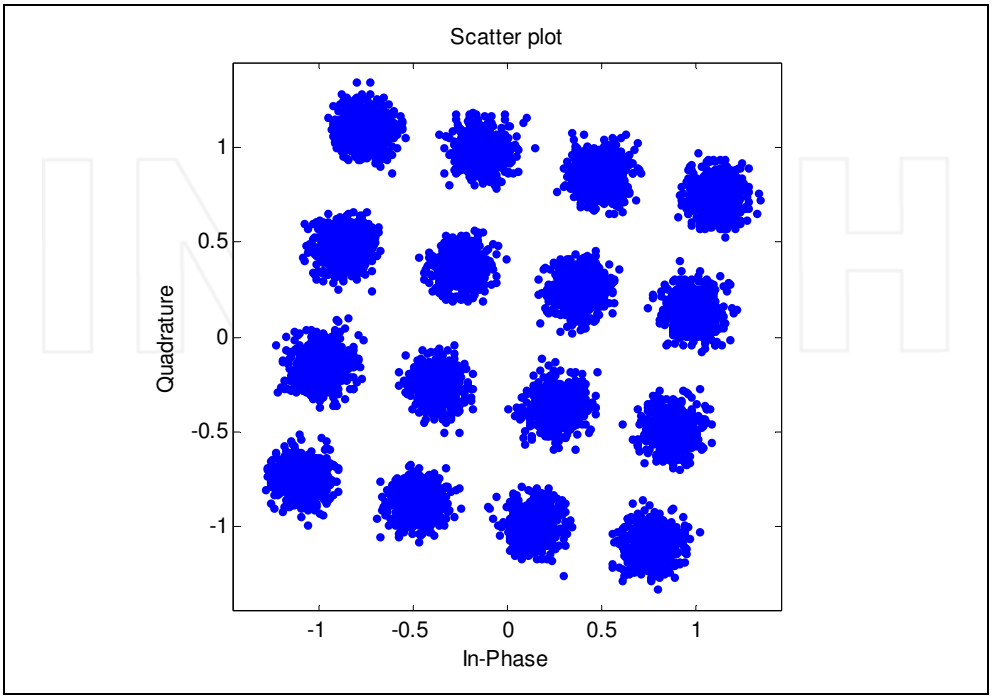


Fig. 9. The constellation of demodulated signal with timing offset

2. Effect of CFO

Considering CFO effects, following CP removing, the received N samples of SC-FDMA symbol for demodulation are

$$r_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{K-1} H_{k+k_0} x_k exp\left(j2\pi n\left(k_0 + k + \varepsilon\right)/N\right) + w_n, \quad n = 0,1,...,N-1 \tag{18}$$

where w_n is the complex-valued AWGN on the n-th time-domain sample. H_k is the channel frequency response (CFR) at the k-th sub-carrier, and $H_k = \sum_{i=0}^{L-1} h_i \cdot \exp\left(-j2\pi k\tau_i / N\right)$. By DFT and single sub-carrier equalization, the output is

$$y_{k'} = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{k=0}^{K-1} W_{k+k_0} H_{k+k_0} x_k exp\left(j2\pi n\left(k + k_0 + \varepsilon - k'\right)/N\right), \quad k' = 0,1,...,N-1 \tag{19}$$

where $x_k = \frac{1}{\sqrt{K}} \sum_{m=0}^{K-1} a_m exp\left(-j2\pi mk / K\right), \quad 0 \le k \le K-1$.

Then, after the sub-carrier demapping, generally, the K elements extracted from the N-sample output of DFT are processed by a K-point IDFT, and yields the estimated symbols

$$\hat{a}_{m'} = \frac{1}{\sqrt{K}} \sum_{k'=0}^{K-1} y_{k'+k_0} exp\left(j2\pi m'k'/K\right), \quad m' = 0,1,...,K-1 \tag{20}$$

Because the energy of ICI inducing by CFO is distributed in all the sub-carriers, as shown in Fig.10, the CFO brings about not only the ICI and the linear phase rotation, but also the ISI and the attenuation on the demodulated symbols.

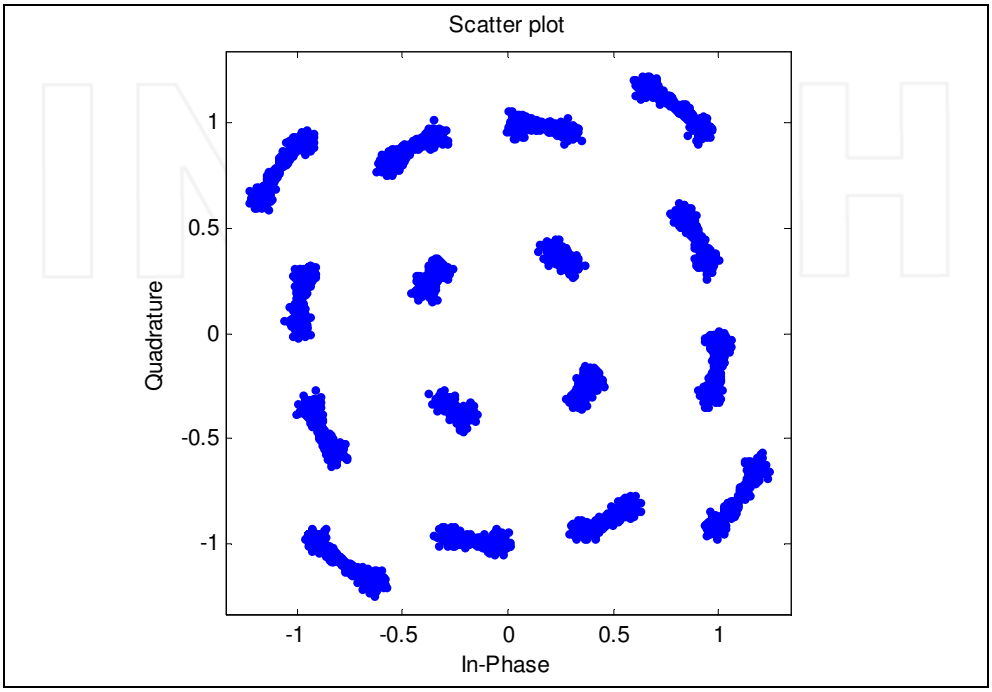


Fig. 10. The constellation of demodulated signal with CFO = 0.05

3.3 Performances analysis and numerical results

a. SIR effects caused by the carrier frequency-offset

For DFT-S-OFDM systems with CFO and over a flat fading channel, the demodulated symbol can be given as [10]

$$\hat{a}_{m'} = a_{m'}\alpha_{m'} exp\left(j2\pi m'\varepsilon/K\right) + ISI_{m'} + ICI_{m'} \tag{21}$$

where $\alpha_{m'}$, $ISI_{m'}$ and $ICI_{m'}$ are the attenuation term, inter-symbol interference and inter-sub-carrier interference on the m'-th demodulated symbol respectively.

$$\begin{aligned} \alpha_{m'} = & \frac{1}{K} \sum_{k'=0}^{N-1} \sum_{k=0}^{K-1} exp\left(-j2\pi\left(k-k'+k_0+\varepsilon\right)m'/K\right) \\ & \cdot exp\left(j\pi\left(k-k'+k_0+\varepsilon\right)(N-1)/N\right) \frac{\sin c\left(k-k'+k_0+\varepsilon\right)}{\sin c\left(\left(k-k'+k_0+\varepsilon\right)/N\right)} \end{aligned} \tag{22}$$

$$ISI_{m'} = \frac{1}{K}\sum_{k'=0}^{N-1}\sum_{k=0}^{K-1}\sum_{\substack{m=0 \\ m\neq m'}}^{K-1} a_m exp\left(-j2\pi\left(mk - m'k' + m'k_0\right)/K\right) \cdot exp\left(j\pi\left(k-k'+k_0+\varepsilon\right)\left(N-1\right)/N\right)\frac{\sin c\left(k-k'+k_0+\varepsilon\right)}{\sin c\left(\left(k-k'+k_0+\varepsilon\right)/N\right)} \tag{23}$$

$$ICI_{m'} = \frac{1}{NK}\sum_{k'=K}^{N-1}\sum_{n=0}^{N-1}\sum_{k=0}^{K-1}\sum_{m=0}^{K-1} a_m exp\left(-j2\pi mk/K\right)exp\left(j2\pi n\left(k+k_0\right)/N\right) \cdot \exp\left(j2\pi\varepsilon n/N\right)exp\left(-j2\pi nk'/N\right)exp\left(j2\pi m'\left(k'-k_0\right)/K\right) \tag{24}$$

The SIR of the m'-th demodulated symbol is

$$SIR_{m'}^{(2)} = \frac{\left|\alpha_{m'}\right|^2}{E\left[\left|ISI_{m'} + ICI_{m'}\right|^2\right]} \tag{25}$$

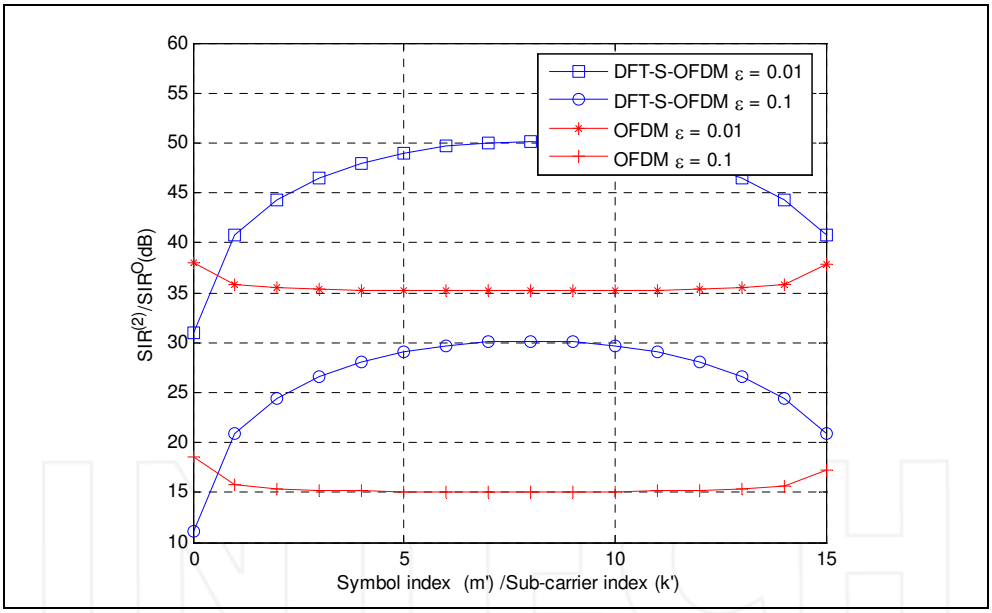


Fig. 11. SIR comparison between DFT-S-OFDM and OFDM system with CFO

As shown in the figure, except the first symbol, all other demodulated symbols of DFT-S-OFDM system have a much higher SIR than that of OFDM system under the same CFO condition.

b. Uncoded BER Performance of CFO Effect [16]

We first derive an exact closed-form BER expression for the DFT-S-OFDM system without channel coding. As we know, an arbitrary rectangular QAM can be viewed as two independent pulse amplitude modulation (PAM), i.e., I-ary and J-ary PAM's, through two

quadrature branches. As a result, the average bit probability of the detected symbol $\tilde{a}_{m',j}$ in the presence of CFO can be obtained by averaging the bit error probabilities from [7]

$$P_{m'}^{j} = \frac{1}{\log_2(I \cdot J)} \left(\sum_{k=1}^{\log_2 I} P_I(k) + \sum_{l=1}^{\log_2 J} P_J(l) \right) \tag{26}$$

where

$$P_I(k) = \frac{1}{I} \sum_{i=0}^{(1-2^{-k})I-1} (-1)^{\left\lfloor \frac{i \cdot 2^{k-1}}{I} \right\rfloor} \left\lfloor 2^{k-1} - \left\lfloor \frac{i \cdot 2^{k-1}}{I} + \frac{1}{2} \right\rfloor \right\rfloor \cdot \operatorname{erfc}\left((2i+1) \sqrt{\frac{3\log_2(I \cdot J) SINR_{m'}^{j}}{(I^2 + J^2 - 2)M}} \right) \tag{27}$$

where M is the number of bits for the special modulated symbol, and $SINR_{m'}^{j}$ denotes the SINR per modulated symbol, which can be obtained from the above subsection. In addition, $\operatorname{erfc}(\cdot)$ is the complementary error function, and $\lfloor x \rfloor$ denotes the largest integer to x. Similarly, the $P_J(l)$ can be denoted as the above. Note that for $I = 2$ and $J = 1$, equation (26) reduces to the BER of a BPSK signal.

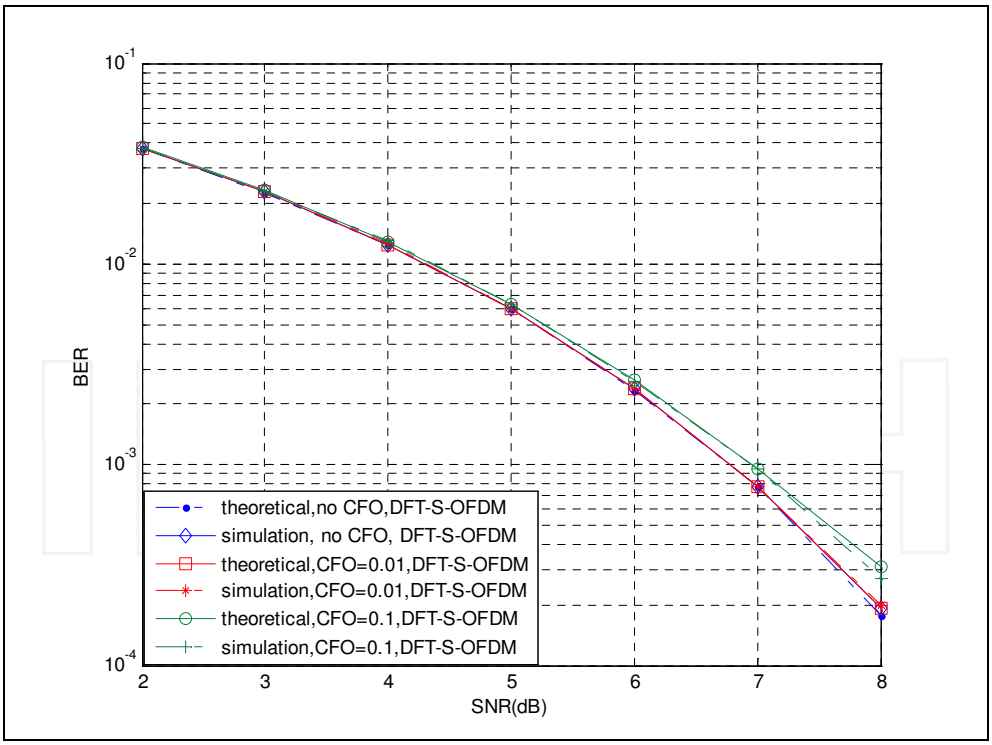


Fig. 12. BER comparison of the CFO effect for uncoded DFT-S-OFDM systems

Therefore, based on the derived SINR for the demodulated symbol in the above section, the closed-form BER expression in the presence of CFO for DFT-S-OFDM system can be given as

$$P_{DFT\text{-}S\text{-}OFDM}^{\text{uncoded}} = \frac{1}{Q}\sum_{j=0}^{N-1}\sum_{m'=0}^{D-1} P_{m'}^{j} \tag{28}$$

Then, using the average BER of the DFT-S-OFDM system, we can compute the effective SINR by the mapping function as shown in (29). For the individual modulation, such as BPSK, the effective SINR can be denoted as

$$SINR_{eff}^{uncoded} = [erfc^{-1}(P_{DFT\text{-}S\text{-}OFDM}^{uncoded})]^2 \tag{29}$$

As a result, the SNR degradation for DFT-S-OFDM system can be presented as

$$D_{\text{DFT-S-OFDM}} = 10\log_{10}\frac{E_s}{\sigma^2 \cdot SINR_{eff}^{uncoded}} \tag{30}$$

As can be seen from the figure, the theoretical result using the exact expression agrees well with the simulation results. This clearly shows that the exact expression for calculating the effective SINR in equation (29) can be used in order to assess the effect of the carrier frequency offset accurately.

c. Turbo coded BER Performance of CFO Effect

In this section, we will focus on the BER performance in the presence of CFO for the coded DFT-S-OFDM system. Due to the mathematical complexity of the iterative turbo decoding algorithm, the analytical derivation of the BER of turbo codes is not available. To simplify the analysis, the BER of coded system in AWGN channels can be approximated by an expression of the form [8]

$$P_{DFT-S-OFDM}^{\text{Coded}} = \exp(-\frac{\gamma}{\beta}) \tag{31}$$

where γ is the received effective SINR. Parameters β is mode-dependent, and can be obtained by fitting the curves to the exact simulated BER. As a result, the exponential effective SINR mapping (EESM) method is considered to incorporate the SINR of all the detected symbols. The formulation of the incorporated SINR can be expressed as [9]

$$SINR_{\text{eff}}^{\text{coded}} = -\beta\ln\left(\frac{1}{D}\sum_{j=0}^{N-1}\sum_{m'=0}^{D-1}\exp\left\{-\frac{SINR_{m'}^{j}}{\beta}\right\}\right) \tag{32}$$

Similar to (30), the SNR degradation can be also obtained.

Fig.13 shows the BER performance in DFT-S-OFDM systems for turbo coded modulation, where the QPSK modulation and 1/2 coding rate is assumed. In addition, the EESM method in (32) is used to incorporate the SINR of all the detected symbols, which is different in the presence of CFO, and β is achieved through the simulation. As we can see from the figure, the BER mapping curve for CFO $\varepsilon = 0.01$ and $\varepsilon = 0.1$ is very close to that without CFO over the equivalent SINR.

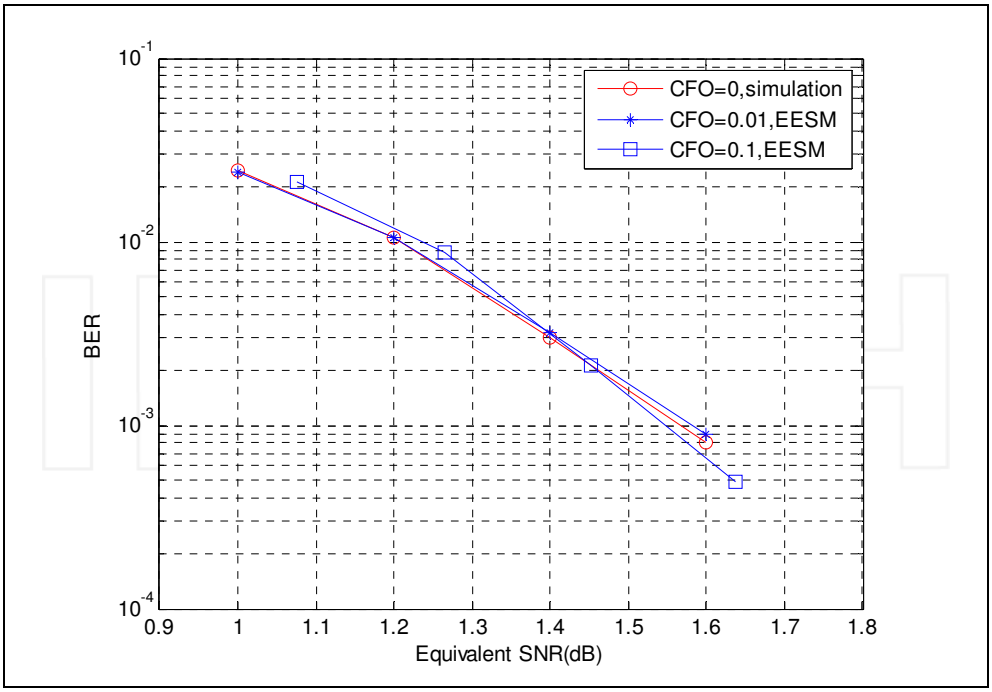


Fig. 13. BER performance with and without the CFO effect

d. Post-processing SINR of DFT-S-OFDM [11]

For DFT-S-OFDM system, the transmit signal vector without CP can be given as

$$S = F_N^H T_{N,M}^m F_M D \tag{33}$$

where $T_{N,M}^m$ is the mapping matrix for sub-carrier assignment, F_M is the M point FFT matrix and F_N^H is the N point IFFT matrix, $D = \begin{bmatrix} d_1 & \cdots & d_M \end{bmatrix}^T$ is the data vector.

Then, at the receiver, the vector of detection metric after FDE is then given as

$$\begin{aligned} \widehat{D} &= F_M^H T_{N,M}^{m\ H} H^H W (H T_{N,M}^m F_M D + Z) \\ &= F_M^H \widetilde{H}_{W,M}^m F_M D + F_M^H \overline{H}_{W,M}^m Z_M \end{aligned} \tag{34}$$

where $W = diag\{\omega_1, \omega_2, \cdots, \omega_N\}$ is the frequency domain equalizer, and

$$\tilde{H}_{W,M}^m = diag\left\{ \left|H_m\right|^2 \omega_m, \left|H_{m+1}\right|^2 \omega_{m+1}, \ldots, \left|H_{m+(M-1)}\right|^2 \omega_{m+(M-1)} \right\} \tag{35}$$

$$\bar{H}_{W,M}^m = diag\left\{ H_m^* \omega_m, H_{m+1}^* \omega_{m+1}, \ldots, H_{m+(M-1)}^* \omega_{m+(M-1)} \right\} \tag{36}$$

Hence, the SINR for DFT-S-OFDM with FDE is given as

$$SINR = \frac{\left| \frac{1}{M} \sum_{k=1}^{M-1} \left| H_{m+k} \right|^2 \omega_{m+k} \right|^2}{\frac{1}{M} \sum_{k=1}^{M-1} \left| H_{m+k} \right|^2 \left| \omega_{m+k} \right|^2 \sigma^2 + \frac{1}{M} \sum_{k=1}^{M-1} \left| H_{m+k} \right|^4 \left| \omega_{m+k} \right|^2 - \left| \frac{1}{M} \sum_{k=1}^{M-1} \left| H_{m+k} \right|^2 \omega_{m+k} \right|^2} \tag{37}$$

Furthermore, for ZF and MMES equalizer, the SINR expression can be simplified respectively as

$$SINR = \left[\frac{1}{M} \sum_{k=0}^{M-1} \frac{\sigma^2}{\left| H_{m+k} \right|^2} \right]^{-1} \tag{38}$$

and

$$SINR = \left[\frac{1}{\frac{1}{M} \sum_{k=0}^{M-1} \frac{\left| H_{m+k} \right|^2}{\left| H_{m+k} \right|^2 + \sigma^2}} - 1 \right]^{-1} \tag{39}$$

4. DFT-S-GMC transmission systems

4.1 DFT-S-GMC system model

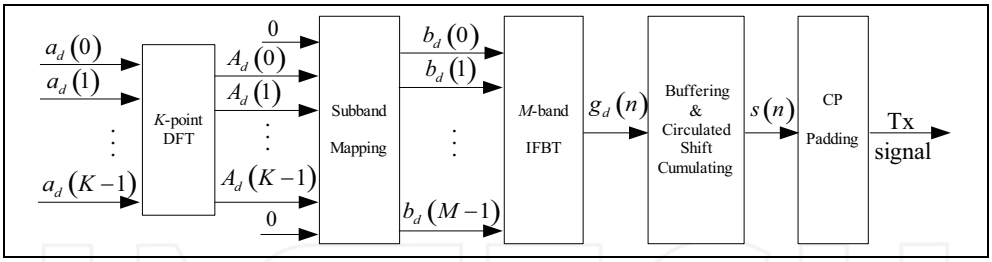


Fig. 14. DFT-S-GMC transmitter

The structure of the DFT-S-GMC transmitter is illustrated in Fig. 14 [12]. Assume that the input parallel modulated constellation symbol sequence is $\{a_d(k)\}$, $0 \le k \le K-1$ and $0 \le d \le D-1$. Note that K is the number of the user-specific occupied sub-bands, and D is the number of inverse filter-bank transform (IFBT) symbols transmitted during each DFT-S-GMC symbol.

The input data sequence is passed through K-point DFT for spectrum spreading, yields the output signal

$$A_d(k') = \frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} a_d(k) \exp\left(-j2\pi kk'/K\right), \quad 0 \le k' \le K-1 \tag{40}$$

Then, the signals are sent to the sub-band mapping module, yields

$$b_d(m)=\begin{cases}A_d(k'), & m=k_0+k'R\\ 0, & m\neq k_0+k'R\end{cases},\quad 0\le m\le M-1 \tag{41}$$

where M is the total number of sub-bands, k_0 is the user-specific sub-band offset, R is the repetition factor. For DFT-S-GMC system, both distributed and localized mapping policy could be supported, which is corresponding to R greater than or equal to one respectively.
After sub-bands mapping, an IFBT is performed on the data sequence, yields the output signal, i.e., the IFBT symbol, as

$$g_d(n)=\sum_{m=0}^{M-1}b_d(m)\exp\left(j2\pi mn/M\right)f_p(n),\quad 0\le n\le L-1 \tag{42}$$

where $f_p(n)$ is the impulse response of the prototype filter for the filter bank, and can be set to a Square Root Raised Cosine (SRRC) function. The prototype filter is with a normalized energy, i.e. $\sum_{n=0}^{L-1}\left|f_p(n)\right|^2=1$ and a length L which is integer times of M.
Then, each IFBT symbol is zero-padded to form a Q -sample data block

$$\tilde{g}_d(n)=\begin{cases}g_d(n), & 0\le n\le L-1\\ 0, & L\le n\le Q-1\end{cases} \tag{43}$$

where $Q=D\times N$.
Next, D consecutive Q-sample data blocks will be processed by buffering and cyclic-shift cumulating with shift interval N, and the output can be described as

$$s(n)=\sum_{d=0}^{D-1}\tilde{g}_d\left(\left(n-dN\right)\right)_Q,\ 0\le n\le Q-1 \tag{44}$$

where $\left((\cdot)\right)_Q$ denotes modulus operation.

4.2 Time-frequency properties of DFT-S-GMC signals

From equation (44), the DFT-S-GMC symbol is formed by cyclically accumulating several IFBT symbols in time domain, and each IFBT symbol has a SRRC waveform. Therefore, as shown in Fig. 15, the spectrum of each sub-band has a Raised Cosine function shape in frequency-domain. Moreover, the sub-band spacing is specially designed and differs from any conventional filter-bank systems in that a certain guard band is inserted between neighbouring sub-bands. By adding some guard band between the sub-bands, the near-orthogonality between neighbouring sub-bands is guaranteed, which further simplifies the detection algorithm greatly. Meanwhile, interferences among successive symbols for each sub-band could be easily mitigated in the receiver because of the shift orthogonality of prototype filter and the narrow-band single carrier transmission. Similar to DFT-S-OFDM systems, DFT-S-GMC systems also exploit DFT based spreading among sub-bands. Therefore, each sub-band contains only a part of spectrum component of transmitted constellation symbols, and transmitted signal over all occupied sub-bands can be viewed as single-carrier signal as a whole.

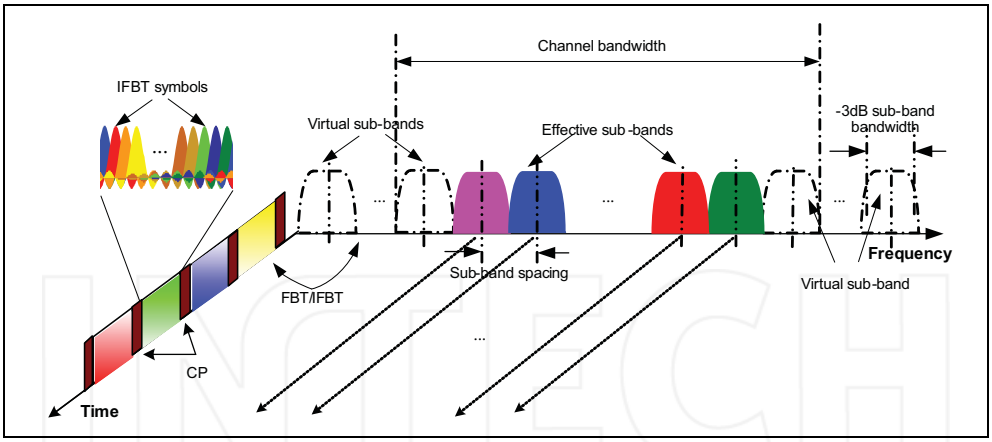


Fig. 15. The time-frequency property of DFT-S-GMC signal

4.3 Frequency-domain implementation structure

a. The frequency-domain equivalent implementation of the DFT-S-GMC transmitter

As described in equation (44), it can be seen that transmitted symbols are multiplexed within each DFT-S-GMC symbol by both time and frequency dimensions. Taking Q-point FFT on the DFT-S-GMC modulation signal, the output signal can be expressed as

$$S(q)=\frac{1}{\sqrt{Q}}\sum_{t=0}^{Q-1}\sum_{n=0}^{D-1}g_{((t-nN))_Q}(n)exp(-j2\pi qt/Q) \tag{45}$$

Since

$$\begin{aligned}&\frac{1}{\sqrt{Q}}\sum_{t=0}^{Q-1}g_{((t-nN))_Q}(n)exp(-j2\pi qt/Q)\\&\quad=\frac{1}{\sqrt{Q}}\sum_{t=0}^{Q-1}g_t(n)exp(-j2\pi qt/Q)exp(-j2\pi qn/D)\end{aligned} \tag{46}$$

$$\begin{aligned}&\frac{1}{\sqrt{Q}}\sum_{t=0}^{Q-1}g_t(n)exp(-j2\pi qt/Q)\\&\quad=\sum_{m=0}^{M-1}b_m(n)\frac{1}{\sqrt{Q}}\sum_{t=0}^{Q-1}f_p(t)exp(j2\pi mt/M)exp(-j2\pi qt/Q)\\&\quad=\sum_{m=0}^{M-1}b_m(n)F_m(q)\end{aligned} \tag{47}$$

where

$$\begin{aligned}F_m(q)&=\frac{1}{\sqrt{Q}}\sum_{t=0}^{Q-1}f_p(t)exp(j2\pi mt/M)exp(-j2\pi qt/Q)\\&,\quad 0\le q\le Q-1;\quad 0\le m\le M-1\end{aligned} \tag{48}$$

$$S(q)=\sum_{m=0}^{M-1}F_m(q)\sum_{n=0}^{D-1}b_m(n)exp(-j2\pi qn/D) \tag{49}$$

In fact, $F_m(q)$ is the frequency response of the prototype filter for m-th sub-band. Therefore, the DFT-S-GMC modulation signal can be expressed as

$$\begin{aligned} s(t)&=\frac{1}{\sqrt{Q}}\sum_{q=0}^{Q-1}S(q)exp(j2\pi qt/Q)\\ &=\frac{1}{\sqrt{Q}}\sum_{q=0}^{Q-1}\sum_{m=0}^{M-1}F_m(q)\underbrace{\sum_{n=0}^{D-1}b_m(n)exp(-j2\pi qn/D)}_{\text{Cyclic extended D-point DFT}}exp(j2\pi qt/Q) \end{aligned} \tag{50}$$

$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad}_{\text{Weighting in frequency-domain}}$$

$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}_{\text{multi-sub-bands summation}}$$

$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}_{\text{Q-point IDFT}}$$

From equation (50), the frequency-domain implementation structure of DFT-S-GMC transmitter can be given by Fig. 16.

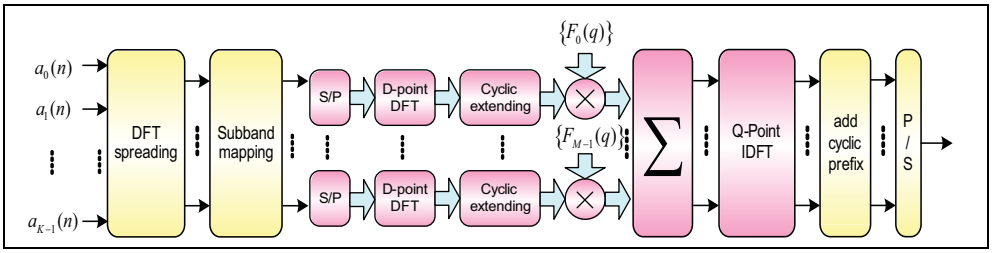


Fig. 16. Frequency-domain implementation structure of DFT-S-GMC transmitter

b. The simplified frequency-domain equivalent implementation of the DFT-S-GMC transceiver

From equation (49), the frequency response of the prototype filter for each sub-band has the energy over all Q tones. However, in fact, the most energy of the frequency response is just over several tones. Therefore, by selecting a set of values with energy greater then a proportional of the total energy of $F_m(q)$, $F_m(q)$ can be simplified as

$$\tilde{F}_m(q)=\begin{cases}F_m(q), & \sum_{q\in\Omega}\left|F_m(q)\right|^2>\xi\sum_{q=0}^{Q-1}\left|F_m(q)\right|^2, \quad q\in\Omega\\ 0, & otherwise\end{cases} \tag{51}$$

With proper designed prototype filter, the frequency response of the prototype filter for each sub-band can be simplified such that the number of total selected values for all sub-bands is equal to the number of total tones Q. By this way, the signals for each sub-band can be even mapped to the tones directly and exclusively, and the simplified implementation structure is shown in Fig.17.

As shown in Fig. 17, by tone mapping, rather than summation processing, the implementation complexity of DFT-S-GMC transmitter can be significantly reduced, and the performance loss is very limited as illustrated by the following simulation results.

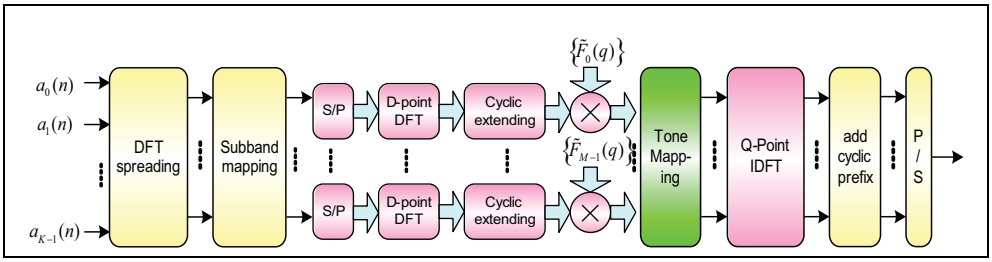


Fig. 17. Simplified implementation structure of DFT-S-GMC transmitter

4.4 Performances analysis and numerical results

a. Post-processing SINR with frequency-domain equalization

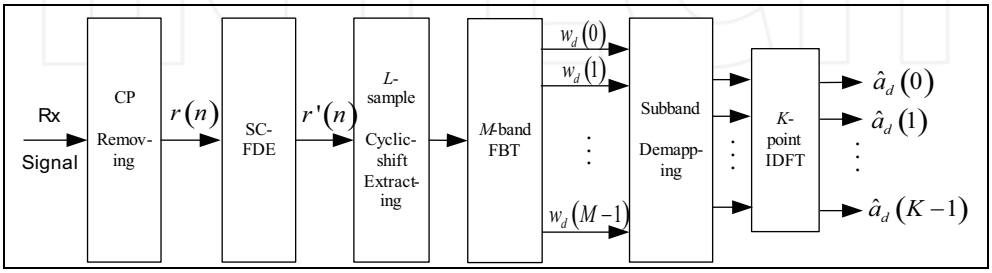


Fig. 18. DFT-S-GMC receiver

At the receiver side, shown in Fig. 18, after removing the CP, and going through the Q-point SC-FDE, the output signal vector can be given as [13]

$$\mathbf{r} = F_Q^H H^H W H F_Q \Omega_{Q,L} \Upsilon_L \Gamma_{L,M} F_M^H T_{M,K} F_K \boldsymbol{\alpha} + F_Q^H H^H W F_Q \mathbf{z} \tag{52}$$

Then, the post-processing SINR can be given by

$$\begin{aligned} SINR &= \frac{E_s}{\sigma_n^2 + \sigma_{ISI}^2 + \sigma_{ITSI}^2} \\ &= \frac{\left| \frac{1}{K} \sum_{k=0}^{K-1} \tilde{H}_k \right|^2}{\frac{\sigma^2}{K} \sum_{k=0}^{K-1} \bar{H}_k + \frac{1}{K} \sum_{k=0}^{K-1} \left| \tilde{H}_k \right|^2 - \left| \frac{1}{K} \sum_{k=0}^{K-1} \tilde{H}_k \right|^2 + \frac{1}{K} \sum_{d=1}^{D-1} \sum_{k=0}^{K-1} \left| \breve{H}_{d;k} \right|^2} \end{aligned} \tag{53}$$

where σ_n^2 is variance of AWGN on demodulated symbols. σ_{ISI}^2 is the variance of the inter-symbol-interference (ISI) within the d-th IFBT symbol, and can be expressed as $\sigma_{ISI}^2 = \sum_{k=1}^{K-1} \left| h_k \right|^2 = \frac{1}{K} \sum_{k=0}^{K-1} \left| \tilde{H}_k \right|^2 - \left| \frac{1}{K} \sum_{k=0}^{K-1} \tilde{H}_k \right|^2$. σ_{ITSI}^2 is the variance of the inter-IFBT symbol interference (ITSI) on the demodulated symbols, and can be expressed as $\sigma_{ITSI}^2 = \frac{1}{K} \sum_{d=1}^{D-1} \sum_{k=0}^{K-1} \left| \breve{H}_{d;k} \right|^2$.

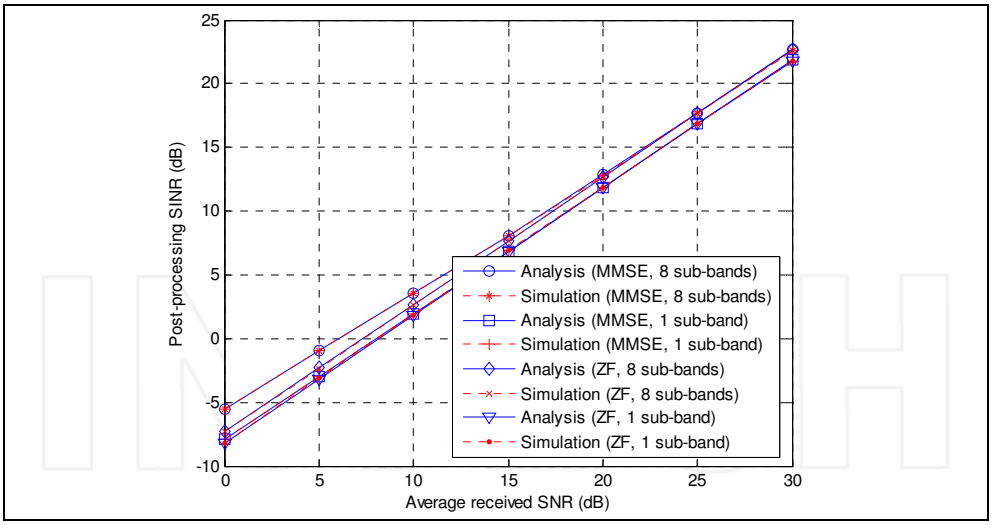


Fig. 19. The post-processing SINR of DFT-S-GMC with SC-FDE

Both the theoretical and simulated post-processing SINR are shown in Fig.19 for the DFT-S-GMC receiver with MMSE and ZF SC-FDE respectively. Over the same channel condition, with MMSE equalization, DFT-S-GMC receiver achieves higher SINR than that with ZF equalization in the low SNR range and with wider band transmission, due to the noise enhancement effects of ZF SC-FDE.

b. Performance of frequency-domain implemented DFT-S-GMC transceiver

System Parameters	Value
Carrier frequency (GHz)	2
Carrier Bandwidth (MHz)	5
Sampling frequency (MHz)	5.6
# of total sub-bands (M)	28
# of useful sub-bands	24
Upsampling rate (N)	32
# of IFBT symbols in each data block (D)	16
Sub-band BW (kHz)	200
Sub-band 3dB-BW (kHz)	175
Occupied BW (MHz)	4.8
FFT size for FDE (Q)	512
Prototype filter type	SRRC
Simulation parameters	
Channel model	PB (3km/h)
Channel coding(coding rate)	Turbo (1/2)
Modulation	QPSK
Equalization	MMSE
# of Tx/Rx antennas	1 / 1

Table 1. Simulation specification

The simulation specification is shown in Table 1. As shown in the Fig.20, the BER performance of frequency-domain implemented DFT-S-GMC transceiver is almost the same as that of time-domain implemented DFT-S-GMC transceiver.

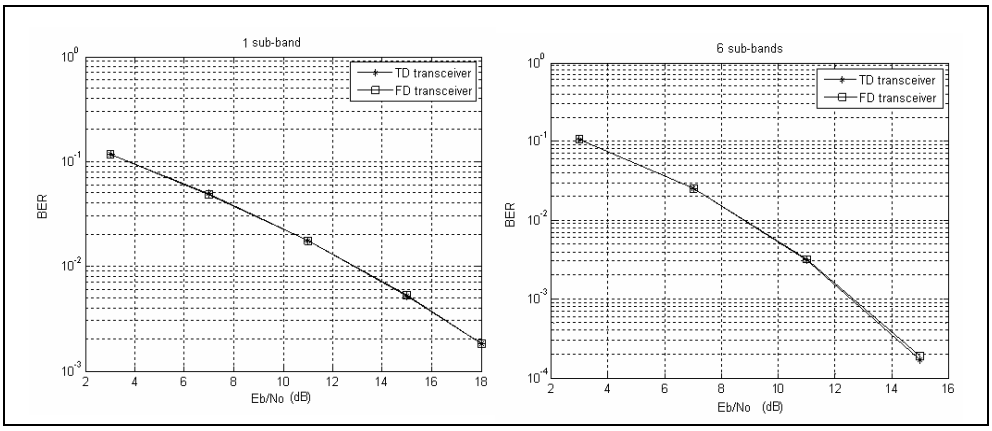


Fig. 20. BER performance comparison of FD transceiver with TD transceiver

As presented in the Table 2, the CM performance of frequency-domain implemented DFT-S-GMC transmitter is very close to that of time-domain implemented DFT-S-GMC transmitter. Moreover, the CM of DFT-S-GMC is smaller 1.7 and 1.1dB than that of OFDM for QPSK and 16QAM modulation respectively.

	Systems					
	Frequency-domain DFT-S-GMC		Time-domain DFT-S-GMC		OFDM	
Used sub-carriers /sub-band(s)	QPSK	16QAM	QPSK	16QAM	QPSK	16QAM
16 / 1	0.5	1.5	0.5	1.5	3.2	3.2
32 / 2	1.4	2.0	1.3	2.0	3.3	3.3
64 / 4	1.6	2.2	1.6	2.2	3.4	3.4
256 / 16	1.7	2.3	1.7	2.3	3.4	3.4

Table 2. Cubic Metric (dB) comparison of frequency- and time-domain implemented DFT-S-GMC system with OFDM system

# of used sub-band(s)	Computation complexity (# of real multiplications)		Percentage of reduction
	TD transmitter/ receiver	FD transmitter/ receiver	
28	18800/58688	15184/31008	19%/47%
1	18432/55404	9416/18868	49%/66%

Table 3. Complexity comparison of frequency- and time-domain implemented DFT-S-GMC transceiver

As shown in the Table 3, with frequency-domain implementation structure, the computation complexity of DFT-S-GMC transceiver with equalization can be reduced significantly, compared with that with time-domain implementation structure. For 28 and 1 sub-band(s) transmission, the computational complexity can be reduced about 47% to 66%.

5. Conclusion

In this chapter, the principle, implementation structure, time-frequency property of three Fourier Transform-based transmission systems, namely OFDM, DFT-S-OFDM and DFT-S-GMC, are presented for broadband wireless communications. For OFDM systems, the spectrum of each sub-carrier has a sinc-function shape, spectrums of all sub-carriers are independent each other which cause high PAPR of transmitted signal; For DFT-S-OFDM systems, each sub-carrier contains only a part of spectrum component of transmitted constellation symbols, and the time-domain waveform can be viewed as a DFT-based interpolation of transmitted constellation symbols, which bring in lower PAPR of transmitted signal; For DFT-S-GMC systems, each DFT-S-GMC symbol is formed by cyclically accumulating IFBT symbols with SRRC waveform in the time domain, hence, the spectrum of each sub-band has a Raised Cosine function shape, and due to DFT based spreading among sub-bands , the transmitted signal over all occupied sub-bands can be viewed as single-carrier signal as a whole. Moreover, the effects of time and frequency offset on OFDM and DFT-S-OFDM systems are analyzed quantitatively. Theoretical analysis and simulation results show that except the first symbol, all other demodulated symbols of DFT-S-OFDM system have a better SIR than that of OFDM system under the same CFO condition. Furthermore, the post-processing SINR of DFT-S-OFDM and DFT-S-GMC are addressed for different equalizer. The closed-from expressions of SINR are presented and verified by the simulation results.

6. References

[1] S,B ; Weinstein and Ebert, P.(1971). Data transmission by frequency-division multiplexing using the discrete Fourier transform. *IEEE Transactions on Communications*, Vol. 19, No.5, (Oct. 1971) 628-634, ISSN: 0018-9332

[2] Myung, H.; Lim, J.; and Goodman, D.(2006). Single Carrier FDMA for Uplink Wireless Transmission. *IEEE Vehicular Technology Magazine*, Vol. 1, No. 3, (Sep. 2006) 30-38, ISSN: 1556-6072

[3] Cherubini, G.; Eleftheriou, E.; Cioffi, J.; Oker, S.(2000). Filter bank modulation techniques for very high speed digital subscriber lines. *IEEE Communications Magazine*, Vol. 38 , No. 5, (May 2000) 98-104, ISSN: 0163-6804

[4] Qiu, RC.; Liu, H.; Shen, X. (2005). Ultra-Wideband For Multiple Access Communications. *IEEE Communications Magazine,* Vol. 43, No.2, (February 2005) 80-87, ISSN: 0163-6804

[5] 3GPP TS 36.211 v8.7.0. Evolved Universal Terrestrial Radio Access (E-UTRA); *Physical Channels and Modulation(Release 8)*，May 2009.

[6] 3GPP TR 36.913 V8.0.1（2009-03） 3rd Generation Partnership Project, Technical Specification Group Radio Access Network, Requirements for further advancements for Evolved Universal Terrestrial Radio Access (E-UTRA) (LTE-A) (Release 8).

[7] Yoon, D. and Cho, K.(2002) . On the general BER expression of one- and two dimensional amplitude modulations. *IEEE Transactions on Communications*, Vol. 50, No.7, (July 2002) 1074–1080, ISSN: 0018-9332

[8]Liu, Q.; Zhou,S.; and Giannakis, G.(2004). Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links. *IEEE Transactions on Wireless Communication.*, Vol. 3, No.5, (sept. 2004) 1746-755, ISSN: 1536-1276

[9] 3GPP TSG-RAN1 WG1 \#36, R1-040189, New results on realistic OFDM interference, Malaga, Spain, Feb. 16-20, 2004.

[10] M, Li., Y, Rui. (2009). Analysis of CFO effects on and phase compensation method for SC-FDMA systems, *Science in China Series F: Information Sciences*, Vol.52, No. 12, (Dec. 2009) 2397-2405, ISSN: 1674-733X

[11] 3GPP TSG-RAN1 WG1 \#42, R1-050719, Simulation Methodology for EUTRA UL: IFDMA and DFT Spread-OFDMA, London, U.K., Aug. 29 – Sept. 2, 2005.

[12] Zhang, X. ; Li, M.; Hu, H.; Wang, H.; Zhou, B.; You, X.(2006). DFT Spread Generalized Multi-Carrier Scheme for Broadband Mobile, *IEEE PIMRC Communications.* pp. 1-5, ISBN: 1-4244-0329-4, Helsinki , Sept. 2006.

[13] Li, M. ; Zhang, X.(2009). Performance analysis of DFT spread generalized multi-carrier Systems, *Science in China Series F: Information Sciences,* Vol.52, No.12, (Dec. 2009) 2385-2396, ISSN: 1674-733X

[14] Liu, P.; Bar-Ness, Y.(2006). Comparing the effect of Carrier Frequency Offset on OFDM and Single-Carrier Block Transmission in AWGN Channels, in *Proc IEEE GLOBECOM* 2006, pp. 1-5, ISBN: 1930-529X, San Francisco, Nov. 2006.

[15] Moose, Paul H.(1994). A technique for orthogonal frequency division multiplexing frequency offset correction, *IEEE Transactions on communications*, Vol.42, No.10, (Oct. 1994) 2908-2914, ISSN: 0090-6778

[16] Rui, Y.; Hu, H.; Li, M.; et al.(2009). Comparing Effects of Carrier Frequency Offset on Generalized Multi-carrier and OFDM Systems, *in Proc IEEE ICC* 2009, pp. 1-6, ISBN: 1938-1883, Dresden, Jun. 2009

12

Applications of Fiber Optic Coupled-Grazing Angle Probe Reflection-Absorption FTIR Spectroscopy

Oliva M. Primera-Pedrozo, Leonardo C. Pacheco-Londoño and Samuel P. Hernandez-Rivera
ALERT DHS Center of Excellence for Explosives
Center for Chemical Sensors Development
Department of Chemistry
University of Puerto Rico-Mayagüez
PO Box 9000, Mayagüez, PR, 00681
Puerto Rico

1. Introduction

The recent increase in attention of detection of chemical threats, explosives and narcotics has led to the development of instruments and sensors that can be effective in a variety of operating environments. Various approaches can be used for in situ analysis of explosives, including the widely used technique, Ion Mobility Spectrometry (IMS). The major advantages of IMS are its sensitivity in the picogram range, its continuous real time monitoring capability, its reasonable price due to instrumental simplicity and the ease of automation (Salleras, 1995). A main disadvantage of IMS is its limited linear range and that it cannot be used for quantitative analysis (Salleras, 1995). It is relatively easy to overload an IMS and, therefore, sample size must be controlled with care (Brambilla, 1997). Another weakness is the response variation that occurs with different background gas compositions and with different sample compositions (Salleras, 1995). However, spectroscopic techniques have the potential to afford the best selectivity for explosives. The infrared spectra of molecules can provide an information-rich fingerprint that allows for near unambiguous identification. A few years ago, direct detection by infrared absorption spectroscopy was not possible because of the limited sensitivity of this method. Fourier Transform Infrared Reflection Absorption Spectroscopy (IRRAS), operating at the grazing-angle, is the most sensitive optical absorption technique available for measuring low concentrations of chemical compounds adhered to reflective surfaces such as metals (Griffiths, 1986). The disadvantage of conventional spectroscopic techniques for applications such as explosives detection is that the test materials must be placed physically within the spectrometer's sample compartment for measurement. FT-IRRAS combined with grazing angle probe (GAP) can now be used outside the boundaries of the sample compartment. Fiber-optic cables (FOCs) that transmit in the mid-IR (MIR) range have made it possible to develop a range of spectroscopic probes for in situ analysis (Melling, 2001; Melling, 2002; Mehta, 2003; Bacci, 2001). Thus, FTIR spectroscopy can now be effectively used outside the confinement

of the sample compartment, making it available for field work (Mizaikoff, 2002). The attractive features of this technique include portability, simple design and rugged design, high sensitivity and short analysis time. These features lead to potential uses of Mid FTIR for airport screening and within the military.

Sample preparation methodology constitutes a critical a step in the detection scheme pursed because a uniformly covered, thin film is needed for preparation of the standard samples on which precision relies on. Several approaches have been used in the lab for sample preparation including: standard preparation using an airbrush aerosol spray, sample smearing and direct transfer from solutions using micropipettes (Primera-Pedrozo, 2004). In the smearing method 20 μL aliquots of standard solutions were deposited over the plate then the solution is spread using a Teflon sheet. Smearing transfer method has led to detect TATP over stainless steel surface. This transfer method let to detect and quantify TATP for first time on metallic surfaces. Despite the fact of tendency towards sublimation of TATP, a limiting value of 8 μg/cm^2 could be detected. Samples ranging from micrograms/cm^2 to nanograms/cm^2 of 2,6-dinitrotuelene (DNT), 2,4,6-trinitrotoluene (TNT), pentaerythritol tetranitrate (PETN), nitroglycerine (NG) and triacetone triperoxide (TATP) have been detected using this new method of deposition. A smearing deposition was used for depositing the target explosives over substrates to be used as standards. The sample transfer method gave good sample distribution, reduced sample loss on transfer and was easy to manipulate giving good reproducible distributions (Primera-Pedrozo, 2004; Primera-Pedrozo, 2009; Primera-Pedrozo, 2010; Pacheco-Londoño, 2010). Although, the smearing technique has given good results for explosives detection, many samples are needed for sample preparation transfer method because it depends on human error since the sample is placed using a piece of Teflon sheet and is distributed with the hands on the surface and sometimes good distribution is not found. In this case other samples have to be prepared. Another disadvantage of this transfer method is solvent interference since various solvents are adhered to the stainless steel plated producing poor distribution of the material over the surfaces. The use of slow evaporating solvents makes the sample preparation more complicated. For these reasons the development of an automatic method for explosives transfers on the surfaces must be devised. Thermal ink jet (TIJ) was selected as transfer technique to avoid human errors during preparation of standards and to decrease the time for sample preparation. In thermal ink jet a thin film resistor superheats less that 0.5% of the fluid in the chamber to form a gas bubble. This bubble rapidly expands (less than ten microseconds) and forces a drop to be ejected through an orifice (Beeson, 1998).

When comparing the inkjet based method of sample transfer to the smearing method used in previous works (Primera-Pedrozo, 2004), the former has notable advantages. The loading concentration of the sample on the surface can be controlled by varying parameters such as: number of passes, dispensing frequency, applied energy, and pen architecture. Precise delivery of the number of droplets with known volume and concentration controls the mass deposited. Also only one solution needs to be used, avoiding dilutions that can increase the analytical errors.

2. Description of methodology

FOC-GAP FTIR spectroscopy has made possible to develop new methods for detection of traces of chemical compounds on surfaces. Thermal inkjet (TIJ) technology is able to deposit very small amounts of chemical compounds, including energetic materials, in a specific

location on a surface (Primera-Pedrozo, 2005). Aliquots of TNT and RDX solutions were deposited on stainless steel film. A thin coating of the explosives can be produced by controlling the concentration of TNT or RDX, the number of drops dispensed and the distribution of drops on the surface. A Vector 22, a Bruker Optics FTIR fiber coupled to a Remspec Corp. Grazing Angle Probe head was used for the experiments. The spectra were recorded at 4 cm^{-1} resolution and 50 scans. The results of the experiments gave intense absorption bands in the fingerprint region of the infrared spectra that were used for quantification. Chemometrics routines were applied for enhancing quantitative analysis.
The sample analysis setup is schematically presented in Fig. 1. A Remspec mid-IR grazing angle probe was used to collect the spectra. The grazing-angle head uses carefully aligned mirrors to deliver the mid-IR beam to the sample surface at the grazing angle (approximately 80° from normal), to collect the reflected beam, and to return it to a mid-IR detector (in this case, external liquid nitrogen cooled MCT detector). The signal is delivered from the spectrometer to the head by IR transmitting fiber optic cables. The grazing angle accessory is connected to the external beam port of the Bruker Vector 22 spectrometer by a 1.5 m, 19–fiber chalcogenide glass optical bundle in the As-Se-Te system, which transmits throughout the mid–IR with the exception of a strong H-Se absorbance band at 2200 cm^{-1}. The IR footprint produced by the grazing angle probe is elliptical with the intensity decaying from the middle towards the edges. The specially configured head illuminates a large spot on the sample surface. The spot is an ellipse 1 inch by six inches that is defined by a Gaussian distribution with a center spot about 1/8 inch by an inch. The electric signal from the MCT is delivered to the FTIR using an amplifier.

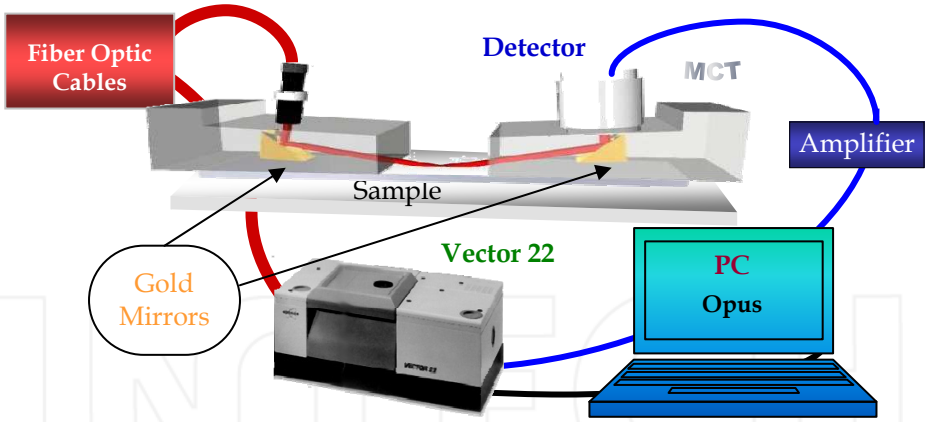
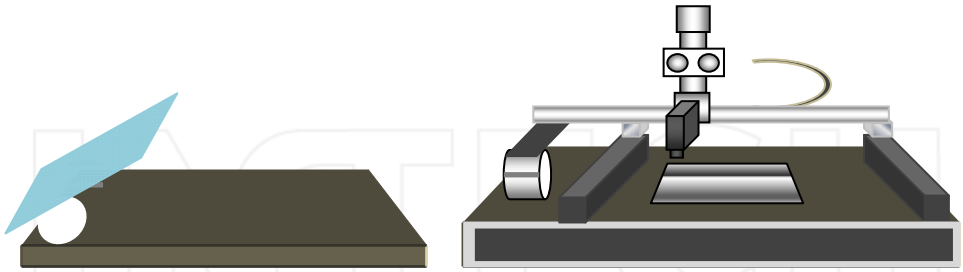


Fig. 1. Experimental setup for Fiber Optic Coupled-Grazing Angle Probe FTIR

3. Evaluation of samples and standards

FOP-GAP FTIR absorption-reflection spectroscopy may be successfully implemented in assessing sample loading distribution of solids deposited on substrates (Primera-Pedrozo, 2008; Primera-Pedrozo, 2009). Two methods were used to prepare the samples by depositing the analytes onto the test surfaces: sample smearing and thermal inkjet transfer. Stainless steel (SS, non-magnetic, type 316) metal sheets with an effective or area for coverage of 46.3 cm^2 (3 cm x 15.4 cm) were used. The plates were cleaned with HPLC grade methanol and

air-dried at room temperature before the experiments. Aliquots of 20 μL of standard solutions were placed at one side of the SS plate. A Teflon sheet was inclined towards the right or left and the smearing was done quickly. Fig. 2-a illustrates how the smearing is done.



(a) (b)

Fig. 2. Methods used for transferring a solid sample onto a substrate: (a) TIJ; (b) sample smearing

This method is rapid and easily executed without specialized equipment. The amount is readily controlled and can be calculated without the need for an independent analysis. Once the solvent had evaporated, the spectrum of the sample was collected immediately. Solutions were dispensed using an ImTech Imaging System model I-Jet 312S, (ImTech, OR, USA) equipped with a HP 51645A inkjet cartridge, illustrated in Fig. 2-b. Aliquots of 10 mL were placed into the inkjet cartridge and the backpressure was set to 3 inches of water using an external backpressure controller. The solutions were dispensed over stainless plates at zero dot spacing (space between drops using HP ink) using a printing resolution of 600 dots per inch (dpi). Once the solvent had evaporated, the spectrum of the sample was collected.

4. Cleaning validation of pharmaceutical batch reactors

FOC-GAP FTIR spectroscopy can be used in cleaning validation applications for active pharmaceutical agents on metallic surfaces (Mirza, 1999); Mehta, 2002; Primera-Pedrozo, 2005-b; Fierro-Mercado, 2010). A method based on smearing a known amount of the sample in solution was used for preparing samples and standards to develop cleaning validation methodologies using IR spectroscopy. The samples were deposited on the surface using smearing transfer method. Using this method ibuprofen was detected on stainless steel plates, a common material on pharmaceuticals reactors. This new technology combining to smearing can decrease the consuming time in cleaning validation process, being advantageous in an in process laboratory. Detection limit for this compound was 0.5 $\mu g/cm^2$ loading concentration. Grazing angle spectra of samples were collected for surface concentrations in the range of 0.1 – 20 $\mu g/cm^2$ FT-IR spectra were collected from each plate using the grazing angle probe. The spectra are shown in Figure 3. Fingerprint signals intensities of the spectra decrease with lower ibuprofen loading concentrations. This amplified region was used for the chemometrics calculations (Beebe, 1998). The most intense band of ibuprofen in the region of 1760-1650 cm^{-1} was used for peak area and peak height calibration curve generation. This band is assigned to C=O stretch (Griffith, 1986; Lin-Vien, 1991).

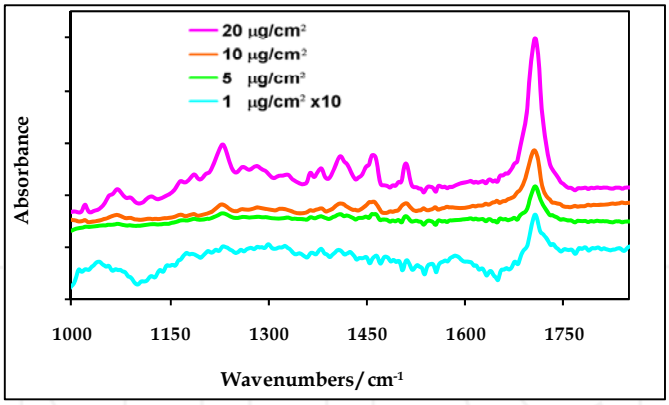


Fig. 3. Grazing angle FTIR spectra of ibuprofen on stainless steel for various surface loadings

For quantification studies, two types of calibration curves were generated using two methods: measurement of the absorbance peak heights and integration of areas spectral region within the 1760 to 1650 cm^{-1} range. Fig. 3 shows the calibration curves of the absorbance peak heights. Results for peak areas are not shown. These plots exhibit a high degree of linear correlation. The calibration curve graph using height peaks shows better in linearity. However, the errors are higher for the calibration curve using peak areas. The calibration graphs using peak areas represents a better choice for quantitative analysis when compared to peak height analysis (Lavine, 2002; Kramer, 1998).

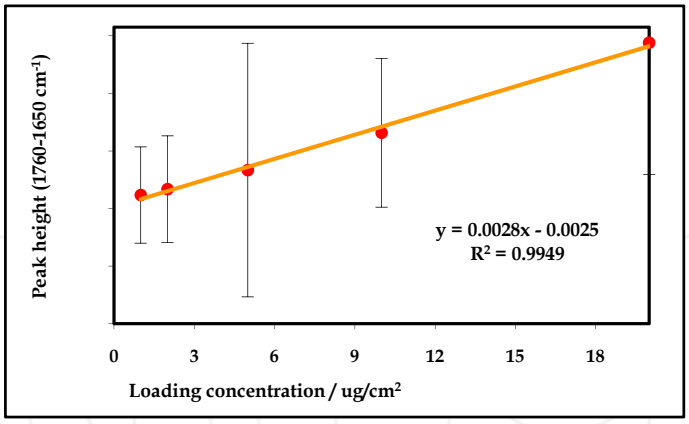


Fig. 4. Calibration curve of ibuprofen using peak height analysis

The calibration model was built using the Quant 2 package, an add-on software package to the OPUS™ (Bruker Optics) data acquisition and analysis software. In this study, the model parameters were optimized in the spectral region 1770 – 1016 cm^{-1} and 3104 – 2750 cm^{-1}. No spectral data preprocessing was done. The resulting model was cross-validated using the ''leave one out'' method in which each spectrum is omitted in turn from the training set and then tested against the model built with the remaining spectra. The results are illustrated

graphically in Fig. 5. The root mean square error of the cross validation was 0.401, and R^2 was 0.9952. Clearly, low levels of ibuprofen can be detected and measured on a metal surface with quantitative results.

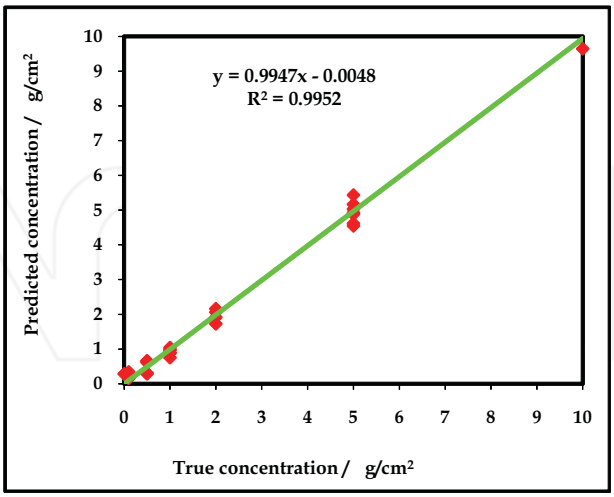


Fig. 5. Cross validation for Ibuprofen quantification on metallic surfaces

Discriminant analysis (Huberty, 1994) was also performed to classify ibuprofen loading concentration in two groups (Fig. 6). The first one corresponds to concentrations < 2 µg/cm² and the second one to concentrations > 2 µg/cm². Peak areas of signals in the range of 1273 – 1978 cm^{-1} were used for the discrimination. This model was generated using 10 PLS and submitting the data to a pre-processing of straight line subtraction. Results show that discriminant analysis can be used to classify ibuprofen samples according to their surface concentration on the metal surfaces. The minimum amount of this API on the reactor after cleaning must be considered for future works. This will allow having a real model of discrimination.

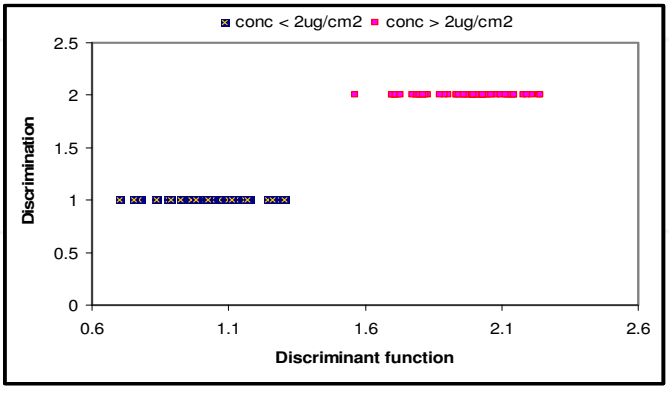


Fig. 6. Discrimination study for ibuprofen: samples were separated according to surface loadings

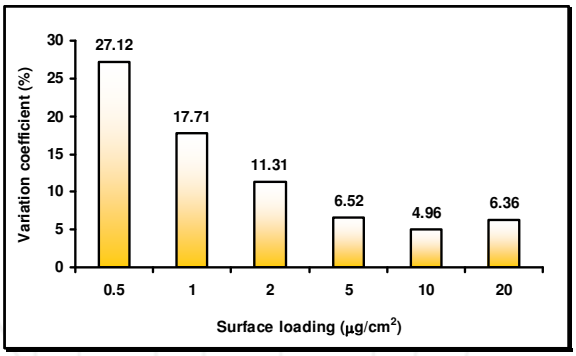


Fig. 7. Variation coefficient with surface loadings for detection and quantification limit calculations

The variation coefficient (Fig. 7) increases at low surface loadings. For 0.5 µg/cm² loading concentration, the variation was higher than the others. This indicates that at this surface loading, the analytical response gets its minimum value and can be confused with the noise. This value can be considered as the detection limit. In order to verify this, an ANOVA test was performed for surface concentrations of 0.5, 1 y 2 µg/cm², and for a 95.0% confidence level there was a statistically significant difference between these values. However at 99.0% confidence level, there was no difference for samples of 0.5 and 1 µg/cm² (Statgraphics Plus for Windows™, 1999).

5. Detection of explosives

FOC-GAP-FTIR spectroscopy is suitable for development of methods for detection of traces of explosives on surfaces. A smearing transfer method can be used for depositing the target explosives on the substrates to be used as standards and samples. The sample transfer method is appropriate to compare with other methods of sample preparation due the fact that a mass balance is not needed in order to know the amount of the sample on the surface. Besides that, many plates were prepared, good reproducible distributions were found (the analyte was distributed almost homogeneous on the surface). Samples ranging from micrograms/cm² to nanograms/cm² of 2,6-dinitrotuelene (DNT), 2,4,6-trinitrotoluene (TNT), pentaerythritol tetranitrate (PETN), nitroglycerine (NG) and triacetone triperoxide (TATP) were deposited as on stainless steel surface. Methanol, acetone and acetonitrile were used as transfer solvents. The IR reflectance spectra were recorded at 4 cm^{-1} resolution and 50 scans. The results of the experiments gave intense absorption bands in the fingerprint region of the IR spectra that were used to calculate the detection limit for each of the target explosives. The nitro band can be used for explosives detection since it acts as a vibrational signature of several classes of explosives: nitroaromatic, nitroaliphatic, nitramines and nitrate esters. Figs. 8 and 9 show the prominent signal of nitro explosives deposited on stainless steel surfaces. Only one signal in the range 1200 – 1400 cm^{-1} was significant for quantitative and qualitative analysis. This band can be attributed NO_2 stretching vibration. Nitro stretching vibration of PETN and NG appears in the 1250 – 1320 cm^{-1} region. For nitroaromatic explosives such as TNT and 2,6-DNT the band appears at 1320 – 1360 cm^{-1} [42, 43].

This difference can be explained in terms of the fact that the group NO_2 in PETN and NG is attached to an oxygen atom. However, in DNT and TNT, this nitro group is directly attached to the aromatic ring. The high electronegativity of the oxygen atom in PETN and NG attracts electron density from the nitro group leading to a lowering of the oscillator strength and causing a shift to lower frequencies. This effect is lower or not present in the aromatic ring for TNT and DNT.

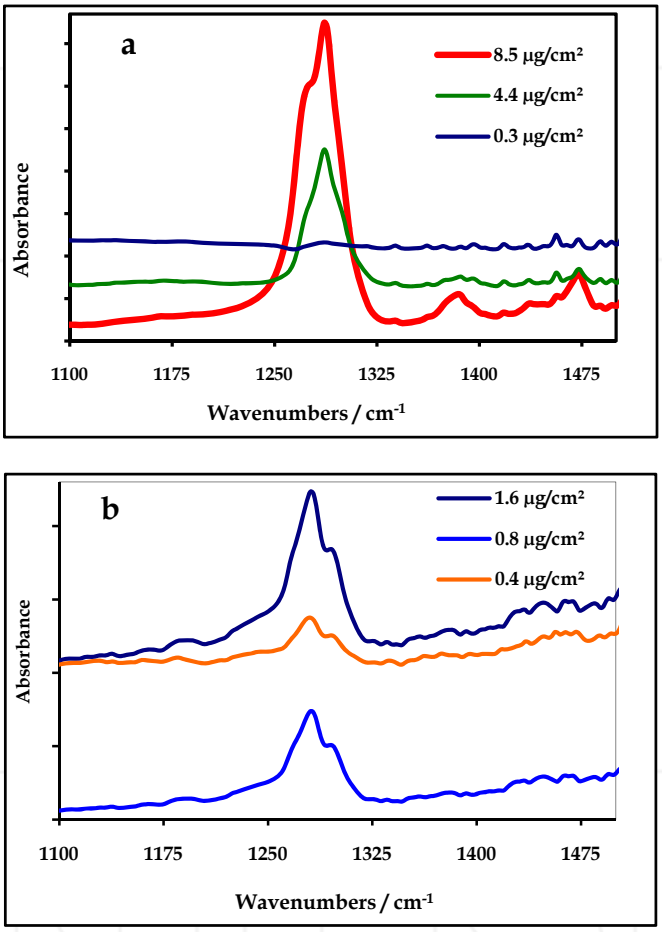


Fig. 8. Grazing angle spectra of nitroexplosives: a. PETN; b. NG

Fig. 10 includes the calculated classical detection limits for some nitroexplosives. For surface loadings near the detection limit only the NO_2 signal can be observed without aid of software. The detection limit varies according to macro properties. Properties such vapor pressure, physical adsorption, sublimation rate and surface-adsorbate thermodynamics can influence the detection limit. A close relation between vapor pressure and limit detection is shown for nitro explosives. Table 1 shows the values for the vapor pressures near room temperature of the explosives studied. The amount of explosive on surface of stainless steel

and the residence time depends on this property, because at this loading concentration the explosive goes to vapor phase fast. This phenomenon is remarkably observed in TATP with detection limit of 8 μg/cm². Some macro properties also can affect sublimation at room temperature and TATP detection limit.

Fig. 11 illustrates that several TATP bands can be used for its detection with accuracy. It shows the prominent presence of the most intense bands of TATP in the IR fingerprint region. The band at 1205 cm^{-1} belongs to the C-O stretch, 1365 cm^{-1} is a deformation of CH_3 group and the band at 1471 cm^{-1} is an asymmetric deformation of CH_3 group. The spectroscopic window used for TATP detection was spectral range of 1320-1407 cm^{-1}. Forward selection analysis of variable significance affirms that the significant peaks were contained within the spectral range of 1330-1407 cm^{-1} (Beeson, 1998; Demuth, 1998). The best discriminant model was done using peak areas. It was selected based on statistical significance and the percent of cases correctly classified. The statistical significance value (p-value) was $p < 0.0001$ and the percent of cases correctly classified was 90.6%.

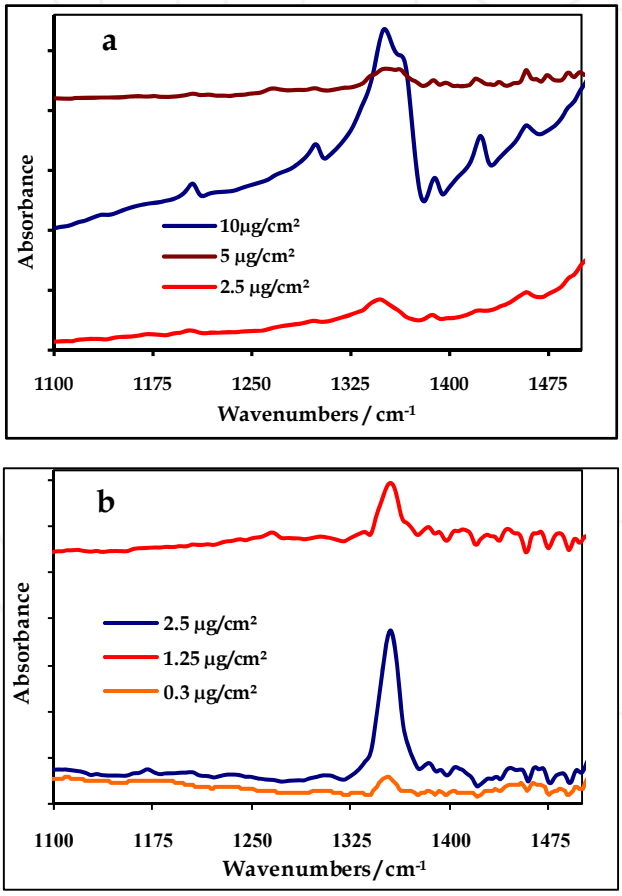


Fig. 9. Grazing angle spectra of nitroexplosives: a. 2,6-DNT; b. TNT

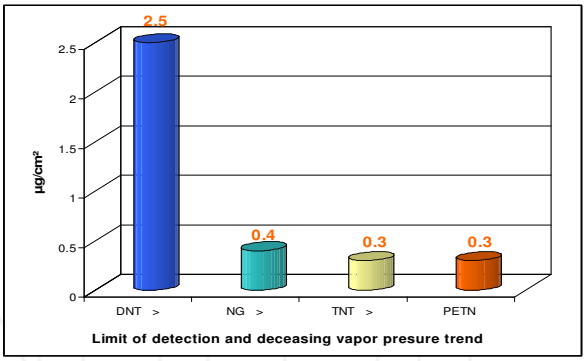


Fig. 10. Detection limits for selected nitroexplosives and correlation with decreasing vapor pressure of the energetic compound

Explosive	Vapor pressure(mm Hg)
2,6-dinitrotoluene	5.67×10^{-4} at 25 °C
2,4,6-trinitrotoluene	1.99×10^{-4} at 20 °C
nitroglycerin	2.0×10^{-4} at 25 °C
pentaerythritol tetranitrate	1.035×10^{-10} at 25°C
TATP	5.25×10^{-2} at 25 °C

Table 1. Vapor pressure of nitroexplosives and TATP

It is important to emphasize that measuring surface concentrations using the peak area method is conceptually simple and easy to use, but it has limitations. The method is univariate (the concentration is determined with a single spectral peak) and depends on a linear correlation between the concentration and the spectral response. The results can, therefore, be undermined by perturbations such as fluctuations caused by detector noise, temperature variations, or molecular interactions. Statistically based, multivariate calibrations use spectral features over a wide range. Information from a calibration spectral set (a training set) was compared to independently determined concentration data using partial least squares (PLS) regression. This method is based on the assumption that systematic variations in the spectra are a consequence of concentration changes. A calibration model for analysis of 2,6-DNT was built using the Quant 2 package, an add-on software package to the OPUS™ data acquisition and analysis software (Bruker Optics). The best spectral region was in the range of 1702 – 1269 cm^{-1}. This range was used for model generation. No spectral data preprocessing was applied to the spectra. The results are illustrated graphically in Fig. 12. The root mean square error of the cross validation (RMSCV) was 0.957, and % R^2 was 97.75. This calibration was used in order to predict unknown loading concentrations. For a deposited loading concentration 3.78 and 7.56 $\mu g/cm^2$, 3.58 and 7.35 $\mu g/cm^2$ were detected, respectively. Clearly, low levels of explosives can be detected and measured on a metal surface with good results. So, chemometrics easily leads to a powerful technique for surface contamination detection and measurement. Moreover, classical detection limits do not apply any longer. Thus the reported values are on the conservative side.

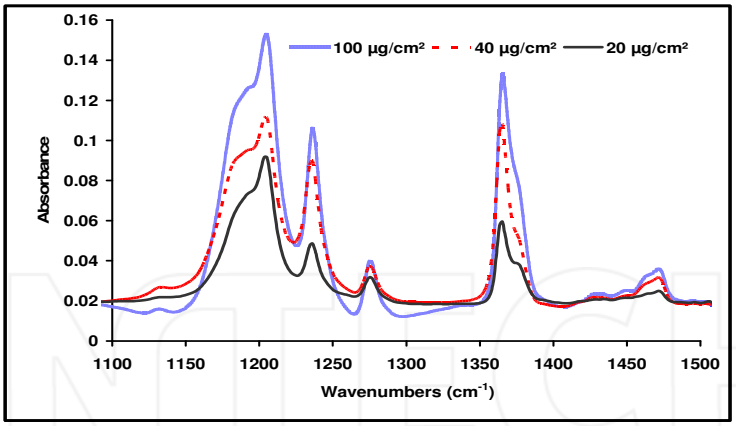


Fig. 11. GAP spectra of TATP at three surface loadings: 20, 40 and 100 µg/cm²

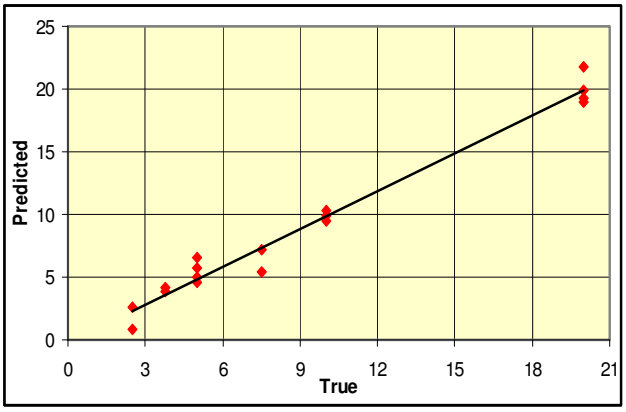


Fig. 12. Partial least squares regression cross validation plot for 2,6-DNT

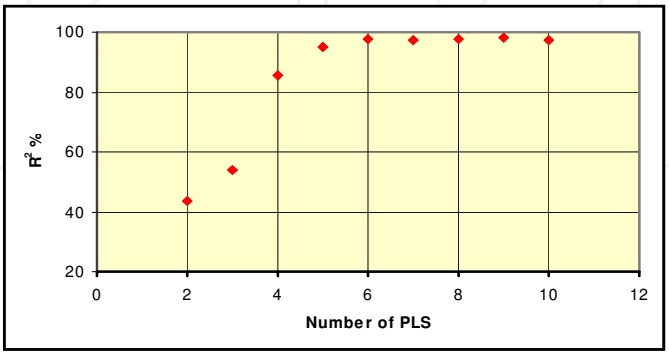
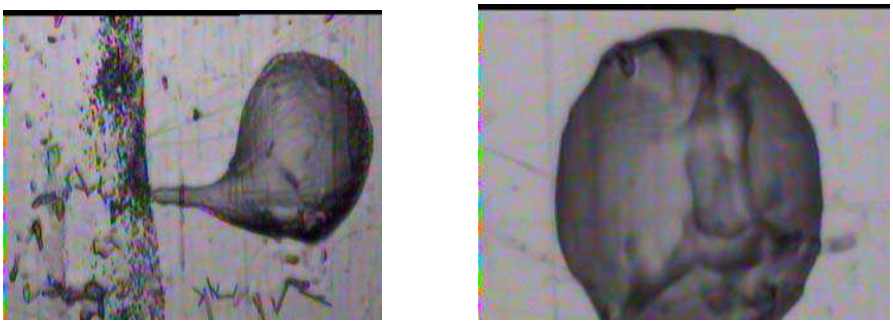


Fig. 13. Variation of % R² with the number of PLS carried out

The importance of applying PLS that it is used to design and build robust calibration models for quantitative analysis. PLS regression is a quantitative spectral decomposition technique that is closely related to principal component (PC) regression. It uses the concentration information during the decomposition process. This causes spectra containing higher constituent concentrations to be weighted more heavily than those with low concentrations. The main idea of PLS is to get as much concentration information as possible into the first few loading vectors or number of PLS (Kramer, 1998; Otto, 1999).

Fig. 13 shows how the calibration model improves with the addition of PLS. Five PLS executions were necessary to build a good calibration model. This indicates that the relationship that exists between the loading concentration and the spectral absorbance in this technique is multidimensional. The robustness of model calibration was evaluated using internal jackknifing validation [44]. Model with lower PLS than 6 was not capable to predict new data with good precision. Fig. 14 shows the appearance of TNT deposits under high magnification of a white light microscope. At these loading concentrations almost all the stainless steel surface was covered by crystals. Using this transfer method, positive and inverted bands were observed (Fig. 15). For loading concentrations > 8 μg/cm², only positive bands were observed. This fact can be attributed some changes in the refraction index of TNT.



a. 20 μg/cm² 50 x b. 40 μg/cm² 100 x

Fig. 14. Optical images for TNT deposited on stainless steel using smearing method

Substance	Sample Preparation Method	Solvent	Detection Limit (μg/cm²)
TNT	Smearing	dichloromethane	0.3
2,6-DNT	Smearing	acetone	0.3
PETN	Smearing	methanol	0.3
NG	Smearing	acetonitrile	0.4
TATP	Smearing	dichloromethane	8.0

Table 2. Detection limits of Explosives using FT-RAIRS on stainless steel

The calibration model was built using the Quant 2 package, an add-on software package to the OPUS™ (Bruker Optics) data acquisition and analysis software. In this study, the model parameters were optimized in the spectral region 1668-1045 cm^{-1}. No spectral data preprocessing was done. The resulting model was cross-validated using the ''leave one out''

method in which each spectrum is omitted in turn from the training set and then tested against the model built with the remaining spectra. The results are illustrated graphically in Fig. 16. The root mean square error of the cross validation was 0.918, and R^2 was 0.9858. TNT can be detected and quantified on metallic surfaces and low concentrations as 500 ng/cm^2. Table 2 shows a summary of the detection limits of the explosives using smearing transfer method. The detection limit depends on the vapor pressure.

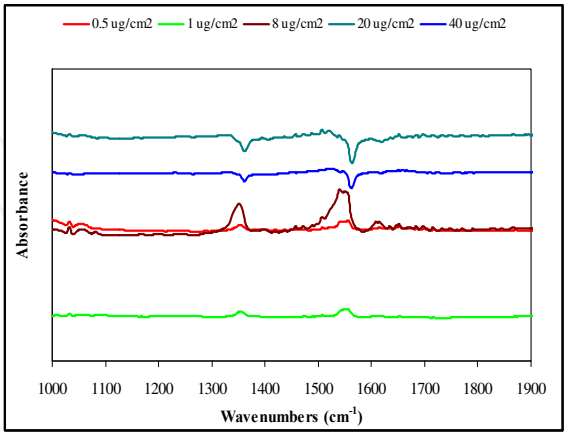


Fig. 15. FT absorption-reflection IR spectra of TNT on stainless steel using smearing transfer method

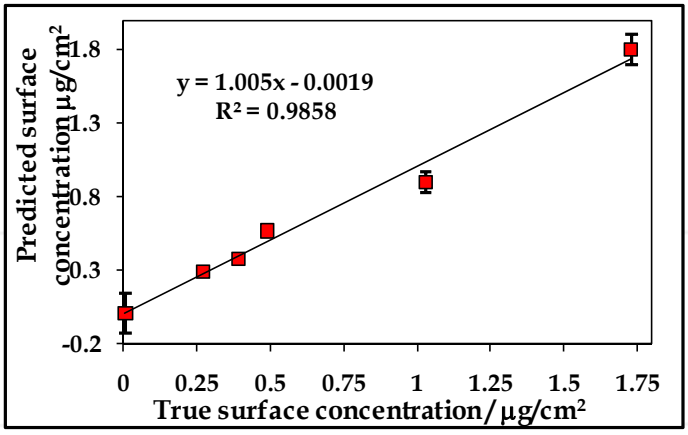


Fig. 16. Leave one out cross validation for TNT surface concentration deposited on stainless steel

6. Sublimation studies of TATP deposited on metal surfaces

The rapid sublimation of the TATP under normal atmospheric conditions presents a significant challenge in applying the fiber-optic grazing-angle method to surface detection

of the compound. Even under controlled laboratory conditions, where spectroscopy was carried out on freshly prepared samples, it proved impossible to develop PLS1 calibrations that met a high standard of robustness or usability. The rapid change in the surface concentration of TATP made it impractical to collect more than one spectrum from each sample and this further limited the possibilities of building a statistically useful data set. The samples were deposited on the surface using a smearing method. To carry out the experiments, TATP was synthesized in the laboratory. For the calibration curves TATP was dissolved in dichloromethane. A solution with an initial concentration of 0.23 g/mL was prepared and then dilutions were made until obtain 0.23 g/mL. The resulting average surface concentrations of TATP ranged from 8 to 200 μg/cm^2. Since dichloromethane evaporates very fast (Boiling point = 39.8 °C), a thin sample film was observed after smearing. Once the solvent had evaporated, the spectrum of the sample was collected immediately to minimize the impact on the calibration of rapid TATP sublimation. The data was analyzed using chemometrics routines; in particular multivariate PLS was used. In other experiments done with stainless steel plates coated with 25-100 μg/cm^2 TATP, spectra were recorded every 27 seconds at 20-30 °C and the sublimation behavior at the studied temperatures was observed.

The readiness with which TATP sublimates (vapor pressure at room temperature = 7 Pa) under normal atmospheric conditions complicates the task of calibrating the detection of the compound on surfaces, as the surface concentration of TATP decreases over the timescale of the experiment. To explore this effect, experiments were performed with stainless steel plates initially given a nominal loading of 100, 80, 50, 20, and 10 μg/cm^2 TATP. Immediately after deposition of the TATP, the probe head was positioned on the surface, and spectra were collected every 27 seconds. Fig. 17 shows successive spectra taken at 27 s intervals from an initial loading of 100 μg/cm^2 TATP.

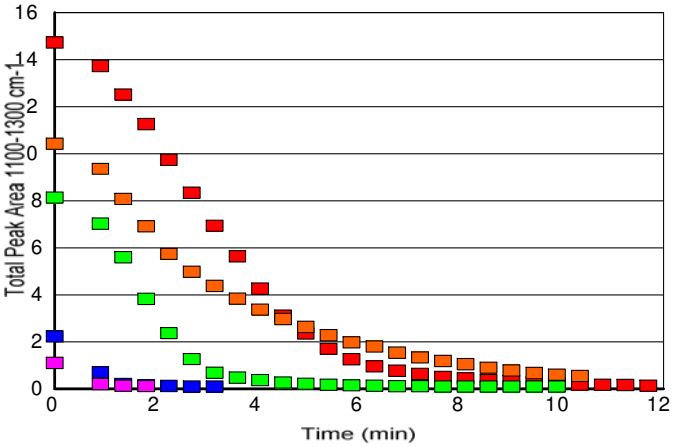


Fig. 17. Peak areas for the grazing-angle mid-IR spectra in the range from 100-1300 cm^{-1}, for initial loadings of 100, 80, 50, 20, and 10 μg/cm^2

A peak-fitting model of the spectral region from approximately 1300-1100 cm^{-1} was fitted against each of the spectra in turn to give the total peak area for the selected region. Fig. 18

shows the result for samples of stainless steel substrates initially loaded with 10-100 µg/cm^2, represented as a graph of peak areas versus time. The amount of TATP detected on the surface drops below the apparent detection limit of the technique within about 9 min.

Standard were prepared as described in the Experimental Section. Grazing angle FT-IR spectra of freshly prepared samples were collected for a series of different surface concentrations, as shown in Table 1. When a PLS1 model was built from all of the 79 spectra listed, using the spectral region from 1066-1506 cm^{-1} and no spectral preprocessing, it proved impossible to develop a model that met a reasonable standard (the maximum obtainable value for R^2 was about 0.75). In this study, the model parameters were optimized in the spectral region 1498 – 1113 cm^{-1}. No spectral data preprocessing was done. When data used for the model was limited to loadings below 40 µg/cm^2, it was possible to build a calibration with R^2 = 0.869, and root mean square error of cross validation (RMSECV) = 3.69 (obtained from a leave-one-out cross validation); the results are shown graphically in Fig. 18. The graph shows the degree of scattering in the data. While some of this scattering may be attributable to variations in the amount of TATP deposited on each coupon, sublimation of the TATP during the experiment is another likely contributing factor. Given these limitations, the quality of the calibration that has been developed is surprisingly good and it is clearly quite possible to detect microgram quantities of TATP on metal surfaces using grazing angle FTIR methods. This is in agreement with previous results for a range of organic compounds on metal and glass surfaces (White, 1992; Lin-Vien, 1991).

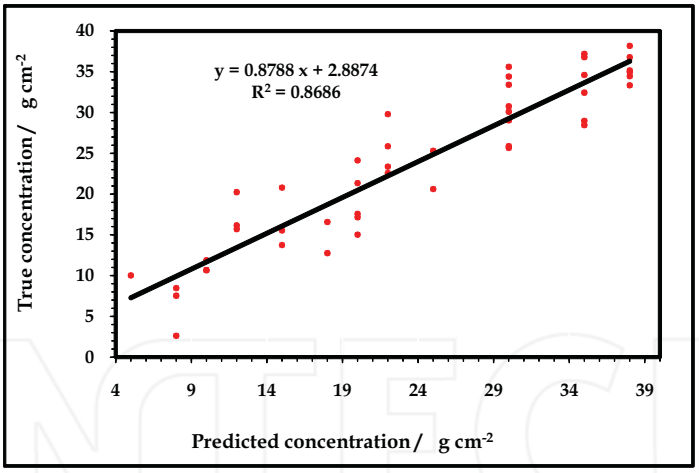


Fig. 18. "Leave-one-out cross" validation of predicted values vs. true values for TATP on stainless steel substrates. All data are shown illustrating point scattering at each value measured

TATP quantification was done using the calibration generated by chemometrics. These results are shown in Table 3. Clearly, low levels of TATP can be detected and measured on a metal surface with quantitative results (Table 4). Attempts to build a separate calibration for the loadings of 40 µg/cm^2 and above were unsuccessful may be attributable to the change in the nature of the surface coating at high loadings, from a thin film capable of generating a double-pass transmission spectrum to a bulk material generating diffuse surface reflectance

independent of the thickness (Blaudez, 1998; Bradshaw, 1988; Hayden, 1987). Calibration curve is shown in Fig. 19.

A true "transflectance" experiment is possible only when the layer of organic material is "very" thin, so that the IR radiation can pass all the way through it to be reflected from the substrate (e.g. the metal). If the coating is thicker than about 1 or 2 micrometers, the IR radiation does not reach the substrate and is reflected from the top surface of the organic material (as if it were a "bulk" sample of the organic material) to give diffuse reflectance (Kaihara, 2001). This changes the spectrum that is obtained from the organic material from the "transflectance" spectrum to a diffuse reflectance spectrum. The diffuse reflectance spectrum is not affected by the thickness of the coating, since it does not come from the whole coating but only from the top layer, and so it does not contribute to a useful calibration.

TATP Loading (μg/cm²)	Number of Spectra	Included in Calibration
5	1	No
8	3	Yes
10	4	Yes
12	4	Yes
15	4	Yes
18	3	Yes
20	7	Yes
22	5	Yes
25	3	Yes
30	9	Yes
35	7	Yes
38	7	Yes
40	8	No
50	2	No
60	2	No
80	2	No
90	2	No
150	1	No
200	1	No

Table 3. List of TATP spectra collected for calibration

Deposited (μg/cm²)	Detected by grazing angle (μg/cm²)	Difference
14.96	13.8	1.16
17.96	16.5	1.46
24.94	25.3	-0.36
34.91	32.6	2.36

Table 4. Quantification of TATP on metal plates by grazing angle Fiber optic FT-IR

Discriminant analysis was also performed to classify the TATP loading concentration in two groups (Fig. 20). The first one corresponds to concentrations lower than 25 μg/cm² and the second one to concentrations higher than 25 μg/cm². Peak areas of signals in the range of 1330-1407 cm^{-1} and 1407-1503 cm^{-1} were used for the discrimination. Forward selection analysis of variable significance affirms that the significant peaks were contained within the spectral range of 1330-1407 cm^{-1}. The best discriminant model was selected based on statistical significance and the percentage of cases correctly classified. The percentage of cases correctly classified was 90.6% and the significance statistical p-value was < 0.0001.

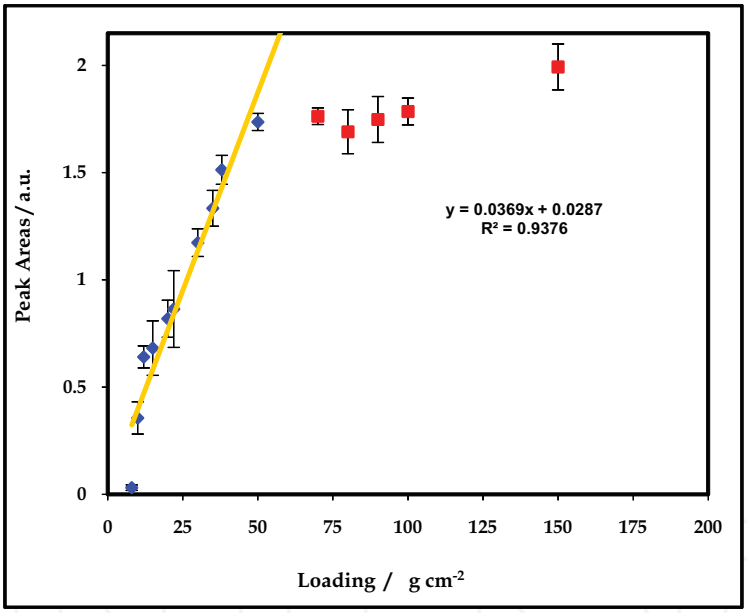


Fig. 19. Calibration curve for TATP on SS surface using peak areas (1330-1407 cm^{-1})

TATP sublimation behavior depends of the temperature as is expected. This is shown in Fig. 21. For temperatures lower than 20 °C TATP sublimation rate is significantly low. Beyond this point the sublimation rate starts to increase very fast. The rate of sublimation for TATP was calculated in units of peak-area/second for the slope of the curve, taking into account the linear range of the graphs (Fig. 21-a and -b). Since the surface concentration is well approximated by the measured peak area in this range because these are proportional (Fig. , the mass transferred to the vapor phase is reasonably well estimated by the decrease in peak

areas in the MIR. This range was selected because here the rate decays more rapidly. These results can be confirmed in Fig. 19. The estimated value of the rate sublimation with temperature in the range of 20-30 °C is -0.0013 peak area-s^{-1} $°C^{-1}$. This is equivalent to ~ -0.81 µg cm^{-2} s^{-1} $°C^{-1}$ in the range of 20-30 °C. These experiments confirm the fact that TATP sublimates very fast in the absence of vapor pressure reducing and stabilizing agents.

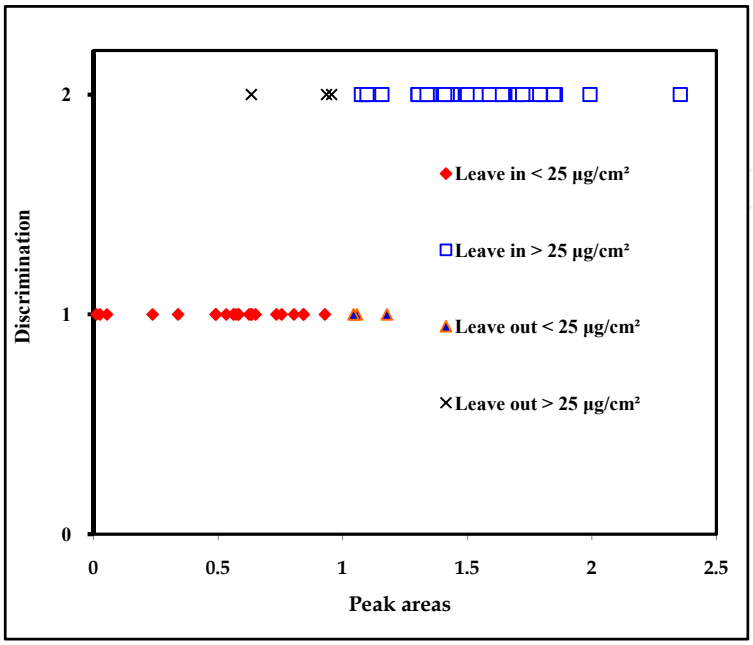


Fig. 20. Plot of peak area selected to construct the discrimination for grouping 67 concentrations

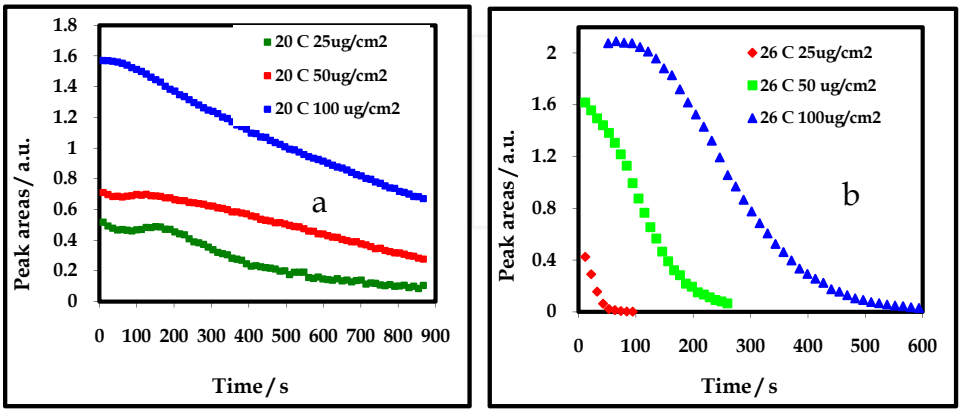


Fig. 21. Plots of peak areas versus time (second) for 25, 50 and 100 µg/cm² at: a: 20 °C; b: 26 °C

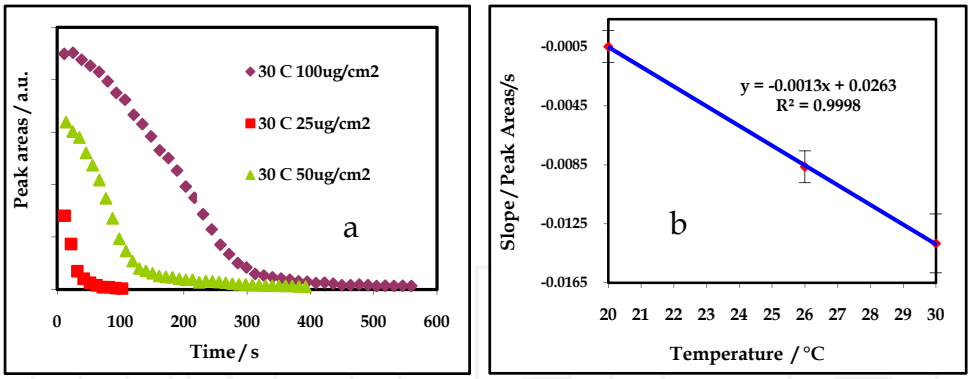


Fig. 22. a: Plot of peak areas versus time (s) for 25, 50 and 100 µg/cm² at 30 °C; b: sublimation rate for different temperatures

7. Characterization studies of TNT deposited on metal surfaces by TIJ technology

Although smearing transfer method gave almost a homogenous distribution of the explosives on surfaces, the method of mass transfer required the preparation of many samples. This was due to the fact that the sample transfer method is prone to uncontrolled operator errors. To circumvent the problem, Thermal inkjet technology (TIJ) was used for the experiments. TNT was selected as an explosive for dispensing on TIJ.

Thermal Inkjet Technology is able to deposit very small amounts of chemical compounds, including energetic materials, in a specific location on a surface. Aliquots of TNT solutions were deposited on stainless steel film. A coating of TNT can be produced by controlling the concentration of TNT, the number of drops dispensed and the distribution of drops over the surface. The loading concentration of the sample on the surface can be controlled by varying parameters such as: number of passes, dispensing frequency, applied energy and pen architecture. Precise delivery of known number of droplets with known mass and concentration are known. Also only one solution can be used, avoiding dilutions that can increase the analytical errors.

The precise amount of TNT on the surface can be known in different ways: one can be known if the drop weight of the solutions is known and the other using an alternative method for quantification (GC, UV or HPLC). In our study, the stainless steel sheets loaded with TNT samples were rinsed with 25-100 mL of acetonitrile to remove the entire TNT that was delivered on the surface. Then solutions were transferred a volumetric flask and filled to mark with solvent. Gas chromatography was used in our experiments as an alternative method in order to determine the surface loading concentration. The analysis was carried out using an Agilent Technologies 6890N, Network GC System with A micro cell ^{63}Ni Electron Capture Detector (µECD). For GC separation, a capillary column was used: RTX-5 (cross bonded 5% diphenil-95% diethyl polysiloxane) 15 m x 0.25 mm ID x 0.25 µm df, Restek Corp, Bellefonte, PA. The GC oven was held at 80 °C for 1 min and then programmed at 10 °C/min to 180 °C, followed by a 30 °C/min ramp to 300 °C. The temperature at the injection port was 250 °C.

Calibration curves for GC-μECD were prepared with 1000 ppm standard solutions of TNT obtained from Restek Corp. Stock solutions of concentrations: 1, 0.5, 0.1, 0.05, 0.01 ppm diluted in HPLC grade acetonitrile (Sigma-Aldrich Chemical Co., Milwaukee) were prepared. All solutions were injected directly using glass syringes (Hamilton Series 7101) into the injection port, this analysis were carried out with three replicates for each concentration. Using TIJ method it was possible to generate standards samples with more uniform coverage, and one advantage is the fact that the surface loading concentration could be varied by changing the numbers of passes delivered to the sample without the need for serial dilutions. Fig. 23 shows the linear correlation of the variation of the number of passes for a 5292 ppm solution. A direct relationship is observed.

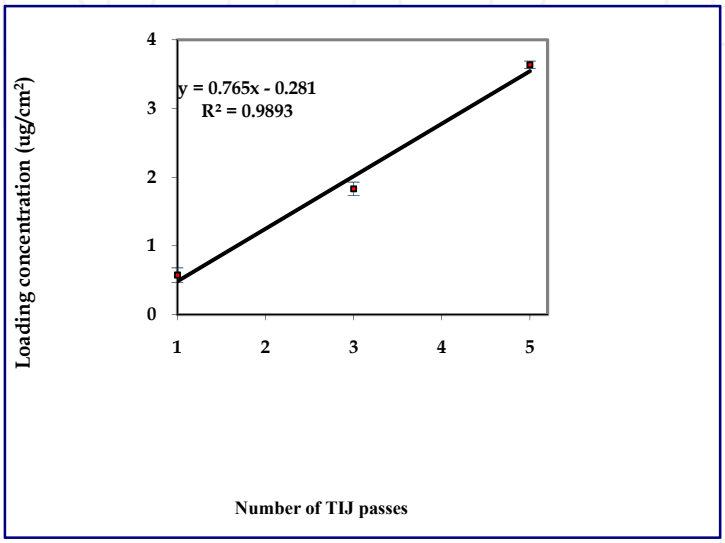
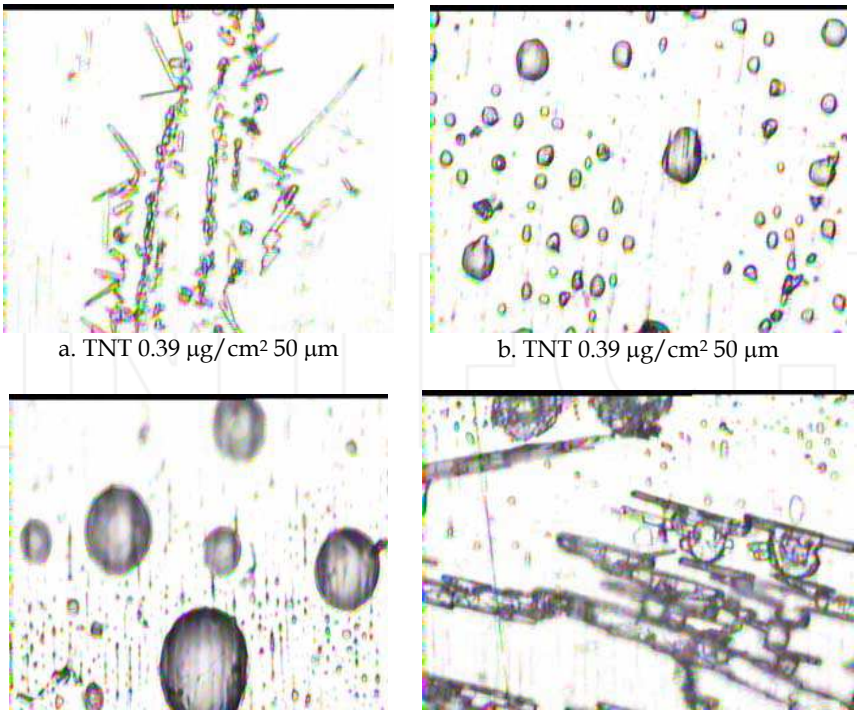


Fig. 23. Variation of the number of passes delivered to the surface for a TNT 5292 ppm solution

For visualization of substrates after sample deposition, optical microscopy images were captured using Leica microscope (model LS). These images can be seen in Fig. 24. The images clearly reveal that TNT is collecting on the stainless steel surface as droplets and crystals. As the number of passes increases on the surface, the size of the droplets increases. At 1.25 μg/cm² loading concentration, the surface is practically covered by crystals.

The way how the sample is distributed on the surface plays an important role in the task of calibrating the detection of the TNT or other compound on surfaces. To explore this effect, experiments were performed by positioning the probe head on different parts on the surface, and spectra were collected. Fig. 25 shows how the analytical response varies with the loading concentration. A peak-fitting model of the spectral region from approximately 1583-1396 and 1418-1220 cm^{-1} was fitted against each of the spectra in turn to give the total peak area for the selected region. This confirmed that although the coating is not homogenous a pattern with TIJ is generated, giving a few errors.



a. TNT 0.39 μg/cm² 50 μm b. TNT 0.39 μg/cm² 50 μm

c. TNT 1.25 μg/cm² 50 μm d. TNT 1.25 μg/cm² 50 μm

Fig. 24. White light images of TNT deposits on stainless steel substrates using TIJ for deposition

The asymmetric and asymmetric vibrational stretches of the nitro (NO_2) group can be used for explosives detection since they act as vibrational signatures of several classes of explosives: nitroaromatic (TNT, DTNT), nitroaliphatic (CH_3NO_2), nitramines (RDX, HMX) and nitrate esters (nitroglycerine, PETN). Fig. 26 shows the prominent signals of TNT deposited on stainless steel surfaces. All nitro signals were significant for quantitative and qualitative analysis. Nitro symmetric stretching vibration of TNT band appears at 1320 – 1360 cm^{-1} and nitro asymmetric stretching vibration is typically located in the wavenumber range: 1477 – 1600 cm^{-1}. A completely inverted spectrum with all bands pointing downwards is finally obtained for 1.73 μg/cm² and 3.58 μg/cm² loading concentration (Fig. 26). Fig. 27 shows TNT IRRAS spectra of two types of deposits. For very low surface loadings and for high surface concentrations (3.58 μg/cm² shown in trace of Fig. 27-A), the typical downward looking percent transmission profiles are observed. However, IRRAS spectroscopic features of thins films on surfaces frequently show strongly variations of the relative intensities and wavelength shifts as the surface loadings changes. These effects were observed when the loading concentration of TNT was bellow 1.73 μg/cm² (Fig. 27-B). Significant wavenumber shifts (> 10 cm^{-1}) were found for the nitro vibrations bands. An inverted upward looking profile similar to an absorption spectrum of the bands is clearly seen in lower trace in Fig.

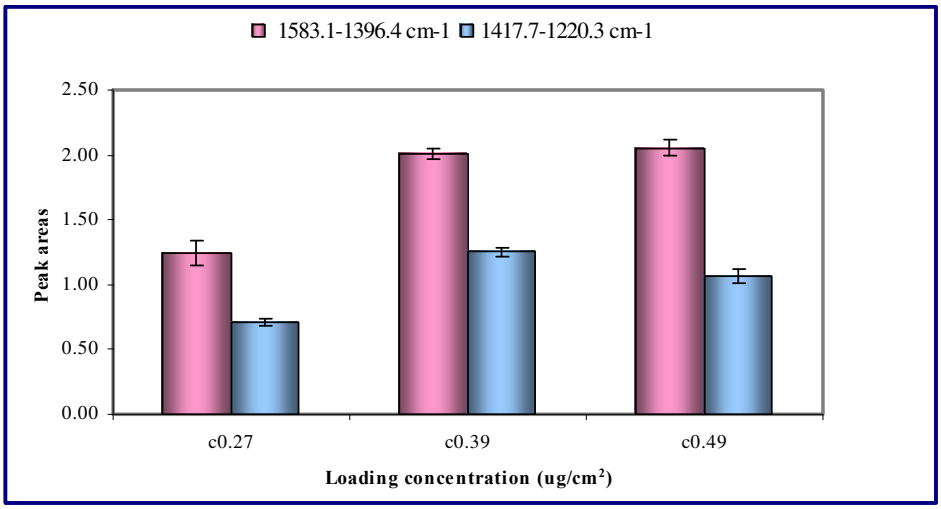


Fig. 25. Variation of the distribution of the TNT deposited on stainless steel using TIJ with the analytical response

The reflected radiation depends of the nature of the film as well as the incident angle of the IR beam reaching the surface. The properties of the reflected IR light depend critically on localized optical characteristics of the film such as the local index of refraction. Sample heterogeneities give rise to local changes in the refractive index across the sample and affect the reflectance spectra. Typical Infrared Reflection Absorption Spectroscopy (IRRAS) spectra are reported in terms the measured absorbance defined as (Bradshaw, 1988; Hayden, 1987):

$$A = \text{Log}\,(Rr/Rs) \tag{1}$$

where: Rr is the reflectance of the of the plate (substrate) and Rs is the reflectance of the sample-substrate [22]. Fig. 26-A shows "normal" (vertical up) for low surface loadings (0.27- 0.47 μg/cm²). Fig. 26-B shows corresponding RAIS spectra for 1.73 and 3.58 μg/cm² illustrating downward pointing peaks. Figs. 27-A and B contains expanded views in the wavenumber range for Rr and Rs of 0.27 μg/cm² and 3.58 μg/cm². The metallic substrate reflects more than the sample (TNT), it means that Rr > Rs and a normal peak is found. This can be observed in Fig. 27-B. The inverted peak is found in 1561 because the difference between Rr and Rs in this frequency value is lower than in other values around this one in the frequency range 1548-1565 cm^{-1}. This means that the sample is reflecting almost at the same level as the metal substrate. This fact can be explained based on changes in the refractive at the film-metallic surface interphase or by the way that TNT layers are packed on the surface. Another reason can be argued is the fact that these inverted bands are found when most of the coverage of the surface is TNT crystals and not amorphous TNT droplets.

It is very important to emphasize that measuring surface concentrations using the peak area method is conceptually simple and easy to use, but it has some limitations. The method is univariate: the concentration is determined with a single spectral peak and it

depends on a linear correlation between the concentration and the spectral response. The results can, therefore, be undermined by perturbations such as fluctuations caused by detector noise, temperature variations, or molecular interactions. Statistically based, multivariate calibrations use spectral features over a wider range. Information from a calibration spectral set (a training set) was compared to independently determined concentration data using partial least squares regression (PLS). The method is based on the assumption that systematic variations in the spectra are a consequence of concentration changes.

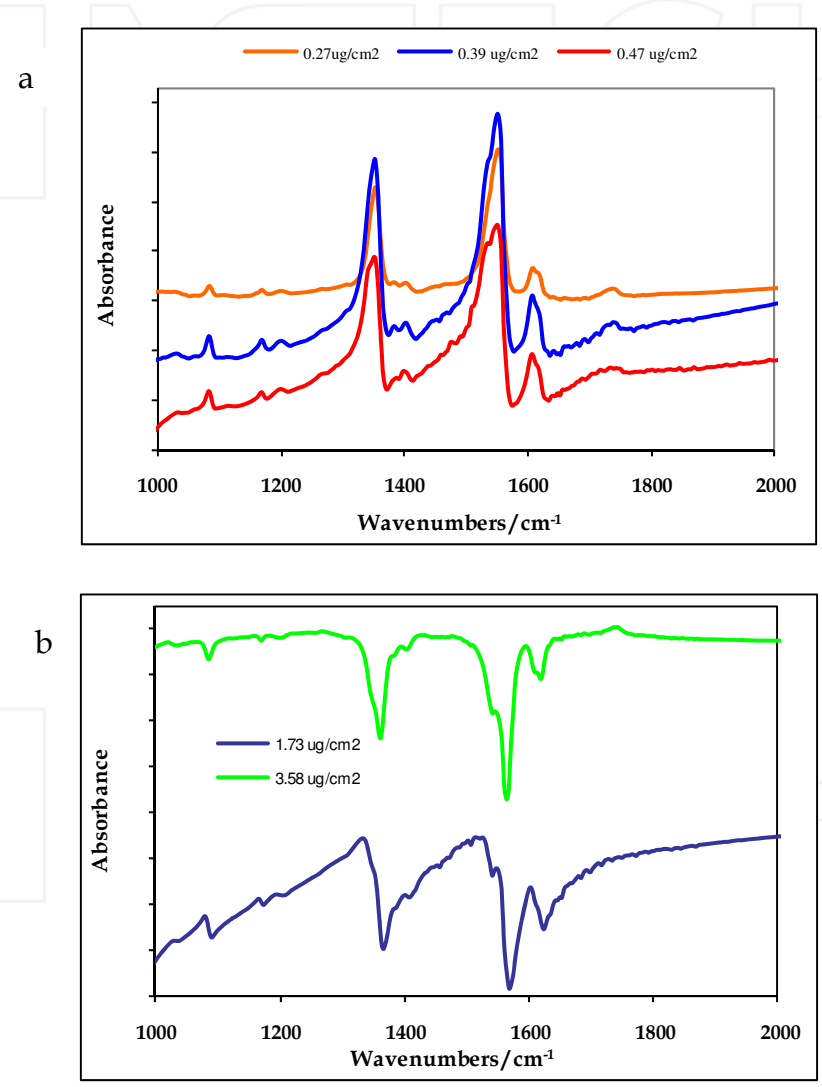


Fig. 26. Grazing angle spectra of TNT: a. Positive bands (normal). b. Inverted bands

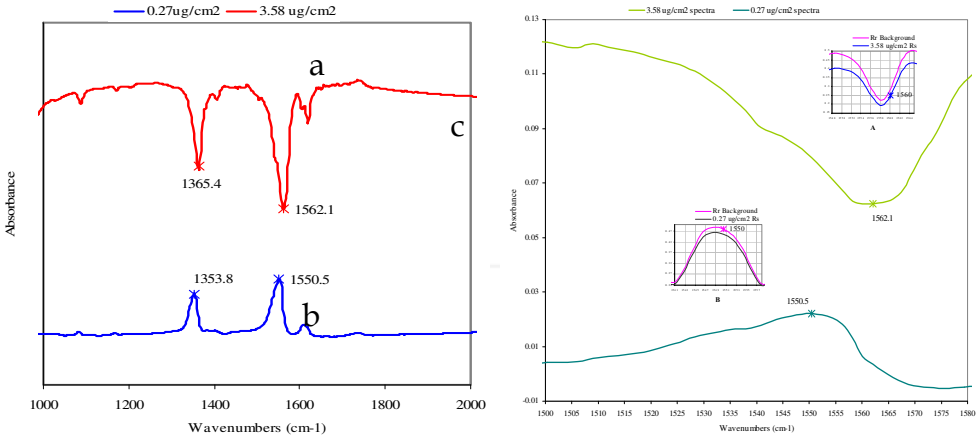


Fig. 27. IRRAS symmetric and asymmetric Nitro stretching vibrations of TNT a: normal (downward pointing) %T profile; b: inverted (upward pointing) %T profile; c: behavior of Rs and Rr for the peak formation

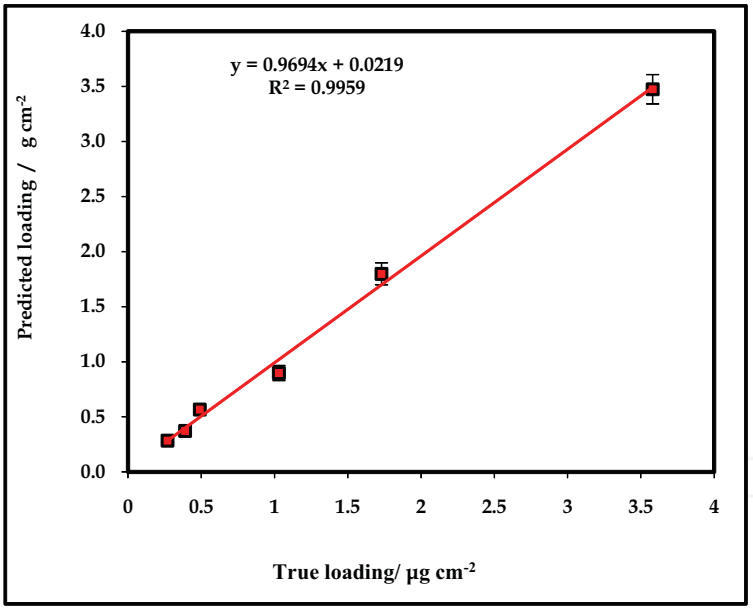


Fig. 28. Graph of averaged predicted surface loading values versus true values obtained by the "leave-one-out" cross validation of a Quant2 calibration for TNT on steel surfaces

Grazing angle FTIR spectra of prepared samples were collected for a series of different surface concentrations. When a PLS1 model was built from all of the 84 spectra listed, using the spectral region from 1028-1713 cm^{-1} and no spectral preprocessing, it was possible to build a calibration with R^2 = 0.9959, and root mean square error of cross validation

(RMSECV) = 0.258 (obtained from a "leave-one-out" cross validation); the results are shown graphically in Fig. 28. According to these results TIJ-IRRAS combined with PLS can be used for TNT detection and quantification on metallic surfaces.

8. Conclusion

Fiber optic coupled FTIR spectroscopy has been shown to provide a useful technique for methods development for *in situ* detection explosives and active pharmaceutical ingredients and excipients on metallic surfaces. Very low detection limits (~ 0.3 μg/cm²) were found for nitroexplosives and for organic peroxides. The present study was limited to highly reflective metallic surfaces such as stainless steel. However, since it is based on a grazing angle probe that is sensitive to adsorbates deposited on substrates regardless of the nature of the surface, and at least to first order, the methodology should be applicable for trace detection of explosives on other types of surfaces. The coupling of data acquisition to powerful non linear, multivariate analysis based on chemometrics routines such as partial least squares resulted in an even more robust analytical methodology. Explosives detection for this technique as applied to the target molecules used was found to be limited by the residence time of the substance on the surface. At low loading concentrations, high vapor pressure explosives escape to the vapor phase by sublimation even faster, limiting the low limit of detection achievable. The detection limit for TATP, high vapor pressure organic peroxide, was 8 μg/cm². For nitroexplosives such as 2,6-DNT, NG TNT, and PETN LODs found were 2.5, 0.4, 0.3, and 0.3 μg/cm² respectively. The results are in good agreement with a decreasing low limit of detection when arranged in the order of decreasing vapor pressure.

In this work we present the first method for detection and quantification of TATP on metallic surfaces. The method entails the coupling of optical fibers operating in the mid infrared (MIR) to a high performance interferometer. A grazing angle head coupled to liquid nitrogen cooled, MCT detector was used to excite and collect IR absorption spectra from stainless steel surfaces loaded with the organic peroxide. A smearing technique was used to prepare standards for surface loading of the explosive.

The novel aspects of the methodology involve *in situ*, remote sensed, fiber optic coupling of a grazing angle probe to a high sensitivity FTIR interferometer. It must be noted that the sample must to be measured immediately after the smearing to avoid sublimation of TATP from the metal surface. The robustness of the new methodology relies on utilizing powerful Chemometrics routines and Discriminant Analysis for statistical enhancement of data. Discriminant Analysis can be used to quantify TATP by classifying loading concentration in two groups: the first one corresponds to concentrations lower than 25 μg/cm² and the second one to concentrations higher than 25 μg/cm².

This work describes the first demonstration of a method for direct, *in-situ* detection and quantification of TATP on metallic surfaces. The rapid sublimation of TATP from the test surface has been demonstrated by tracking the decreasing intensities of the IR spectrum of TATP; the signal falls below the detection limit after about 9 minutes, from an initial value of 100 μg/cm². This phenomenon limits the extent to which the surface concentration of TATP can be calibrated. However, it has been possible to show that TATP on metallic surfaces can be quantitatively detected in a matter of seconds using the grazing-angle FTIR approach.

The first method for detection and quantification of TNT on metallic surfaces using thermal inkjet technology as a transfer method was discussed. Spectroscopic characterization and thin layer deposits quantification was achieved using the powerful Grazing Angle Probe Fiber Coupled-FTIR developed for surface analysis. Chemometrics routines were applied as data enhancers in order to accomplish fully the difficult task proposed. Inkjet printing of explosives demonstrated to have the following important characteristics: precision in sample deposit, drop delivery with non-contact fluid transfer and high reproducibility. The methodology could be applied for development of standards in Trace Explosive Reference materials on surfaces.

Sample transfer methods coupled to RAIRS promises to be an excellent support means when used as sensor for explosives detection on surfaces. A grazing angle head coupled to liquid nitrogen cooled, MCT detector was used to excite and collect IR absorption spectra from stainless steel surfaces loaded with the explosive. This case study suggests that MIR reflectance spectroscopy using a fiber-optic probe with grazing angle head is a viable method for detecting and measuring low (μg/cm2) quantities of organic contaminants on metallic surfaces. Adding chemometrics algorithms, which are developed and automated easily, leads to powerful techniques for surface contamination detection.

9. Acknowledgments

Parts of the work presented in this contribution were supported by the U.S. Department of Defense, University Research Initiative Multidisciplinary University Research Initiative (URI)-MURI Program, under grant number **DAAD19-02-1-0257**. The authors also acknowledge contributions from Mr. Aaron LaPointe from Night Vision and Electronic Sensors Directorate, Fort Belvoir, VA, Department of Defense, Dr. Jennifer Becker MURI Program Manager, Army Research Office, DoD and Dr. Stephen J. Lee Chief Scientist, Science and Technology, Office of the Director, Army Research Office/Army Research Laboratory, DoD.

Support from the U.S. Department of Homeland Security under Award Number **2008-ST-061-ED0001** is also acknowledged. However, the views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

Thanks are due to Dr. Luis. F De La Torre-Quintana for his assistance with statistical analysis and coupling with chemometrics routines and for his support and collaboration in design and preparation some figures.

10. References

Akhavan, J. (2004). *The Chemistry of Explosives*. 2nd ed. T J International Ltd, Padstow, Cornwall, UK,

Bacci, M.; Fabbri, M.; Picollo, M.; Porcinai, S. (2001). Non invasive fibre optic Fourier Transform-infrared reflectance spectroscopy on painted layers identification of materials by means of principal component analysis and Mahalanobis distance. *Anal. Chim. Acta.* 446, 15-21.

Beebe, K.R.; Pell, R.J.; Seasholtz, M.B. (1998). *Chemometrics: A Practical Guide*; John Wiley & Sons, Inc. New York, NY.

Beeson, R. and Skip, R.D., (1998). *Thermal ink jet technology –review and outlook*, Rung Research Laboratory, Inkjet Business Unit, Hewlett-Packard Company, Corvallis Oregon, US 3rd International Conference on Imaging Science. Chongqing, China.

Bellamy, A.J. (1999). Triacetone triperoxide: its chemical destruction. *J. Forensic Sci*. 44, 603-608.

Blaudez, D.; Buffeteau, T.; Desbat, B.; Fournier, P.; Ritcey, A.; Pezolet, M. (1998). Infrared refection – absorption spectroscopy of thin organics films on nonmetallic substrates: optimal angle of incidence. *J. Phys. Chem. B*. 102, 99-105.

Bradshaw, A.M.; Schweitzer, E.; R.J.H. Clark.; Hester, R.E. (1988). *Infrared reflection absorption spectroscopy of adsorbed molecules, in spectroscopy of surfaces*. *John Wiley & Sons*, New York, NY.

Brambilla, G., Loizzo, A., Fontana, L., Strozzi, M., Guarino, A. and Soprano, V. (1997). *J Am Med Assoc*., 278, 635.

Demuth, H.; Beale, M. (1998). *Neural Network Toolbox for use with MATLAB User's Guide Version* 3.0; 5th ed.; The MathWorks, Inc., Natick, MA.

Diem, M. (1993). *Introduction to Modern Vibrational Spectroscopy*, John Wiley, New York, NY.

Evans, H.K.; Tulleners, F.A.J.; Sanchez, B.L.; Rasmussen, C.A. An Unusual Explosive, Triacetonetriperoxide (TATP). *J. Forensic Sci*. 1986, 31, 1119-1125 .

Fierro-Mercado, P.M., Primera-Pedrozo, O.M., Hornedo, A., Hernández-Rivera, S.P., An In situ FTIR Fiber Optic Method for the Detection of Active Pharmaceutical Ingredients and Excipients on Metallic Substrates, in "Fourier Transform Infrared Spectroscopy: Developments, Techniques and Applications", Rees, O.J., ed., Chemical Engineering Methods and Technology Series, Nova Science Publishers, Inc. Hauppauge, NY, third quarter 2010, ISBN: 978-1-61668-835-6.

Griffiths, P.R. and De Haseth, J.A. (1986). *Fourier- Transform Infrared Spectrometry*. John Wiley, New York, NY.

Hamilton, M.L.; Perston, B.B.; Harland, P.W.; Williamson, B.E.; Thomson, M.A.; Melling, P.J. (2005). *Org. Process Res. Devel. 9*, 337−343.

Helmut, G.; Hans-Ulrich, G. (2002). *IR Spectroscopy: An introduction*. Wiley-VCH Federal Republic of Germany.

Hollins, P. (1992). The influence of surface defects on the infrared spectra of adsorbed species. *Surf. Sci. Rep.* 16, 51-94

Huberty, C. J. (1994). *Applied Discriminant Analysis*. Wiley-Interscience, NJ.

Hayden, B.E.; Yates, J.T.; Madey, T.E. (1987). *Reflection Absorption infrared spectroscopy, in Vibrational Spectroscopy of molecules on surfaces*. Vol. I. John Wiley & Sons, New York, N.Y.

http://chesapeake.towson.edu/data/all_electro.asp

http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB.htm

http://www.wcer.wisc.edu/step/ep301/Spr2000/esselman/IBUb.html

Kaihara, M.; Sato, Y.; Sato, T.; Takahashi, T.; Takahashi, N. (2001). Chemometrics for FT-IR reflectance spectroscopy. The distinction between the specular and diffuse reflection. *Analytical Sciences.* 17, 1701-1704.

Katona, Z.; Vincze, L.; Vegh, Z.; Trompler, A.; Ferenczi-Fodor, K. (2000). Cleaning validation procedure eased by using over pressured layer chromatographic. *J. Pharmaceutical and Biomedical Analysis.* 22*, 349-353.*

Kevin, J, Kochlowski.; Arbuckle-Keil, G. (2000). Infrared spectroscopy. *Anal Chem.* 72, 71-79R.

Kramer, R., (1998). *Chemometric Techniques for Quantitative Analysis*, Marcel Dekker, New York, NY.

Kurth, D. (1998). Analysis of artifacts in infrared spectroscopy of thin organic films on metallic substrates. *Langmuir*. 14, 6987-6991.

Lavine, B., and Workman, J. Jr. (2002). Chemometrics. *Anal Chem.* 74, 2763-2770.

Lin-Vien, D.; Colthup, N. B.; Fateley, W. G. and Grasselli, J. G. (1991).*The Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules*; Academic Press, San Diego, CA.

Mehta, N.K., Goenaga-Polo, J.E., Hernández-Rivera, S.P., and Hernández, D., Thomson, M.A. and Melling, P.J. (2002). Development of an *In-Situ* Spectroscopic Method for Cleaning Validation Using Mid-IR Fiber Optics. *Bio Pharm*. 15, 36-42.

Mehta, N.; Goenaga, J.; Hernandez-Rivera, S.P.; Hernandez, D.; Thomson, M.; Melling, P. (2002). Development of an in situ spectroscopic method for cleaning validation using mid-IR fiber optics. *Spectroscopy*. 18 (4),14-19.

Melling, P.J. and Shelley, P., (2001). Spectroscopic Accessory for Examining Films and Coatings on Solid Surfaces. U.S Patent 6,3,10,348, US Patent and Trademark Office, Washington, DC.

Melling, P, and Thomson, M. (2002). *Fiber optic probes for mid-infrared spectrometry*. John Wiley & Sons Ltd., UK.

Mirza, T.; Lunn, M.; Keeley, F.; George, R.; Bodenmiller, J. (1999).Cleaning level acceptance criteria and high pressure liquid chromatography procedure for the assay of Meclizine Hydrochloride residue in swabs collected from pharmaceutical manufacturing equipment surfaces. *Journal of Pharmaceutical and Biomedical Analysis*. 19, 747-756.

Mizaikoff, B. (2002). Sensory Systems Based on Mid-infrared Transparent Fibers. In *Handbook of Vibrational Spectroscopy,* Vol. 2, Eds. Chalmers, J.M. and Griffiths, P.R. John Wiley & Sons, Chichester, UK.

Nozal, M.; Bernal, J.; Toribio, L.; Martin, M.; Diez, F. (2001). Validation of a liquid chromatographic method for the determination of ranitidine hydrocloridine residues on surfaces in the manufacture of pharmaceuticals. *J. Chromatogr. A*. 919, 87-93.

Otto, M. (1999). *Chemometrics. Statistics and Computer Application in Analytical Chemistry*. Wiley-VCH. Federal Republic of Germany.

Pacheco-Londono, L., Santiago, A., Pujols, J., Primera-Pedrozo, O.M., Mattei, A., Ortiz, W., Ruiz, O., Ramirez, M., Hernandez-Rivera, S.P., (2007). Characterization of layers of Tetryl, TNB and HMX on metal surfaces using fiber optics coupled grazing angle-FTIR. *Proc. SPIE Int. Soc. Opt. Eng.* 6542, 65423-65410.

Pacheco-Londoño, L.C., Primera-Pedrozo, O.M., Hernández-Rivera, S.P., (2010). Evaluation of Samples and Standards of Energetic Materials on Surfaces by Grazing Angle-FTIR Spectroscopy in "Fourier Transform Infrared Spectroscopy: Developments,

Techniques and Applications", Rees, O.J., ed., Chemical Engineering Methods and Technology Series, Nova Science Publishers, Inc. Hauppauge, NY, ISBN: 978-1-61668-835-6.

Primera-Pedrozo, O.M., Pacheco-Londono, L.C., De la Torre-Quintana, L.F., Hernandez-Rivera, S.P., Chamberlain, R.T., Lareau, R.T. (2004). Use of fiber optic coupled FT-IR in detection of explosives on surfaces. *Proc. SPIE Int. Soc. Opt. Eng.* 5403, 237-245.

Primera-Pedrozo, O.M., Pacheco-Londoño, L., Ruiz, O., Ramirez, M., Soto-Feliciano, Y.M., De La Torre-Quintana, L.F., Hernandez-Rivera, S.P., (2005-a). Characterization of thermal Inkjet technology TNT deposits by fiber optic-grazing angle probe FTIR spectroscopy. *Proc. SPIE Int. Soc. Opt. Eng.* 543-552.

Primera-Pedrozo, O.M. (2005-b). MS Thesis, University of Puerto Rico, Mayaguez Campus, Mayaguez, PR.

Primera-Pedrozo, O.M., Rodriguez, N., Pacheco-Londono, L., Hernandez-Rivera, S.P., (2007). Detection of 2,4,6-trinitrotoluene on non-traditional surfaces using fiber optic coupled grazing angle probe-FTIR. *Proc. SPIE Int. Soc. Opt. Eng.* 6542, 65423-65410.

Primera-Pedrozo, O., Soto-Feliciano, Y., Pacheco-Londoño, L., Hernández-Rivera, S. (2008). High Explosives Mixtures Detection Using Fiber Optics Coupled: Grazing Angle Probe/Fourier Transform Reflection Absorption Infrared Spectroscopy. *Sens Imaging: Int. J.* 9 (3), 27-40.

Primera-Pedrozo, O., Soto-Feliciano, Y., Pacheco-Londoño, L., Hernández-Rivera, S. (2009). Detection of High Explosives Using Reflection Absorption Infrared Spectroscopy with Fiber Coupled Grazing Angle Probe/FTIR. *Sens Imaging: Int. J.* 10 (1), 1-13.

Primera-Pedrozo, O.M., Soto-Feliciano, Y.M., Pacheco-Londoño, L.C., Hernández-Rivera, S.P., (2010). Fiber Optic-Coupled Grazing Angle Probe-Fourier Transform Reflection Absorption Infrared Spectroscopy for Analysis of Energetic Materials on Surfaces, in "Fourier Transform Infrared Spectroscopy: Developments, Techniques and Applications", Rees, O.J., ed., Chemical Engineering Methods and Technology Series, Nova Science Publishers, Inc. Hauppauge, NY, ISBN: 978-1-61668-835-6.

Salleras, L., Donguez, A., Mata, E., Taberrer, J.L., Moro, I. and Salva, P. (1995). *Public Health Rep*., Vol 110 pp338-342.

Schrader, B. (1995). *Infrared and Raman spectroscopy: Methods and applications*, Schrader, B., Ed. VCH, New York, NY, pp. 215.

Schulte-Ladbeck, R.; Kolla, P.; Karst U. (2003). Trace Analysis of Peroxide-Based Explosives. *Anal. Chem*. 75,731-735.

Schulte-Ladbeck, U, Karst. A field test for the detection of peroxide-based explosives. (2002). *Analyst*. 127, 1152-1154.

Statgraphics Plus for Windows (1999). Statistical graphics System, User's Guide version 3.0. Statistical Graphics Corporation.

Smith, B. Fundamentals of Fourier Transform Infrared Spectroscopy, CRC Press. Boca Raton, Florida, 2000.

Soto-Feliciano, Y., Primera-Pedrozo, O.M., Pacheco-Londono, L., Hernandez-Rivera, S.P., (2006). Temperature dependence of detection limits of TNT on metallic surfaces using fiber optic coupled FTIR. *Proc. SPIE Int. Soc. Opt. Eng.* 6201, 62012-62019.

Urbanski, T. (1964). *Chemistry and Technology of Explosives*. Vol. 1, Macmillan Co.: New York, NY.

Zitrin, S.; Kraus, S.; Glasttstein, B. (1984). in: Proceedings of the International Symposium, US Government Printing Office, Washington, DC. pp 137-141.

White, G.M. (1992). An Explosive Drug Case. *J. Forensic Sci*. 37, 652-656.

13

Fourier Transform Photocurrent Spectroscopy on Non-Crystalline Semiconductors

Jakub Holovský
Institute of Physics AS CR, v. v. i.
Czech Republic

1. Introduction

Spectroscopic methods are massively used in material physics, because as seen already from Einstein's explanation of photoeffect phenomenon material response to light is wavelength sensitive. Such sensitivity is one of key characteristics for semiconductors and therefore spectral dependence of conductivity on light excitation is of strong interest. Band structure of electron configuration in semiconductors projects into the spectra of light absorption, emission or photocurrent. In Fig. 1. such projection into spectrum of absorption coefficient (details will be explained in paragraph 4.1) can be seen. Although absorption can be measured optically, very low light absorptance cannot easily be measured so the method of measurement of photocurrent upon light illumination is used instead. Basically the aim of the photocurrent spectroscopy is to obtain the curve like the one on the right of the Fig. 1.

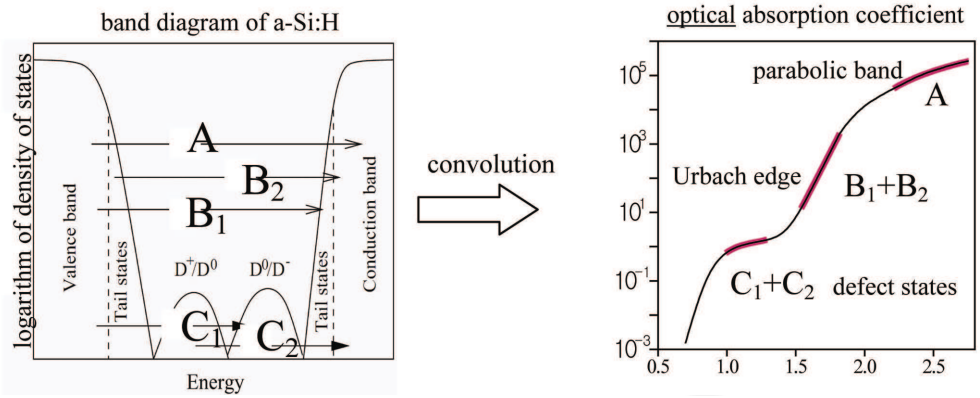


Fig. 1. On the left: band structure of density of electronic states of amorphous silicon
On the right: optical absorption coefficient curve with indicated regions attributed to different electron transitions - desired result of photocurrent spectroscopy

Photocurrent method based on the Fourier transform (F-T) and called FTPS (Fourier Transform Photocurrent Spectroscopy) is a spectroscopic method where monochromator is replaced by FTIR (Fourier Transform Infrared) spectrophotometer. This method was firstly

reported in 1997 (Tomm et al., 1997) and later (Poruba et al., 2001), (Vaněček et al., 2002) well promoted mainly for the purposes of R&D of thin film silicon photovoltaics where the quality of semiconducting layers are monitored. In earlier method CPM (Constant Photocurrent Method) keeping the photocurrent constant during the measurement by complicated modulation of light intensity was necessary (Vaněček et al., 1981). In FTPS method this condition is fulfilled automatically and also the measurement is much faster. The method can be more sensitive and more reproducible. Due to its first publications mainly on microcrystalline silicon people usually understand the FTPS simply as a method used for quality analysis of this material although its potential is not limited only to microcrystalline silicon and its use on amorphous silicon and many other non-crystalline semiconductors was successfully proven. Many practical issues of the use of FTPS are discussed in this chapter too.

In general Fourier-transform (F-T) represents in optical spectroscopy strong alternative to the classical approach based on light dispersion (monochromator). One advantage (Felgett) is that in F-T spectroscopy light of single wavelength is not isolated but only "labeled" or "encoded" so that the measurement can be performed for all wavelengths simultaneously. The spectral distribution of measured effect is obtained subsequently after mathematic decoding. Second advantage (Jacquinot) is much higher limit for resolution with the same light throughput than in monochromators. These are main reasons why F-T spectroscopy is used for example in FTIR vibrational analysis. The FTIR measurements have standardized procedures and usually not deep understanding of principles is necessary. Since FTPS is not a standard application of FTIR instrument, many other issues except the main advantages have to be considered for correct measurement and interpretation. Understanding of them is based on some fundamental principles that are in condensed form discussed in next section.

2. Principles of Fourier-transform spectroscopy

In this paragraph we want to discuss in condensed form the principal issues important for correct measurement. For F-T spectrometers many alternatives exist, but most common and instructive is the F-T spectrometer based on Michelson's interferometer, or 'modulator' as depicted in Fig. 2. Incoming light from source is partly transmitted and partly reflected by beamsplitter. Each part continues into separate delay line with mirrors at the ends. After reflection on the mirrors two beams superpose at the beamsplitter again but with mutual phase shift given by product of wavenumber[1] ν and retardation Δ, i.e. mutual path difference of the two beams. One of the mirror moves linearly with velocity u, so that retardation is time dependent: $\Delta=u(t-t_0)$. It follows from theory of light coherence that the intensity of light superposition will depend on retardation and thus will change in time t as cosine function:

$$I(\nu,t) = B(\nu)\left(1+\cos(2\pi\cdot 2\nu u(t-t_0))\right) \tag{1}$$

The factor B(ν) is baseline and represents compound spectrum of additional effects i.e.: lamp radiance, transmittance and reflectance of beamsplitter and effects of other optical elements in the instrument. Formula 1 can be regarded as 'coding key' that attributes to each wavenumber ν harmonic modulation in time with frequency f that depends linearly on the

[1] Used in infrared spectroscopy, unit is inverse centimeter cm^{-1}, $\nu(cm^{-1})=10^7/\lambda(nm)$

wavenumber according to formula 2. The modulation is for typical conditions (λ=500-1600nm, u=0.16cm/s) in the range from 2 kHz to 6 kHz. If some of the factor in the experiment exhibits frequency dependence correction to that has to be done, see paragraph 3.6.

$$f = 2\nu u \tag{2}$$

When the vavenumber is 'encoded' into modulation frequency the light beam is ready for measurement. Then other elements in the measurement should be introduced: optical filter with spectral transmittance $F(\nu)$ and detector that transforms light intensity into electrical current. Its spectral dependence can be labeled as $D(\nu)$. We make one more step and we will also go from one discrete wavenumber to continuous spectrum so that we will add the signs of integration over ν. Then we get formula 3 where the electrical current J is a function only of time t. The example of such time evolution called interferogram is in Fig. 2.

$$J(t) = J_0 + \int_{-\infty}^{\infty} D(\nu)\cdot F(\nu)\cdot B(\nu)\cdot\cos\left[2\pi\cdot 2\nu u(t-t_0)\right]\mathrm{d}\,\nu \tag{3}$$

It is very important that the detector is linear so that the current is a linear function of intensity and can be therefore represented in formula 3 linearly (y=D·x). Non-linear detectors (y=D·x+D´·x²+…) would introduce higher powers of cosine and would lead to parasitic contributions to the modulation at higher multiples of ν (higher harmonics) !
In formula 3 we can easily recognize Fourier-transform (for precision we renormalize time as $t'=(t-t_0)\cdot 4\pi u$), so that we can directly write bidirectional formulae between time domain and wavenumber domain (with **FT** as a symbol for Fourier transform):

$$J(t') = \sqrt{\pi/2}\cdot\mathbf{FT}\left(B(\nu)\cdot F(\nu)\cdot D(\nu)\right) \tag{4}$$

$$B(\nu)\cdot F(\nu)\cdot D(\nu) = \sqrt{2/\pi}\cdot\mathbf{FT}\left(J(t')\right) \tag{5}$$

Formula no. 5 finally gives simple instruction to calculate spectra from recorded interferogram. However only theoretically, because the interferogram is obviously not infinite in time and is sampled only in finite steps. The finite length has then effect on resolution through Rayleigh criterion: Two beams can be distinguished only if they differ in full retardation by full wavelength. Which transformed to domain of wavenumbers means that the theoretical resolution (in cm^{-1}) is inverse of max. retardation Δ_{MAX} (in cm). In reality due to the Gibbs phenomenon and often used triangular apodization[2] the theoretical limit of resolution will for given max. retardation Δ_{MAX} (Δ_{MAX} is twice the mirror path) be ~ $2/\Delta_{MAX}$, that can go well below $0.1cm^{-1}$. These issues are important in FTIR but not in FTPS (typically resolution $32cm^{-1}$ and triangular apodization is used). According to Jacquinot advantage the F-T spectrometer can have up to 100 times higher light throughput than dispersive instrument for given resolution (Griffiths 1977). In reality the built-in sources are optimized only for high resolution where the light throughput is comparable with monochromator with low required resolution (for thin film Si resolution Δν/ν or Δλ/λ around 0.01 is

[2] In F-T spectroscopy the spectra are not corrupted by convolution with instrument function but with F-T of envelope function that can be artificially reshaped by so called apodization so that the convolution has better representation of sharp features, but lower resolution.

enough) and thus lower resolution with high throughput is possible only with external light source, see paragraph 3.3.

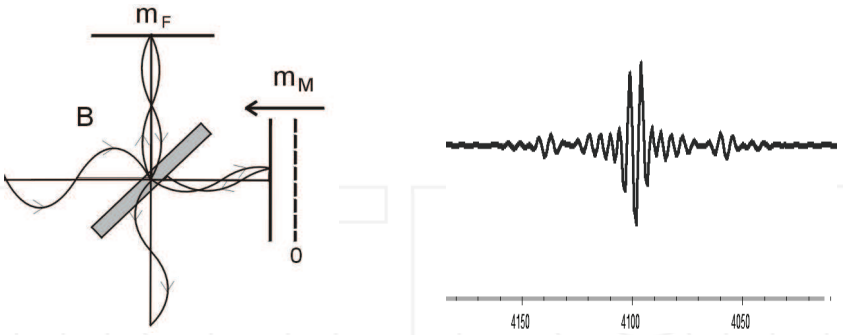


Fig. 2. On the left: Michelson's interferometer. Light enters from left side, is splitted by beamsplitter (**B**), reflected at fixed ($\mathbf{m_F}$) and movable ($\mathbf{m_M}$) mirrors and then exits downwards. On the right: Example of interferogram – amplitude of signal in domain of time/mirror position

The relation between resolution and spectral range is bidirectional and so it concerns the finite sampling density (sampling 'resolution'). Due to Nyquist-Shannon-Kotelnikov theorem that says roughly that the maximal distinguishable frequency is half of the sampling frequency, F-T spectrometer is unable to distinguish between multiples of $15798/g$ cm^{-1}, where *15798* is wavenumber of red laser and g is parameter called sample spacing[3]. This constraint unfortunately requires division of the spectrum by appropriate cut-off filters into parts from $m \cdot 15798/g$ cm^{-1} to $(m+1) \cdot 15798/g$ cm^{-1} that have to be measured separately. Fortunately modern instruments have $g=0.5$ and allow measurement up to *31600*cm^{-1} which with using halogen light as a source needs no filtering and spectrum can be measured at once. But for example measurement of FTPS of thin films silicon with $g=1$ and halogen source requires adding long-pass filter with edge above 633nm ('red glass').

Felgett (or multiplexing) advantage is basically the fact that we can measure all wavelengths simultaneously and thus measure much faster. It is also the reason why the FTPS can replace CPM method (see paragraph 4.5). These advantages have one important disadvantage. Like most measuring instruments F-T spectrometer has limited dynamic range, i.e. the ratio between highest and lowest values measured together and with small relative error. This is especially important in the case of steep absorption edge[4] in photocurrent measurement of semiconductors that is investigated always in logarithmic scale. The effective dynamic range is approximately 100 but can still vary a little according to the shape. In the case of slow slopes the dynamic range can be higher whereas in abrupt steep edges the dynamic range can be lower. Therefore for measurement of higher dynamic range spectra have to be measured with optical filters that reduce or eliminate strong parts of the spectra against weaker parts. Regarding the noise we can bring one argument that

[3] In FTIR the sampling is made with the help of red laser ($15798cm^{-1}$) interference. Sample spacing expresses distance of sampling points as a fraction of distance between 2 maxima of red laser interference.

[4] Due to the Gibbs phenomenon sharp steps are not perfectly represented by truncated F-T.

also supports the Felgett advantage: Noise is proportional to the square root of spectral range, but signal as we can see from formula 3 is proportional linearly to the spectral range, so the signal to noise ratio in turn increases with spectral range (Griffiths 1977). Therefore it is better for enhancing the dynamic range to use filters that are not cut-off so they don't totally eliminate high signal region but are 'smooth' and only reduce the high signal region and maintain nonzero signal in broad region.

Summary:

- each wavelength is modulated by different frequency $f=2vu$
- resolution is limited by mirror path length and aperture dimensions and can easily go below $0.1cm^{-1}$
- signal linearity is essential condition
- optical cut-offs are necessary for combination of halogen source with sample spacing greater than 0.5
- dynamic range is ~100 and has to be enhanced by additional optical filtering
- filtering by smooth filters is better than by cut-off filters

3. FTPS experiment

FTPS method is atractive due to its simple implementation to the research grade FTIR (Fourier-transform infrared) spectrometer that has option to external (e.g. photoacoustic) detector. FTIR spectrophotometers are since 1990's widely used for optical vibrational spectroscopy (range from 400 to 25000 cm^{-1}). They are user-friendly compact instruments including source, modulator and detector. Advantage is if the instrument has an availability of sample spacing 0.5. Then, only a high-quality low-noise current preamplifier (with voltage output), suitable optical filters and sample holder and cables are necessary for measurement of solar cells. For layers on glass additional voltage source has to be used.

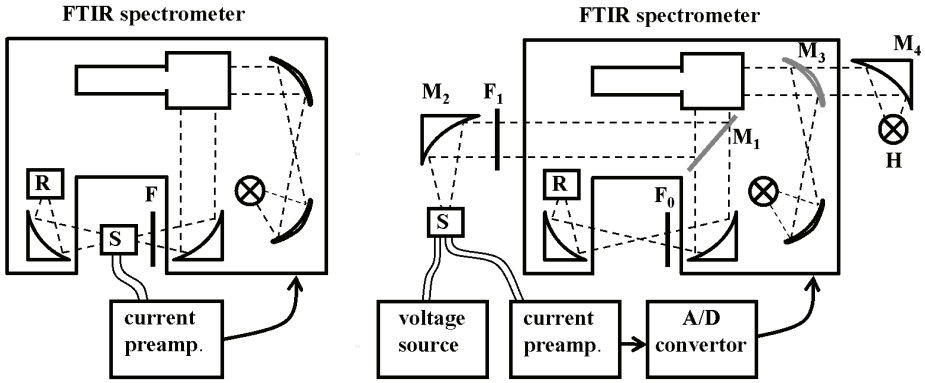


Fig. 3. On the left: 'simplest' FTPS setup where sample **S** is inside the sample area, reference **R** is measured without sample and with filter **F** (generally different for sample and reference). On the right: 'richest' FTPS setup with beam reflected out by sliding mirror $\mathbf{M_1}$ to external focusing mirror $\mathbf{M_2}$, two different positions for filter $\mathbf{F_0}$, $\mathbf{F_1}$, external source **H**, external mirror $\mathbf{M_4}$, rotating mirror $\mathbf{M_3}$, external A/D convertor and voltage source

Other components such as external focusing mirror, external A/D converter or external light source can be used optionally, see Fig. 3. Special features (not shown here) are internal motorized filter wheel and grey filter wheel. Measurement of a sample (layer on glass) in sample area has advantage of possibility of measurement of transmittance directly by internal reference detector. Measurement is performed as a sequence of inteferogram scans. The time for one scan is based on mirror velocity and max. retardation. Recorded interferograms are corrected for the phase and summed together. Signal to noise of the calculated spectrum increases with number of scans. 1 scan at velocity 0.16cm/s and resolution 32 cm^{-1} takes approx. 1 second and typically few hundreds of scans are necessary for the most 'difficult' samples. The evaluation is performed according to simple logic that is obvious from the formula 5: We compare two measurements of compound spectrum $B(\nu)\cdot F(\nu)\cdot D(\nu)$ one from sample (indexed by 1) and one from reference (indexed by 0), see Fig. 3 and formula 6. The sample and reference can play a role of any factor in the formula. In FTIR the sample (and reference) plays a role of optical filter $F(\nu)$, in FTPS the sample plays a role of detector $D(\nu)$. Baseline is for both measurement the same and its effect cancels out, as seen in formula (6). Filters can be for both measurements generally different. We correct mathematically afterwards their effects by dividing the signal by the transmittance spectrum of the filter that was used for the measurement. If we don't know the transmittances of the filters, we have to use same filters for sample as for reference which in general means a limitation.

$$D_1(\nu) = \frac{\mathbf{FT}\left(J_1(t')\right)}{\mathbf{FT}\left(J_0(t')\right)} \cdot \frac{F_0(\nu)}{F_1(\nu)} \cdot D_0(\nu) \tag{6}$$

Obviously, the response of the reference detector $D_0(\nu)$ has to be known too. Since in optics we calculate with photon fluxes rather than energy fluxes D means quantum efficiency of the photon - electron generation in the sample or detector. Absolute quantum efficiency of sample and reference detector is difficult to find out, so we always measure it relatively and additional procedures have to be used for absolute scaling, e.g. according to additional optical transmittance and reflectance measurement in medium absorption region. For the purposes of FTIR measurement, the knowledge of quantum efficiency of reference detector is not necessary and therefore it is usually unknown and for purposes of FTPS has to be found out. Typically the detectors are pyrodetectors, so they have flat response in energy flux (amperes per watt). By multiplying their response by photon energy (in electronvolts) we obtain quantum efficiency (no. of electrons per one photon). By this we obtain quantum efficiency for modulation frequencies close to zero. For obtaining quantum efficiency at real conditions of wavelength dependent modulation at 2-6 kHz frequency we could either do an approximate frequency dependence correction (paragraph 3.6) or rather do a calibration by frequency independent detector[5].

3.1 Sample preparation

Generally three types of samples can be measured: 1) Layer of semiconductor on low alkaline glass (low-cost glass will cause problems with charging of impurities) is measured

[5] Commercial calibrated c-Si detectors are unfortunately strongly frequency dependent, but can be used for low frequency calibration of e.g. thin film solar cells with low frequency dependence and these can then be used for calibration of reference detector in FTPS.

in coplanar configurations, i.e. two contacts on layer are evaporated by Al or NiCr or alternatively drawn by graphite paste. For smooth layers interdigitated contacts with thin gaps can be used, but for scattering layers spacing at least 1.5 mm is necessary (Poruba et. al 2000), details in paragraph 5.1. Voltage typically in range from 50V to 500V is applied. Light is perpendicularly focused to sample so that it has to entirely fill the gap between contacts. Sample should be illuminated from layer side, only for ideal smooth layers or for comparative measurement illumination from substrate side is possible (paragraph 4.2). Advantage of these samples is possibility of transmittance measurement (Fig. 3) at the same time and easy and accurate interpretation of the results, details in paragraph 4.2. Disadvantage is that these samples do not correspond to material grown on real substrates. 2) More close to real conditions e.g. in solar cell technology might be layers grown on conductive substrates as ZnO or SnO_2 coated glass. These have to be measured in sandwich configuration when measured semiconductor layer is sandwiched between the conductive substrate and another planar electrode deposited on top. At least one of the electrodes has to be transparent for illumination. In this case the current flows perpendicularly to surface and thus only very small distance is between the electrodes. Usually also band bending effects are presented and so small voltages (up to 2V) have to be carefully chosen to compensate such effects (Poruba et al. 2003), (Holovský et al. 2010). 3) Solar cells on the other hand don't require any preparation and typically are measured without any voltage applied (Poruba et al. 2001). Application of voltage is used to simulate real working conditions of solar cells (Bailat 2004). Monolithic multijunctions e.g. tandems or modules have to be selectively light-biased, more information in paragraph 5.3.

3.2 Choice of the FTIR instrument

The main requirements for the FTIR instrument used for FTPS are related to sample spacing, mirror velocity, beamsplitter, light source, optical windows and detector. As described above, the lowest possible sample spacing of 0.5 is strongly recommended and spacing of 1 is necessity. Lowest mirror velocity at the lowest sample spacing in true linear mode[6] should be 0.16 cm/s. FTPS measurement is in the visible and near-infrared range (0.4-2.5 μm) and so the choice of material for beamsplitter and windows is mainly quartz or sapphire, detector should be pyrodetector (e.g. thermoelectrically cooled deuterated triglycin sulphate), mirrors should be aluminum and light source should be halogen lamp. Choice of the halogen lamp is especially advantageous due to its intensity increase towards regions of low absorptance of semiconductors. Usually beamsplitter, source and detector are exchangeable so that the spectrophotometer can still be used for FTIR infrared range.

3.3 External focusing mirror, external A/D converter and external light source

The external A/D converters supplied by the same manufacturers can have better performance that the built-in A/D converters in terms of number of bits, possibility of gain ranging and signal to noise ratio that can be higher significantly. The use of external A/D converter can be for some instruments impossible for measurement in the sample area and has to be combined only with external focusing mirror. External focusing mirror is more flexible in terms of size of samples, additional light biasing etc. see Fig. 4. Also the beam can

[6] Sometimes low speed modes are not true linear scans, but in principle fast step-scan modes (for step-scan mode see paragraph 6)

be focused to smaller spot to get higher intensity. Typical focal length of internal mirror is ~15 cm with the focused beam spot size 2x8 mm².

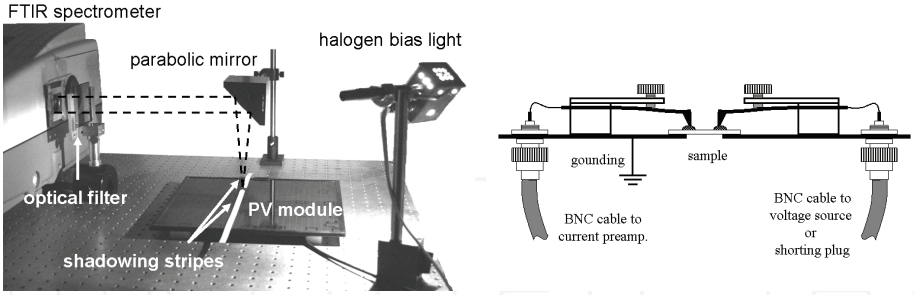


Fig. 4. On the left: FTPS setup with the use of external focusing mirror for measurement of maps of density of defects of PV modules. For selecting one spot shadowing stripes and bias light was used. On the right: sample holder for layers on glass and solar cells that have both contacts on the same side of glass substrate

The external light source can significantly increase intensity of light and also enhance its intensity in blue region. Internal halogen source has typically power of 25W, dimensions of filament approximately 2x8 mm², is intentionally kept at lower voltage and is coupled to the bench by mirror with high focal length for better resolution (~15cm). The photon flux at focus is around $2x10^{17}cm^{-2}s^{-1}$. If we use halogen lamp with 75W power with filament of comparably equal size[7] and with combination of a mirror with lower focal length (~10cm) we get approximately 7 times higher intensity. Then the resolution is no more guaranteed by automatically controlled internal aperture and has to be checked according to formula 6 where d is largest dimension of light source (or entrance slit), f is focal distance and ν_{MAX} is maximal wavenumber in the spectrum. For details see (Griffiths & de Haseth 1986).

$$\Delta\nu = (d/2f)^2 \nu_{MAX} \tag{6}$$

For described external source and for ν_{MAX}=20000cm^{-1} resolution is $\Delta\nu$~32 cm^{-1} that is what is normally used. For maximizing intensity going to resolution $\Delta\nu$ ~100cm^{-1} is possible and intensity grows 4 times. Conservative estimate of potential intensity increase of external source is 20 times. Advantage of external source is that the FTIR instrument is more thermally stable. Disadvantage is that the bench alignment can be done only for either source, see paragraph 3.7.

3.4 Current preamplifier, voltage source and sample holder

The choice of these components will strongly affect the signal to noise ratio and thus the sensitivity and speed of the measurement. For the voltage source there is no necessity to buy expensive sources. For low voltages simple battery source is better than expensive programmable sources. For high voltages (up to 500V) some high voltage source has to be used. The important issue is the scheme of serial connection of the voltage source, sample

[7] 12V halogen lamp nominally 100W designed for vehicles from manufacturer NARVA

and current preamplifier in terms of electromagnetic noise. The use of BNC 50Ω coaxial cables has proven to be much better compared to connection by twisted pair cables. The connection in series can be easily realized by the design of a sample holder where both outer wirings are connected to metal frame and each inner wire contacts one pole of the sample, see Fig. 4. Grounding of the metal frame usually reduces noise level too. In the case of measurement without voltage source (solar cells) shorting plug is used at one of the BNC connector.

Most important is the choice of current preamplifier and state-of-the art instruments are recommended. Important parameters are: noise level, frequency cut-offs, input impedances and dynamic reserve. The noise level of preamplifier at high amplification is more critical for highly resistive samples of layers on glass (1-100GΩ). For amplification 10^7 V/A broadband noise should not be above 1 picoamper. On the other hand for measurement of solar cells and modules with generally low resisitivity (10-100Ω) dynamic reserve[8] is critical. The frequency cut-offs are inequitable in all preamplifiers at high amplifications and cause frequency dependence of signal (see paragraph 3.6) and prevent from using high amplifications. Modulation frequency range is given by formula 2 and for our spectral range and range of mirror velocities (0.16– 0.47cm/s) frequency range in low signal (=high amplification) region is 1kHz – 10kHz. High amplifications cannot be used also due to increasing input impedance. Two different preamplifiers one with lower noise level and one with higher dynamic reserve were tested on solar cell and layers on glass and it was impossible to make any unambiguous preference, see Fig. 5.

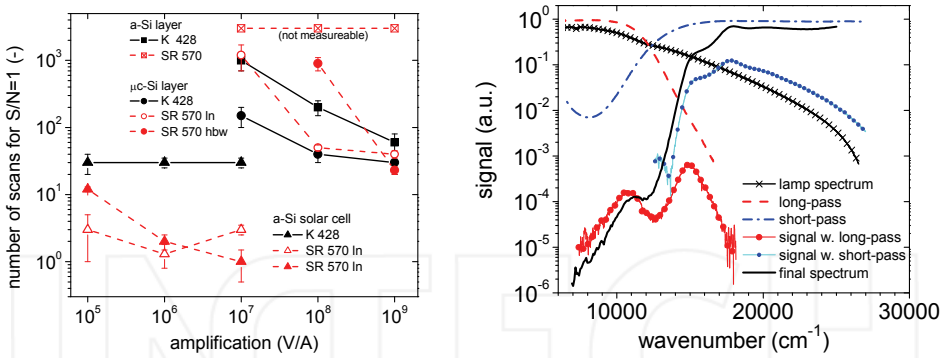


Fig. 5. Left: comparison of number of scans needed for signal to noise ratio S/N=1 for preamplifiers Standford Research 570 and KEITHLEY 428, data are based on real measurements of layers of amorphous (a-Si) and microcrystalline (μc-Si) silicon on glass and for a-Si solar cell. ('ln' = low noise mode, 'hbw'= high bandwidth mode).
Right: FTPS spectra of 250nm thick amorphous silicon measured only with 2 filters (long-pass and short-pass), transmittances of the filters and source intensity spectrum included

[8] Dynamic reserve express how larger can be broadband noise than the signal before overloading – depends on setting and location of frequency band filters

3.5 Optical filters

There are at least four reasons for using filters in FTPS: 1) Measurement with sample spacing greater than 0.5 requires using cut-off filters for both sample and reference (for details see paragraph 2). 2) Filters are necessary for improving dynamic range too. One possibility is to use cut-off filters, such as 2mm polished Si wafer for measurement of microcrystalline silicon (Poruba et al. 2002) or even set of filters for amorphous (Melskens et al. 2007) to eliminate more and more from high absorption region and to get to lower and lower absorptances. Alternative is to use less filters or even only one that only reduces light intensity at high absorption region. If the minimum of signal S that we want to measure is S_{MIN} then transmittance of ideal filter would be 1 in region where $S<100 \cdot S_{MIN}$ and $100 \cdot S_{MIN}/S$ where $S>100 \cdot S_{MIN}$ then we would measure whole curve with dynamic range only 100. For example see spectrum of long-pass filter in Fig. 5. 3) Third reason for filtering is that F-T spectroscopy doesn't reproduce spectrum where sharp step occur (absorption drop on the bandgap in amorphous Si at 1.75eV) and better results are obtained when using the filter that reduces the signal before the abrupt drop comes, see short-pass filter in Fig. 5. 4) Fourth reason for filtering is when it is necessary to keep low signal conditions. For that color or neutral density filters can be used (see section 4.5). The use of optical filters complicates strongly the automation of the FTPS because generally for different materials different filters have to be used.

3.6 Frequency dependence correction

Here we come to the main issue of FTPS. The frequency of modulation ranges from 1kHz to 10kHz and is wavenumber dependent as seen from formula 2. If performance of any part of experiment (sample, reference, preamplifier) is frequency dependent a correction for this dependence has to be done in order to obtain results same as measured at modulation frequency close to zero. Because sample and reference has different frequency dependence, formula 5 can be used only after such correction. We can suppose that the signal S at certain wavenumber ν and certain frequency f can depend on a frequency either according to formula 7 or formula 8, that means that the function Φ describing frequency dependence can be wavenumber dependent and so has wavenumber as parameter.

$$S(\nu, f(\nu)) = S(\nu, 0) \cdot \Phi(f(\nu)) \tag{7}$$

$$S(\nu, f(\nu)) = S(\nu, 0) \cdot \Phi(f(\nu), \nu) \tag{8}$$

Function Φ is unknown and can be investigated only experimentally for each sample. In FTPS the way of investigating function Φ is to make additional measurements at different velocities and observe the signal change with velocity. But because we can go from velocity 0.16 cm/s only to higher ones, function Φ can be found only around some central frequency, say 5kHz. Based on formulae 7 or 8 we can calculate signal at this frequency and so get from measured signal spectrum $S(\nu,f(\nu))$ to new spectrum $S(\nu,$ 5kHz$)$ that corresponds to spectrum as if measured at one wavelength independent frequency 5kHz. Obviously we can't get further to zero frequency. For some samples we can suppose that formula 7 holds and from it follows that $S(\nu,$ 5kHz$) = S(\nu, 0) \cdot constant$ and so we can get correct, but relative spectrum. This explains why even for 'good' frequency dependence (formula 7) we can't measure absolutely. For high preamplification, major part of frequency dependence is caused by preamplifier for which formula 7 holds. In practice it has only sense to assume

wavelength independent frequency dependence obeying formula 7. As we will see in paragraph 4.3 and 4.6 wavelength dependence as in formula 8 can occur only for photothermal ionization effects or for samples where carrier diffusion dominates. For correction we measure spectrum at three velocities and then in set of reference wavenumber points we quantify and fit (typically by exponential) the relative decrease of signal with frequency (calculated for each point by formula 2). In next step the average of the parameter of relative decrease is made and finally one of the three curves is corrected (virtually to 5kHz). Only the curve that we want to correct (usually lowest velocity) has to be measured finely with enough scans but for the two other much less scans are necessary. Frequency dependence can be sometimes significant and can lead to shift of one end of the spectrum compared to opposite end as high as by 50%, so it is clear that the accuracy is strongly influenced by frequency dependence and so the error is usually at least few percents.

3.7 Bench alignment

For successful measurement correct bench alignment is necessary. Once properly done it does not have to be done for months. Alignment is done in commercial instruments automatically and instrument remembers alignment for each beamsplitter. It was found that the shape of a baseline spectrum depends on alignment and if bench is properly aligned the baseline (roughly spectrum of source) for quartz beamsplitter should monotonously decrease towards high wavenumbers. To do so and to enhance intensity in visible range, alignment in two steps first without filter and then with short-pass glass is necessary.

When using external source, alignement should be done as follows: 1) Make proper alignment for internal source. 2) Mark on a screen that is placed into the focus the precise position of the focal point. 3) Manually align the mirror (M_4 in Fig. 3) into axis of Michelson modulator. 4) Adjust the external source to obtain the same position and dimension of the focal point on the screen. 5) Make new alignment with external source. Then the bench will not be aligned for internal source.

Summary:

- research grade FTIR and low noise preamplifier are only necessary large investments for setting up FTPS method
- high dynamic range or low noise of the preamplifier is more important for measurement of solar cells or layers on glass, respectively
- proper choice of long-pass optical filters is main know-how in FTPS but prevents simple automation of the method
- coaxial cables and grounded sample holder is necessary for low noise signal
- frequency dependence correction is necessary for FTPS on thin film silicon
- with external source the high troughput advantage allows 20 times higher intensity
- proper two-step bench alignement is necessary for accurate measurements

4. Interpretation of FTPS

In the section 1 and Fig 1 we outlined the aim of optical spectroscopy in disordered semiconductors, e.g. amorphous silicon (a-Si) that is to get information about defect density and disorder. In section 3 we covered most of the technical issues regarding the successful measurement of a FTPS spectrum. Nice thing would be to show in this part a formula for FTPS signal as a function of defects and disorder. Unfortunately situation is much more

complex and to puzzle out its complexity a diagram (Fig. 6) is sketched where all the arrows represent one specific relation that will be discussed hereafter. In spite of the complexity of the diagram, the individual effects themselves are quite straightforward and well known in the field of semiconductors.

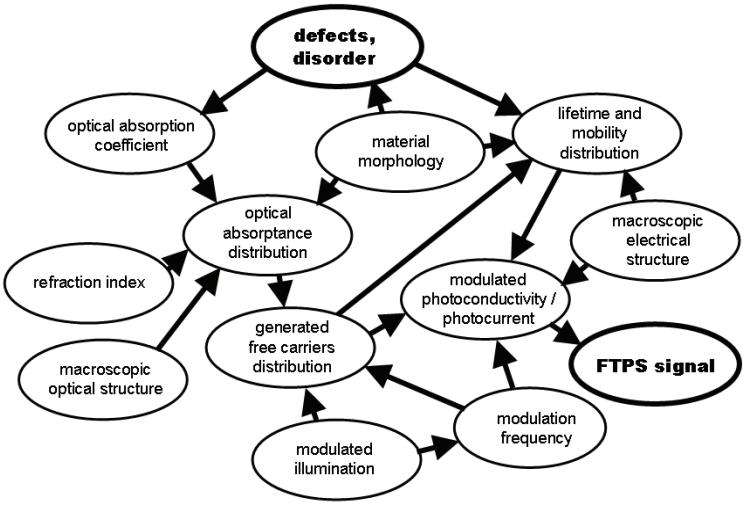


Fig. 6. Diagram shows most of the possible effects playing between defects density and disorder as input that we want to know and FTPS signal as output that we measure

4.1 From defects to absorption coefficient

Defects and disorder (their origin base on material microstructure is beyond the scope of this article) are represented as electronic states in material and characterized by spatial density and energy. In non-crystalline semiconductors these are only two parameters fully characterizing the electronic structure of material, because due to high electron scattering on irregularities no momentum quantum number and its conservation exist. So the conditions for optical excitation between occupied and unoccupied state are only energy conservation and spatial overlap of the initial and final state. Mathematically the absorption coefficient can be expressed as convolution of spatial density of occupied and unoccupied states by formula 9, where $W(\hbar w)$ is matrix element the form of which is not consensual[9].

$$\alpha(\hbar\omega) = W(\hbar\omega)\int N_V(E)\ N_C(E+\hbar\omega)\ dE \tag{9}$$

This frequently used formula holds for undoped material under low light conditions (Fermi level E_F is well defined and is close to the middle of the gap) and also low temperature limit is considered (states above E_F are unoccupied, states below E_F are all occupied).

[9] $W(E)=const$, e.g. in (Davis 1970), $W(E)= const \cdot E$, e.g. in (Jackson et al. 1985), $W(E)=const/E$, e.g. in (Vaněček et al. 1984)

material	defect energy	typ. good value	scaling constant (cm^{-2})	method	reference
a-Si	1.2 eV	<1 cm^{-1}	2.4-5x10^{16}	CPM	(Wyrsch et al. 1991)
		<2 cm^{-1}	1.2-2.5x10^{16}	PDS	
μc-Si	0.8 eV	<0.1 cm^{-1}	~ 1.7x10^{17}	CPM	(Vaněček et al. 2000)
	0.7 eV	<0.1 cm^{-1}	3.6x10^{16}	PDS	(Klein et al. 2007)

Table 1. Defect absorption energies, typical values for good quality materials and factors for recalculation into defect density (cm^{-3}) for values obtained from constant photocurrent method (CPM) and photothermal deflection spectroscopy (PDS)

In the band structure of amorphous materials the Gaussian humps in the middle represent defect states or unsaturated bonds and the tilt of the edges (band tails) represent mainly angular disorder of chemical bonds. Two vertical lines are called mobility edges, separate delocalized extended states and localized tail states and define electrical gap of the material. Y-axis in the Fig. 1 is in logarithm scale and only strong transitions are identified, i.e. only transition involving always at least one state from valence or conduction band. Transitions C_1, C_2, B_1 and B_2 project into region of subbandgap absorption on absorption coefficient. There exist more possibilities of defect density evaluation from this region (Wyrsch et al. 1991). Often used is just taking the value of absorption coefficient at some physically defined energy and multiplying it by scaling constant, see Table 1. Transition B_2 is much stronger than B_1 due to its slower slope. Exponential part of absorption coefficient is according to (Shah 2010) function of both slopes of B1 and B2, but usually is attributed only to the slope of valence band tail. For its slope the Urbach energy E_0 is defined by formula 10 (Street 1991). For good quality non-crystalline silicon material E_0 is never higher than 50 meV.

$$\alpha(\hbar\omega) = c \cdot \exp(\hbar\omega / E_0) \tag{10}$$

4.2 From structure to optical absorptance

For evaluation of optical absorptance, as seen from Fig. 6 not only absorption coefficient discussed in paragraph 4.1, but also refraction index, material and interface morphology and obviously macroscopic sample structure have to be known. Exact knowledge of all parameters and accurate calculation of absorptance for arbitrary structures is insoluble problem and therefore simple well defined samples should be measured if absolute results and not only relative comparison is desired. Simplest example is smooth weakly absorbing layer on nonabsorbing semiinfinite substrate. For the purpose of absorption coefficient measurement approximate formulae no 11, 12 and 13 presented in (Ritter & Weiser 1986) can be used:

$$A \cong \frac{(1-R_1)\left[e^{\alpha d} - 1 + R_2(1-e^{-\alpha d})\right]}{e^{\alpha d} + R_1 R_2 e^{-\alpha d} - 2\sqrt{R_1 R_2}\cos(2\gamma + \delta_1 + \delta_2)} \tag{11}$$

$$T \cong \frac{(1-R_1)(1-R_2)}{e^{\alpha d} + R_1 R_2 e^{-\alpha d} - 2\sqrt{R_1 R_2}\cos(2\gamma + \delta_1 + \delta_2)} \tag{12}$$

$$\alpha = \frac{1}{d}\ln\left(0.5\left\{ (1-R_2)(1+A/T) + \sqrt{(1-R_2)^2(1+A/T)^2 + 4R_2} \right\} \right) \quad (13)$$

Meaning of the symbols are: $r_{01} = \sqrt{R_1} \cdot \exp(i\delta_1)$, $r_{01} = \sqrt{R_1} \cdot \exp(i\delta_1)$, where r_{01}, r_{12} are Fresnell coefficients of perpendicular reflectance on first and second interface, A, T, α, d are absorptance, transmittance, absorption coefficient and thickness respectively. Formula 13 is directly derived from the two previous and contains no *cos* function and so we can profit from advantage of absence of interference patterns in ratio A/T. Therefore it is advantageous for thin films to measure absorptance and transmittance spectra simultaneously. This formula however requires absolute measurement of A and T, that is not always possible and therefore scaling according to additional optical measurement is necessary. Also knowledge of R_2 is necessary. Mathematic fit of e.g. Cauchy formula parameters of refraction index or assumption of some typical spectrum of refraction index (Vaněček 1995) is possible for calculation of R_2. The least sophisticated, but still used is to approximate the formula 11 into simple form 14 or 15. Both formulae do not take interference pattern into account and should therefore be applied to smoothened curve of A. Formula 14 is very often used and indeed by analyzing its error we can find that neglecting the term with *cos* in formula 11 and further neglecting some other terms on a way to formula 14 and attributing the result to curve with averaged interference will result in error in α of -10% in low absorption region for amorphous silicon on glass. This formula however still requires knowledge of R_1 and absolutely scaled A. Thus often R_1 is assumed to be constant and then formula 15 is used. Then only saturation value $A_{REL,MAX}$ at high absorption where $1\text{-}e^{-\alpha d} \rightarrow 1$ and thickness d have to be known. Then resulting α is correct except the region above and close to the maximum point.

$$A \cong (1-R_1)(1-e^{-\alpha d}) \quad (14)$$

$$\alpha \cong -\ln\left(1 - A_{REL} / A_{REL,MAX}\right)/d \quad (15)$$

So far theory for smooth, non scattering layer on glass was used. For measurement of samples with rough interfaces or even with bulk scattering as for example microcrystalline silicon, more complicated theory has to be used (Poruba et al. 2000). Sometimes effect of scattering can be well masked and for example accurate comparison between measurements of total and specular transmittance has to be done (Vaněček et al. 1998). Even more complicated is the evaluation of absorptance in solar cells where the transmittance of transparent conductive oxide layers (TCO) comes into play, but even here correction exists (Python 2009). Otherwise FTPS (and photocurrent in general) on real solar cells is useful mainly for comparative measurements of cells with same optical structure (TCO, roughness, back reflectors).

Presented formulae are for total absorptance, but characterizing light absorption by its total value is not always precise because absorption is generally not homogeneous due to many effects: 1) due to exponential decrease of intensity in absorbing material, 2) due to standing waves in thin films 3) due to inhomogeneity of material and thus inhomogeneity of absorption coefficient itself, 4) due to inhomogeneously distributed defects, for example close to surfaces of grain boundaries. Last two cases are not accounted in above presented calculations and so the measurement in A/T mode is not legitimate and has an effect of non-vanishing interference patterns (Vaněček et al. 1995). Moreover this effect depends on

position of highly absorbing sub-layer in the whole layer and if it is on the back side with respect to illumination, the non-vanishing interferences are not observed, inhomogeneity is masked but generally higher absorptance is observed, see Fig. 9.

4.3 From absorptance to excess carrier generation

Until now the analysis was based purely on optics. Now we want to look closely whether all absorbed photons create free charge carriers that can eventually contribute to photocurrent. In undoped amorphous silicon for example mobility of holes is much smaller and so only the transitions terminating above the conduction mobility edge can in principle contribute to photocurrent. But at room temperature the thermal excitation may release the carriers from small depth below the mobility edge above the edge and so enable them to be electrically collected. This two-step carrier generation is called photothermal ionization. The thermal ionization however has to be fast enough to be registered by phase correlated detection, because FTPS is (like other lock-in methods) based on measurement of modulated response that is correlated with excitation pulses and carriers excited too late are not accounted. In this point we get to the issue of frequency dependence dependent on wavelength (section 3.6, formula 8). Based on theory developed for Modulated Photocurrent Method (Abe et al. 1988), we can describe the effect of thermal excitation as a dispersion of the conduction mobility edge F down to energies ΔE below original mobility edge, analogical to Fermi-Dirac function as in formula 16 where symbols ω, f_0, k_B and T have meaning of modulation frequency, attempt-to-escape frequency (in a range of 10^{12} Hz), Boltzmann constant and temperature, respectively.

$$F(\Delta E) = \left[1 + \left(\frac{\omega}{f_0} \right)^2 \exp\left(\frac{2\Delta E}{k_B T} \right) \right]^{-1} \tag{16}$$

In the work of Abe the electronic states below mobility edge having overlap with F contribute as frequency dependent photocurrent signal. But not only frequency dependent signals, but also frequency independent signals from displacement current from the same electronic state are considered in his theory. So, depending on parameters in formula 16 and depending on the ratio of displacement versus real currents, different transitions will have different effect in photocurrent. In his work mainly transitions C_1 are studied at frequencies from 10Hz to 10kHz with strongest frequency dependence around 500Hz. Finally for studying the same effect directly for FTPS a comparison of photocurrent measurement by FTPS, CPM at 13Hz and DBP[10] method for wide range of frequencies was done (Holovský et al. 2008), see Fig. 8. From the comparison we see the effects of frequency dependence and displacement currents in defect region and beginning of Urbach edge in DBP at different frequencies. We also see difference between FTPS and CPM at defect reference energy 1.2 eV well below factor 2. According to previous papers (Wyrsch et al. 1991), CPM itself gives values of defect absorptance approximately twice lower than optical absorption. Similarly, our measurements systematically show difference in defect absorptance between FTPS and Photothermal Deflection Spectroscopy by factor around 2 at energy 1.2eV for amorphous silicon. It means that photocurrent generation is calculated according to same optical

[10] DBP stands for Dual Beam Photocurrent, method is explained in paragraph 4.5.

absorption coefficient as in paragraph 4.1 except the region of defect absorption where the photocurrent generation in amorphous silicon can be approximately twice lower. This principal discrepancy of different absorption coefficient for light propagation and for photocurrent generation shall not complicate our interpretation, because since the differences are in region of low homogeneous absorption ($1\text{-}exp(-\alpha d) \approx \alpha d$) we make no error by assuming light propagation according to the same absorption coefficient as corresponds to photocurrent generation. Only we have to choose correct recalculation constant in Table 1 (for FTPS same as for CPM method).

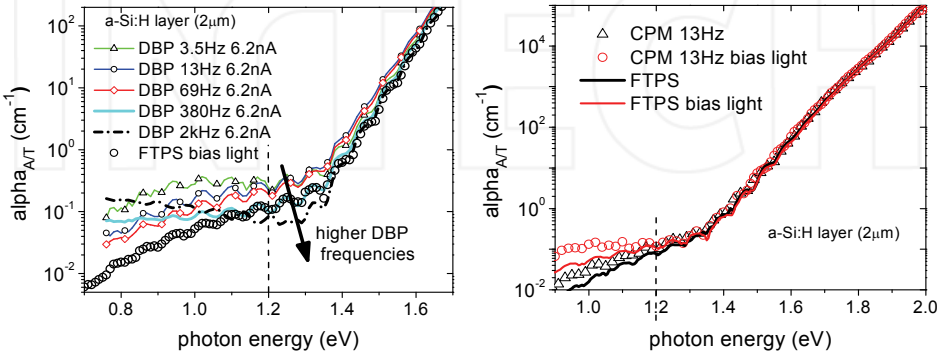


Fig. 8. Left: Dual Beam Photocurrent (DBP) spectra on 2μm amorphous Si on glass converge at higher frequencies to FTPS under bias light, at 380Hz and 2kHz DBP show anomalous deviation due to displacement currents. Right: Similar effect of light bias on FTPS and CPM methods and their comparison on the same sample

4.4 From structure to mobility lifetime distribution

When the excess carriers are generated, their efficient conduction to the collection electrodes depends on ability of the sample to conduct electrical current. As seen in Fig. 6 it depends mainly on microscopic morphology, but also presence of surfaces and so on macroscopic geometry. And also on excess carrier distribution itself, but it will be subject of next paragraph. Photocurrent in semiconductor, i.e. difference between electrical current in dark and under illumination is in simplest case given by excess carrier density n multiplied by carrier mobility μ, see equation 17. Excess carrier density is result of competition between photogeneration G described in previous paragraphs and carrier recombination that depends in simplest approximation linearly on excess carrier density. Such simple equation has steady-state solution saying that excess carrier density is a product of photogeneration G

$$j_{ph} = eE \cdot \mu \cdot n = eE \cdot G \cdot \mu \cdot \tau \tag{17}$$

and value τ called lifetime. Lifetime is a measure of recombination that in non-crystalline semiconductors is dominated by Shockley-Read-Hall recombination through active defects. Carrier mobility μ on the other hand depends on material disorder. So there are basically two material parameters that can influence conduction of generated carriers. In homogeneous samples usually also μ is homogeneous, but near surfaces τ is lower due to

surface defects. And of course in heterogeneous material both μ and τ are generally inhomogeneous.

4.5 Carrier density and lifetime

In last paragraph we defined lifetime as a value which after multiplication with generation gives excess carrier density. Now we will discuss the fact that lifetime is not a constant but a function of the excess carrier density. We discuss this phenomenon separately, because it makes the photocurrent measurement complicated. That is because activity of recombination centers is approximately given by positions of quasi-Fermi levels that depend on excess carrier density. As a consequence, conductivity on large scale is not linear with illumination, but rather sub-linear. Due to this phenomenon it is not possible to measure photocurrent for different monochromatic light of the same intensity because with different wavelength we could get photogeneration changes over many decades and so the lifetime would not be constant. On a small scale on the other hand we still assume linearity so that applying small harmonic light modulation will result in harmonic modulation of photocurrent. But for the non-linearity on the large scales there are at least three methods that solve this problem: 1) Dual Beam Photocurrent (DBP) method applies large DC light with much higher intensity than the modulated monochromatic light so that lifetime is fixed by the level on DC intensity. Drawback of this method is that, as seen in Fig. 8 the spectra still depend on level of illumination. That is due to violating the low light condition defined on the beginning in paragraph 4.1. We assume that for monochromatic light this condition is still satisfied. 2) Constant Photocurrent Method (CPM), (Vaněček 1981) adjusts the intensity of modulated monochromatic light in order to keep photocurrent constant. It will for the constant mobility mean constant excess carriers density and so constant lifetime[11]. 3) FTPS method can be regarded as combination of both CPM and DBP, because as all the wavelengths are measured simultaneously, the level of modulated photocurrent does not change and even the illumination does not change. The only question is the low light condition that would not be normally satisfied. Fortunately this condition is critical in subbandgap region only where the long-pass optical filter eliminating majority of light has to be used (paragraph 3.5).

4.6 From distributions to electrical current

We already discussed all physical properties that play role in photoconductivity: absorption coefficient, amount of absorbed light, generated excess carriers, mobility lifetime product. We already discussed their properties and mutual dependencies. In principle none of these physical values are constant and can be distributed non-uniformly. In this paragraph we should put together these distributions to get the distribution of photocurrent in our sample. So far we have not accounted for time in our thinking because all the processes we have discussed can be regarded as infinitely fast. Carrier conduction through sample is however not always so fast with respect to modulation frequency in FTPS. Treatment of such situation accurately is complicated. In simplified case of uniform doping and uniform temperature without taking into account Poisson equation we solve drift-diffusion equation 18 and continuity equation 19, where n , $\mathbf{j}$, G , τ , $\mathbf{E}$ are excess carrier concentration, current

[11] This implication may not hold only if both electrons and holes play significant role and total photocurrent is combination of both. This is not a case in non-crystalline silicon.

density, generation rate, lifetime and electric field respectively, other symbols have their usual meaning. Time enters the problem by time dependent generation *G(t)*. Solving such equations is not realistically possible and so conditions for any simplification are desirable: 1) time modulation is slow 2) lifetime is uniform in bulk, 3) absorption coefficient is uniform in bulk 4) absorption is low so that (except standing waves) the light intensity is uniform, 5) samples are thick and so without standing waves, 6) there is no additional absorption due to surface defects 7) there is no additional recombination due to surfaces.

$$\mathbf{j} = \mu \cdot \left(e\mathbf{E} \cdot n + k_B T \cdot \nabla n \right) \tag{18}$$

$$\frac{\partial n}{\partial t} = \frac{1}{e} \nabla \cdot \mathbf{j} + G(t) - \frac{n}{\tau} \tag{19}$$

a. If all 7 conditions are satisfied, solution of equations 18 and 19 leads to simple equation 17 and total photocurrent density is equivalent to total generation and so equivalent to total optical absorptance and according to paragraph 4.2 we can use A/T mode and formula 13. The same will practically work even if samples are thin and condition 5 is not satisfied.
b. If condition 7 is not satisfied and neither absorption is low (condition 4) we will get the situation that was for electric field applied perpendicularly to illumination described by theory of DeVore (DeVore 1956). It is the effect of signal loss at high absorption region when carriers are generated close to surface. Same theory however describes well also situation in high absorption for thin samples (when neither conditions 5 is satisfied) see Fig 9.
c. If the conditions 7 and 5 are not satisfied, and absorption is low or moderate (and not high as in previous case) the effect of standing waves with maxima close or far from surface will affect modulation of interference maxima in photocurrent. In this case the A/T mode may give some non-vanishing interference maxima in whole spectrum depending on the direction of illumination with respect to defective surface, see Fig. 9. Instead of A/T approach (formula 13) averaging of interference maxima and formula 14 should be rather used. If instead of condition 7 condition 6 is not satisfied then the effects on non-vanishing inreferences in A/T mode are observed only in low absorption regions where additional absorptance is comparable with bulk absorptance (Vaněček et al. 1995). Arguing that these two effects compensate is not very safe, because they affect different regions, see Fig. 9. In these cases the sample thickness should be chosen big enough (1 μm or more) so that the surface areas with low mobility or with additional absorption have negligible effects compared to bulk absorptance and bulk photoconductivity.
d. If one of the conditions 2 or 3 are not satisfied, for example in microcrystalline silicon, situation is overcomplicated and exotic effects can be studied e.g. by comparison of coplanar and sandwich arrangement (Unold et al. 2000) or by light biased CPM (Siebke et al. 1998).
e. Concerning the 1st condition of slow modulation, we assume that it is satisfied when $\tau << f^{-1}$ (lifetime much shorter than time period of modulation) and then $G(t) \approx G$ and thus $\partial n/\partial t \approx 0$ in equation 18. For FTPS f^{-1}=0.2ms and so for non-crystalline silicon where τ=1-10μs condition is well satisfied whereas for crystalline silicon wafers where

τ=0.1-10ms not and at least in combination with high or moderate absorption (condition 4) this leads to strong frequency dependence, and so for example using FTPS for photoelectric measurement on c-Si wafers is problematic, see Fig. 9.

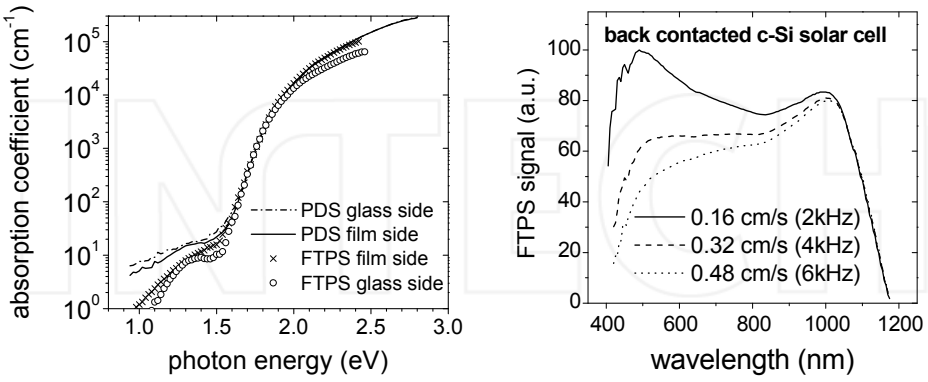


Fig. 9. Left: Result of A/T of FTPS and Photothermal deflection specroscopy (PDS) for a 350nm thick a-Si layer on glass with defective surface layer. We see non-vanishing interferences in film side measurement at low absorption in PDS and in whole range in FTPS. Deviation of PDS and FTPS in high absorption region is partly given by surface recombination. Right: Effect of high frequency modulation for c-Si solar cell observed by use of different mirror velocities

Summary:

- classical absorption coefficient in disordered materials results from convolution of density of states in low temperature limit and under dark conditions
- for optically homogeneous smooth layer A/T ratio has no interference maxima
- generation of conductive excess carriers follows optical absorption coefficient except the region far from mobility edge below the gap where its contribution is frequency dependent
- FTPS gives in defect absorption approximately twice lower value
- photocurrent depends not only on optical absorptance, but also on distribution of mobility lifetime product
- thus only for electrically (mobility, lifetime) and optically (absorption coefficient) homogeneous samples photocurrent is equivalent to optical absorptance and also A/T mode can be used
- to reduce inhomogeneity effects induced by surfaces, thicker samples (1 micrometer or more) should be used
- even for homogeneous samples measured photocurrent still depends on modulation frequency versus sample thickness and so FTPS measurement of thick samples with lifetime well above 0.2ms (c-Si wafers) can be problematic
- moreover lifetime itself is generally dependent on excess carrier density and thus 'constant photocurrent' methods like CPM or FTPS with long-pass filter are used.

5. Review of FTPS applications

Two papers give broad overview about variety of use of o FTPS: (Vaněček & Poruba 2007), (Poruba et al. 2008). In the following we will discuss separately in detail some scientific and also industrial applications.

5.1 Microcrystalline silicon

In the paragraph 4.5 we discussed the capability of use of FTPS for amorphous silicon due to its specific properties. Microcrystalline silicon is heterogeneous material with significant amorphous fraction and therefore basically same arguments can be used in case of FTPS too. Probably the first impulse was relative simplicity and good sensitivity especially on μc-Si that has advantageous shape of absorptance and with using halogen lamp as a source only one additional measurement with silicon filter is needed. The first reports of FTPS was (Poruba et al. 2001) on μc-Si p-i-n solar cells, later (Vaněček et al. 2002) on μc-Si layers on glass and later (Poruba et al. 2003) also on solar modules and layers on ZnO coated glass. As interpretation of FTPS spectra Urbach slope as a measure of disorder and absorption coefficient at energy 0.8eV as a measure of defect concentration (see Table 1) can be used. The effect of surface or bulk scattering in both p-i-n cell and layer on glass can be large and has to be corrected, see Fig.10. For layers on glass the effect depends strongly on spacing of the electrodes and minimal spacing 1.5 mm is recommended, then correction based on known surface roughness (Poruba et al. 2000) is necessary. Method of evaluation of μc-Si solar cells avoiding the effect of ZnO was well developed by Python (Python 2009). Absolute scaling can be in the case of smooth layers on glass made by approach developed for 'absolute CPM' (Vaněček et al. 1995). Without knowledge of thickness or for solar cells the approximate scaling according to crystalline silicon can be used: value at 1.35 eV is either scaled directly to value of c-Si 245cm^{-1} and we can call it normalization back to crystalline fraction (Python 2009) or is scaled to value 245cm^{-1}*Φ_c , where Φ_c is crystallinity[12]. For strong effect of scattering scaling at 1.2eV to 25cm^{-1} can be approximately done because the factor of enhancement due to light scattering changes only a little between 0.8eV and 1.2eV (Poruba et al. 2008). Broad study of microcrystalline silicon by FTPS was made at Université de Neuchâtel: (Bailat 2004), (Sculati-Meillaud 2006), (Python 2009) and correlation between FTPS and intragrain or grain boundary defects and solar cell deterioration was well verified. Microcracks in solar cells as another type of defects in solar cells are however not visible by FTPS method (Python et al. 2010).

5.2 Application of FTPS in industry at Oerlikon Solar

The power output of solar modules depends on several PECVD layers and also on several manufacturing steps before and after the PECVD deposition. To improve the module efficiency each layer and the interface between them has to be optimized. The optimization is simplified if a parameter search for each layer can be done and evaluated independently. A single layer is much faster to coat and in addition the evaluation is not affected by variations in the process steps required to manufacture a cell. The FTPS measurement is a

[12] In region around 1.35 eV α scales approximately linearly with crystallinity, however more accurate is effective medium theory (Vaněček et al. 1998) and according to (Siebke et al. 1998) scaling in broad range of crystallinity is rather exponential

relatively fast method to evaluate the defect density, which is a necessary requirement for a high quality solar cell layer.
The evaluation approach depends on the layer quality. On a coarse level poor depositions are always identified when the absorption at 0.8eV is very high e.g. above 1.0cm^{-1}. For fine tuning of already good quality layers with absorption values in the range of 0.05cm^{-1}, a comparison of only the same deposition setup and sample material is possible. Changes in the substrate e.g. with additional SnO_2 layer or different glass type leads to a change in the absorption value. This is true even if two similar glass types are compared: Schott AF32eco and Schott AF45 are coated with mc-Si in the same run, however, the FTPS data in Figure 10 shows a higher absorption for AF45 glass compared to AF32 type.

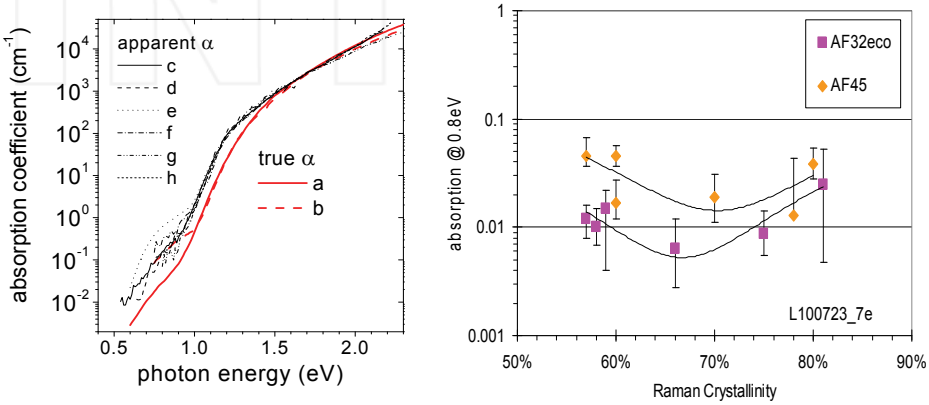


Fig. 10. Left: True absorption coefficient (a,b) and apparent absorption coefficient of 6 different materials (c-h) from different labs that all exhibit similar deviation due to the effect of scattering. Right: Effect of substrate on defect absorption for layers made in a single deposition as an example of using FTPS for module optimization in industry at Oerlikon Solar

Deposition on different substrate materials can have slightly increased absorption while still giving improved cell results. An explanation for this, is that the FTPS measurement is currently measured on AF32 glass samples placed on top off the module glass substrate. The module glass is of a different quality selected for the module requirements. The growth of the layer is affected by the different substrate surfaces. So a deposition parameter leading to a perfect layer on the Schott glass (for FTPS measurement) might lead to a non optimal layer on the large size module glass.

5.3 Amorphous silicon

FTPS on amorphous silicon is slightly more difficult than for µc-Si, because it requires different optical filters than just crystalline silicon, see Fig.5. Moreover Photothermal Deflection Spectroscopy (Jackson & Amer 1982) that is not enough sensitive for good quality µc-Si is usually enough sensitive for a-Si. The first published results of FTPS on a-Si appeared later and originated mainly from University of Delft, e.g. (Melskens et al. 2008).

Principally only FTPS can replace slower method CPM that was especially designed for amorphous silicon (paragraph 4.5) and it was shown by comparison between CPM and FTPS on a-Si (Holovský et al. 2008). FTPS was also performed on layers on a-Si codeposited on different substrates including rough conductive and non transparent Al. Special sandwich arrangement with glass covered by conductive oxide together with transparent conductive liquid (glycerol) as front contact was used. Bias voltage -1.5 V on front contact was used to create homogeneous electrical field and to compensate band bending effects. Recalculation from optical absorption to absorption coefficient was done by Monte-Carlo optical simulator, details in (Holovský et al. 2010). Measurement of multijunction solar cells monolithically interconnected was performed with the help of light biasing of the cells that are not measured, see Fig. 4. Results of such measurements can be found in (Poruba et al. 2001), (Vaněček et al. 2007) and especially in (Holovský et al. 2007) the conditions and limits of measurements tandem cells are analyzed. In tandem cells it is not possible to spatially separate the light bias or measurement beam for individual sub-cells and so the modulated generation occurs in all sub-cells. Limits of discrimination of signal from not measured cells depend on quality of diode and also on modulation frequency vs. capacity of measured cell. This unfortunately might be a problem for FTPS where high frequency modulations are used.

5.4 Nanocrystalline diamond, CIS and organic semiconductors

Together with development of FTPS on μc-Si the method was also used for measurement of subbandgap absorptance of defects and dopants of nanocrystalline diamond layers prepared by MW PECVD (Kravets et al. 2002), (Kravets 2005). Main issue in interpretation of the FTPS on nanodiamond is the effect of photothermal ionization (paragraph 4.3). This effect can be reduced by use of step-scan mode (paragraph 6) of FTIR spectrometer when ultimate sensitivity is reached with much slower speed of measurement (Remeš et al. 2007).
There has been efforts to use FTPS also for subbandgap absorption of chalcopyrite semiconductors (CIS, CIGS). FTPS spectra has been measured on solar cells only (Poruba et al. 2008), whereas measurement on layers on glass under room temperature has seemed impossible due to high dark conductivity. The Urbach slope as a measure of compositional disorder for interpretation of subbandgap region can be used (Wasim et al., 2001).
FTPS has even been used for measurement of photogeneration down to very low values in organic solar cells based on polymer-fullerene blends (Vandewal et al. 2009). Study of structural changes induced by annealing has been done (Poruba et al. 2008). According to (Vandewal et al. 2008), frequency dependence is not an issue in case of organic solar cells.

6. Discussion

In the presented text we did not pointed much to an alternative step-scan mode of FTIR. In step-scan mode the linear motion of mirror is separated into steps and in between them the mirror is stationary. Modulation is then realized by slow vibrations of the fixed mirror (phase modulation) or external chopper with lock-in amplifier (amplitude modulation). This option will certainly solve the problem of frequency dependence (paragraph 4.3), but will slow down the measurement so that the advantage compared to CPM method will be weakened. But in some cases this solution brings ultimate sensitivity that is usually not necessary e.g. for thin film silicon. Usually disadvantage of step-scan mode is sample

spacing (see paragraph 2) that is higher than 0.5. Very promising in the point of reducing modulation frequencies are new approaches based on arrays of different luminescence diodes that are used for measurement of spectra of quantum efficiency of solar cells (Young et al. 2008) or LCD-based F-T spectrophotometer (US patent no 6031609).

7. Summary

The Fourier Transform Photocurrent Spectroscopy being firstly published in 1997 has relatively short history and in this chapter we briefly reviewed its today's level of use and we more basically explained the general principles and conditions of use. Method is largely determined by the use of commercial FTIR spectrophotometer that is basis of the method. It can make the method attractive on one hand, but brings challenges on other hand. Future modern technology might show larger employment of the method. Principal advantages of F-T allow maintaining special condition of measurement such as constant photocurrent or constant illumination or allow achievement of high intensities. Method has so far been used for variety of materials: thin film silicon, nanocrystalline diamond, organic semiconductors and CIS compounds. In a laboratory FTPS can be sometimes replaced by CPM method (in most cases) or by Photothermal Deflection Spectroscopy (only non-absorbing substrates, lower sensitivity requirements) and its advantage over these methods is speed and sensitivity. Its disadvantage is wavelength dependent modulation frequency in range of kHz that requires correction procedure, complicates the interpretation of results and may even disqualify the method for specific types of materials. Despite that FTPS method has been proven to be useful in many scientific and also one industrial application.

8. Acknowledgments

The origin of this method as well as many pioneering results are mainly merits of Milan Vaněček and Aleš Poruba and many colleagues from Institute of Physics of AS CR v. v. i. The valuable feedback from solar cell industry is a merit of Stephan Jost from Oerlikon Solar, Trubbach.
This work was supported by Czech Science Foundation projects no. GA202/09/0417 and no. 202/09/H041 and 7th Framework program, project N2P, no. CP-IP-214134 and MSMT support 7E09057.

9. References

Abe, K.; Okamoto, H.; Nitta , Y.; Tsutsumi , Y.; Hattori , K. & Hamakawa, Y. (1988). Gap states in undoped amorphous silicon studied by below-gap modulated photo-current spectroscopy, *Philosophical Magazine B*, Vol. 58, No. 2, pp. 171-184 , (August 1988) ISSN: 1364-2812

Bailat, J. (2004). Growth, microstructure and electrical performances of thin film microcrystalline silicon solar cells, *PhD thesis*, Université de Neuchâtel, July 2004

Davis, E. A. (1970). Optical absorption, transport and photoconductivity in amorphous selenium. *Journal of Non-Crystalline Solids*, Vol. 4, No. C, April 1970, pp. 107-116 ISSN: 0022-3093

DeVore, H. B. (1956). Spectral distribution of photoconductivity, *Physical Review,* Vol. 102, No. 1, (April 1956) pp. 86-91

Gordijn, A.; Hoďákova, L.; Rath, J. K. & Schropp, R.E.I. (2006). Influence on cell performance of bulk defect density in microcrystalline silicon grown by VHF PECVD. *Journal of Non-Crystalline Solids*, Vol. 352 No. 9-10 (June 2006) pp. 1868–1871 ISSN: 0022-3093

Griffiths, P. R.; Sloane, H. J. & Hannah, R. W. (1977). Interferometers vs Monochromators: Separating the Optical and Digital Advantages, *Applied Spectroscopy*, Vol. 31, No. 6, (1977) pp. 485-495 ISSN: 0003-7028

Griffiths, P. R.; De Haseth, J. A. (1986). *Fourier Transform Infrared Spectrometry*, Wiley, New York

Holovský, J.; Poruba, A.; Bailat, J. & Vaněček, M. (2007). Separation of signals from amorphous and microcrystalline part of a tandem thin film silicon solar cell in Fourier transform photocurrent spectroscopy, Proceedings of *22nd EU-PVSEC*, pp. 1851-1854, ISBN: 3-936338-22-1, Milano, September 2007, WIP Munich

Holovský, J. ; Poruba, A.; Purkrt, A.; Remeš, Z. & Vaněček, M. (2008). Comparison of photocurrent spectra measured by FTPS and CPM for amorphous silicon layers and solar cells, *Journal of Non-Crystalline Solids,* Vol. 354, No. 19-25 (May 2008) pp. 2167 – 2170, ISSN: 0022-3093

Holovský, J. ; Dagkaldiran, Ü. ; Remeš, Z. ; Purkrt, A. ; Ižák, T. ; Poruba, A. & Vaněček, M. (2010). Fourier transform photocurrent measurement of thin silicon films on rough, conductive and opaque substrates, *Physica Status Solidi A,* Vol. 207, No 3, (March 2010) pp. 578-581, ISSN: 1862-6319

Jackson, W. B. & Amer, N. M. (1982). Direct measurement of gap-state absorption in hydrogenated amorphous silicon by photothermal deflection spectroscopy, *Physical Review B*, Vol. 25, No. 8 (April 1982) ISSN 1098-0121

Jackson, W. B.; Kelso, S. M.; Tsai, C. C.; Allen, J. W. & Oh, S.-J. (1985). Energy dependence of the optical matrix element in hydrogenated amorphous and crystalline silicon, *Physical Review B*, Vol. 31 No. 8 April 1985 ISSN: 1098-0121

Melskens, J.; Elzakker, G. van; Li, Y. & Zeman, M. (2008). Analysis of hydrogenated amorphous silicon thin films and solar cells by means of Fourier Transform Photocurrent Spectroscopy. *Thin Solid Films*, Vol. 516, No. 20 , (August 2008) pp. 6877–6881, ISSN: 0040-6090

Klein, S.; Finger, F.; Carius, R.; Dylla, D. & Klomfass J. (2007). Relationship between the optical absorption and the density of deep gap states in microcrystalline silicon, *Journal of Applied Physics*, Vol. 102, No. 10, p. 103501 (November 2007) ISSN: 0021-8979

Kravets, R.; Ogorodniks, V.; Poruba, A.; Moravec, P.; Nesládek, M.; Rosa, J. & Vaněček, M. (2002). Fourier-Transform Photocurrent Spectroscopy of Defects in CVD Diamond Layers. *Physica Status Solidi (a)*, Vol. 193, No. 3, (October 2002) pp. 502–507 (2002) ISSN: 1862-6319

Kravets, R. (2005). Fourier Transform Photocurrent Spectroscopy of CVD diamond, *Ph.D. Thesis*, Czech Technical University in Prague, 2005

Poruba, A.; Fejfar, A.; Remeš, Z.; Špringer, J.; Vaněček, M.; Kočka, J; Meier, J.; Torres, P. & Shah, A. (2000). Optical absorption and light scattering in microcrystalline silicon thin films and solar cells, *Journal of Applied Physiscs*, Vol. 88, No. 1, pp. 148-160 (July 2000) ISSN: 0021-8979

Poruba A.; Vaněček, M.; Rosa, J.; Feitknecht, L.; Wyrsch, N.; Meier, J.; Shah, A.; Repmann T. & Rech B. (2001). Fourier Transform Photocurrent Spectroscopy in thin film silicon solar cells, *Proceeding of 17th European Photovoltaic Solar Energy Conference*, pp. 2981-2984, ISBN: 3-936338-08-6, Munich, October 2001, WIP Munich

Poruba, A.; Špringer,J.; Mullerová, L.; Vaněček, M.; Repmann, T.; Rech, B.; Kuendig, J.; Wyrsch, N. & Shah, A. (2003). Fast and sensitive defect characterization and spectral response measurement of thin film silicon solar structures. *Proceeding of 3rd World Conference on Photovoltaic Energy Conversion (WCPEC3)*, p. 5P-A9-07, ISBN: 4-9901816-3-8, Grand Cube, Osaka, May 2003, WCPEC, Osaka

Poruba, A.; Holovský, J.; Purkrt, A. & Vaněček, M. (2008). Advanced optical characterization of disordered semiconductors by Fourier transform photocurrent spectroscopy, *Journal of Non-Crystalline Solids*, Vol. 354, No. 19-25 (May 2008) pp. 2421-2425, ISSN: 0022-3093

Python, M. (2009). Microcrystalline silicon solar cells: growth and defects, *PhD thesis*, Institut de Microtechnique, Université de Neuchâtel, March 2009

Python, M.; Dominé, D.; Söderström, T.; Meillaud, F. & Ballif, C. (2010) Microcrystalline silicon solar cells: effect of substrate temperature on cracks and their role in post-oxidation *Progress in Photovoltaics: Research and Applications*, Vol. 18, No. 7, November 2010, pp. 491–499 ISSN: 1099-159X

Remeš, Z,; Nesládek, M.; Bergonzo, P.; Barjon J. & Jomard. F (2007). Amplitude modulated step scan Fourier transform photocurrent spectroscopy of partly compensated B-doped CVD diamond thin films *Phisica status solidi (a)* Vol. 204, No. 9, pp. 2950–2956 (September 2007), ISSN: 1862-6319

Sculati-Meillaud, F. (2006). Microcrystalline silicon solar cells: theory, diagnosis and stability, *PhD thesis*, Institut de Microtechnique, Université de Neuchâtel, July 2006

Shah, A. (2010) *Thin-film silicon solar cells,* EPFL Press, ISBN 978-2-940222-36-0, Switzerland

Siebke, F. Yata, S. Hishikawa, Y. Tanaka, M. (1998). Correlation between structure and optoelectronic properties of undoped microcrystalline silicon, *Journal of Non-Crystalline Solids* Vol. 227–230 No. 2 (May 1998) pp. 977–981, ISSN: 0022-3093

Street, R. A. (1991). *Hydrogenated amorphous silicon,* Cambridge University Press, ISBN: 0 521 37156 2, Cambridge

Tomm, J. W.; Jaeger, A.; Bärwolff, A.; Elsaesser, T.; Gerhardt, A. & Donecker, J. (1997). Aging properties of high power laser diode arrays analyzed by Fourier-transform photocurrent measurements. *Applied Physics Letters,* Vol. 71, No. 16 (August 1997) pp. 2233-2235 ISSN 0003-6951

Unold, T.; Brüggemann, R.; Kleider, J.P. & Longeaud, C. (2000) Anisotropy in the transport of microcrystalline silicon, *Journal of Non-Crystalline Solids,* Vol. 266-269, No. 1 (May 2000) pp. 325-330, ISSN: 0022-3093

Vaněček, M.; Kočka, J.; Stuchlík, J. & Tříska, A. (1981). Direct measurement of the gap states and band tail absorption by constant photocurrent method in amorphous silicon, *Solid State Communications*, Vol. 39, No. 11, September 1981, pp. 1199-1202, ISSN: 0038-1098

Vaněček M.; Abrahám, A.; Štika, O.; Stuchlík, J. & Kočka J. (1984). Gap States Density in a-Si:H Deduced from Subgap Optical Absorption Measurement on Schottky Solar Cells, *Physica Status Solidi (a)* Vol. 83, No. 2 p. 617-623 (June 1984) ISSN: 1862-6319

Vaněček, M.; Kočka, J.; Poruba, A. & Fejfar, A. (1995). Direct measurement of the deep defect density in thin amorphous silicon films with the "absolute" constant photocurrent method. *Journal of Applied Physics*, Vol. 78, No. 10, (November 1995) pp. 6203-6210 ISSN: 0021-8979

Vaněček, M.; Poruba, A.; Remeš, Z.; Beck, N. & Nesládek, M. (1998). Optical properties of microcrystalline materials, *Journal of Non-Crystalline Solids* Vol. 227-230, Part 2 (May 1998) pp. 967–972 , ISSN: 0022-3093

Vaněček, M.; Poruba, A.; Remeš, Z.; Rosa, J.; Kamba, S.; Vorlíček, V.; Meier, J. & Shah, A; (2000). Electron spin resonance and optical characterization of defects in microcrystalline silicon, *Journal of Non-Crystalline Solids* Vol. 266-269, No. pp. 519-523 (May 2000), ISSN: 0022-3093

Vaněček, M. & Poruba, A. (2002). Fourier-transform photocurrent spectroscopy of micro-crystalline silicon for solar cells. *Applied Physics Letters,* Vol. 80, No. 5, (February 2002) pp. 719-721 , ISSN 0003-6951

Vaněček, M.; Kravets, R.; Poruba, A.; Rosa, J.; Nesládek, M. & Koizumi, S. (2003). Fourier transform photocurrent spectroscopy of dopants and defects in CVD diamond. *Diamond and Related Materials*, Vol. 12, No. 3-7, (March-July 2003), pp. 521-525, ISSN: 0925-9635

Vandewal, K.; Goris, L.; Haeldermans, I.; Nesládek, M.; Haenen, K.; Wagner, P. & Manca, J. V. (2008) Fourier-Transform Photocurrent Spectroscopy for a fast and highly sensitive spectral characterization of organic and hybrid solar cells, *Thin Solid Films* Vol. 516, No. 20 (August 2008) pp. 7135–7138 ISSN: 0040-6090

Vandewal, K.; Tvingstedt, K.; Gadisa, A.; Inganäs, O. & Manca, J. V. (2009) On the origin of the open-circuit voltage of polymer–fullerene solar cells *Nature Materials*, Vol. 8 No.11, pp. 904 - 909 (November 2009) ISSN: 1476-1122

Wasim, S. M.; Rincón, C.; Marín, G.; Bocaranda, P. & Hernández, E. (2001). Effect of structural disorder on the Urbach energy in Cu ternaries. *Physical Review B*, Vol. 64, No. 19, 195101 (October 2001) ISSN 1098-0121

Wyrsch, N.; Finger, F.; McMahon, T. J. & Vaněček, M. (1991). How to reach more precise interpretation of subgap absorption spectra in terms of deep defect density in a-Si:H *Journal of Non-Crystalline Solids* Vol. 137-138 Part 1 pp. 347-350 (1991) ISSN: 0022-3093

Young, D.L.; Pinegar, S.; Stradins, P. & Egaas, B. (2008) New Real-Time Quantum Efficiency Measurement System *Proceedings of the 33rd IEEE Photovoltaic Specialists Conference*, San Diego, California May 11–16, 2008 Conference Paper NREL/CP-520-42509

14

Fourier Transform Rheology: A New Tool to Characterize Material Properties

Massimiliano Grosso[1] and Pier Luca Maffettone[2]
[1]*Dipartimento di Ingegneria Chimica e Materiali, Università degli Studi di Cagliari*
Piazza D'Armi, I-09123, Cagliari,
[2]*Dipartimento di Ingegneria Chimica, Università degli Studi di Napoli "Federico II"*
Piazzale Tecchio, I-80125, Napoli
Italy

1. Introduction

Liquid multiphase systems such as polymer blends or emulsions are ubiquitous in many applications, including plastic production, food processing, pharmaceutical and cosmetic production. When the constituents of the multiphase system are incompatible the phases are immiscible, and, depending on their relative amount, the microstructure can consist of droplets in a matrix, elongated fibrils or a co-continuous structure (Utracki, 2003) as schematically shown in Figure 1. The morphology of the liquid multiphase system is important in the applications as it strongly affects processing properties, and the properties of the final products. With the term "morphology" we here indicate not only the overall form or shape of the physical structure of the system, but also the distribution and orientation of the phases, the interfacial area, and the volume of the interphase.
Hence, a profound knowledge of the relation between processing parameters, material properties and morphology is essential to optimize the performances of the liquid multiphase systems.

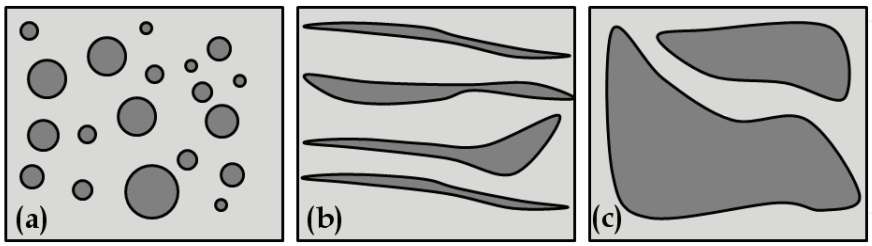


Fig. 1. Different morphologies of immiscible polymer blends (a) dilute droplet blends; (b) elongated fibrils; (c) co-continuous structure

Substantial efforts were done in the last decades to set up experimental protocols aimed at evaluating the morphological properties of polymer blends and emulsions via rheological measurements. So far, the most reliable strategy for morphological characterization through rheological measurements is based on the dynamic small amplitude oscillatory shear

(SAOS) experiment: the samples are subjected to small amplitude shearing oscillations, and the measured shear stress response is used to gain information on the blend properties (e.g. Palierne, 1990).

Here, we present an alternative technique we have recently proposed to characterize the liquid two-phase system morphology. This methodology is based on Large Amplitude Oscillatory Shear (LAOS) flows. This kind of analysis is often referred in the literature as Fourier Transform Rheology (FTR) (Wilhelm et al., 1998), since the stress response is usually analyzed in the Fourier domain.

It will be shown that Fourier Transform Rheology possesses a high sensitivity in the characterization of the morphology, thus allowing evaluation of properties that might otherwise be hardly appreciated with traditional linear methodologies.

2. Rheological oscillatory experiences

2.1 Small Amplitude Oscillatory Shear

A typical tool used for the characterization of complex liquids is based on oscillatory rheometry (Macosko, 1994). The basic working principle of an oscillatory rheological test is to impose a sinusoidal shear deformation, and measuring the resultant shear stress response. In a typical experiment, the sample is placed between two plates (or a cone and plate geometry) (Figure 2): while the bottom plate remains stationary, a rotation is imposed on the top plate, thereby allowing a time-dependent strain deformation on the sample:

$$\gamma(t) = \gamma_0 \sin \omega t \tag{1}$$

where ω is the oscillation frequency and γ_0 is the strain amplitude. The oscillation period is thus $T = 2\pi/\omega$. The resulting time-dependent shear stress, $\sigma(t)$, is quantified by measuring the torque on the top plate. At low strain amplitudes, the stress response can be assumed to depend linearly on the strain deformation:

$$\sigma(t) = A \sin \omega t + B \cos \omega t = \sigma_0 \sin(\omega t + \delta) \tag{2}$$

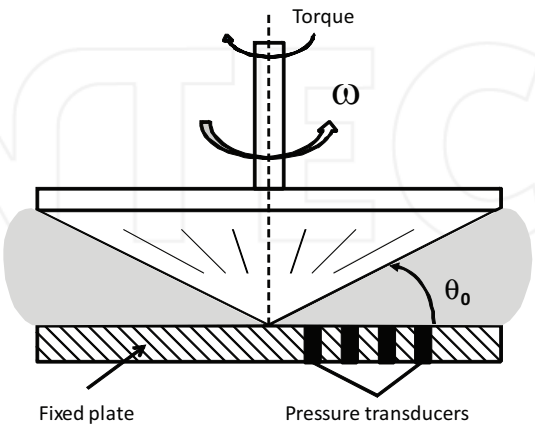


Fig. 2. Cone and plate geometry

For this reason Small Amplitude Oscillatory Shear (SAOS) tests are usually referred as *linear* rheological measurements. If the material behaves as an ideal elastic solid, then the stress is in phase with the imposed deformation wave (i.e. proportional to sin(ωt)), and the proportionality constant is the shear modulus of the material. On the other hand, if the material is a purely viscous fluid, the stress is proportional to the rate of the strain deformation (i.e. proportional to cos(ωt)), and the proportionality constant is the viscosity of the fluid. The applied strain and the measured stress are in this case out of phase with phase angle $\delta = \pi/2$.

Complex materials usually show a response that contains both in-phase and out-of-phase contributions. As a consequence, the total stress response at a given ω is characterized by both the sine and cosine components:

$$\sigma(t) = G'(\omega)\gamma_0 \sin\omega t + G''(\omega)\gamma_0 \cos\omega t \tag{3}$$

In equation 3, G′(ω) is the *storage* modulus which characterizes the solid-like behavior, whereas G″(ω) is the *loss* modulus that takes into account the fluid-like contributions. The *complex* modulus to the frequency can be thus defined:

$$G^*(\omega) = G'(\omega) + iG''(\omega) \tag{4}$$

where i is the imaginary unit. Using the relation between the complex modulus G*(ω) and the complex viscosity η*(ω),

$$G^*(\omega) = \omega\eta^* \tag{5}$$

and

$$\eta^*(\omega) = \eta' + i\eta'' \tag{6}$$

The absolute value of the complex viscosity is of course given by:

$$\left|\eta^*(\omega)\right| = \frac{\left|G^*\right|}{\omega} = \frac{\sqrt{G'^2 + G''^2}}{\omega} \tag{7}$$

The frequency dependence of G′ and G″ provides some important information about the microstructure of a material. For example, gels exhibit G′ that is larger than G″ with both moduli independent of frequency. Polymer melts show G′ and G″ at low frequencies that are dependent on ω^2 and ω, respectively. For viscoelastic materials, the overlap frequency (the frequency at which G′ and G′ curves intersect) gives information about the relaxation time of the system. The plateau modulus, i.e. the value of G′ at high frequency, gives information about the strength of the structures formed in the material.

For the case of dilute blends with Newtonian constituents, the dependence of G′ and G″ can be described in terms of the Palierne model which may quantitatively associate the linear viscoelastic properties of polymer blends to its chemical-physical properties e.g. to the interfacial tension (Palierne, 1990: Graebling et al., 1993a; Graebling et al., 1993b; Lacroix et al., 1996).

$$G^*(\omega) = G_m^*(\omega)\frac{1+3\varphi H\left(\overline{R}_V,\omega\right)}{1-2\varphi H\left(\overline{R}_V,\omega\right)} \tag{8}$$

where

$$H\left(\overline{R}_V,\omega\right) = \frac{\begin{array}{c}4\frac{\alpha}{\overline{R}_V}\left(2G_m^*(\omega)+5G_i^*(\omega)\right)+\\ \left(G_i^*(\omega)-G_m^*(\omega)\right)\left(16G_m^*(\omega)+19G_i^*(\omega)\right)\end{array}}{\begin{array}{c}40\frac{\alpha}{\overline{R}_V}\left(G_m^*(\omega)+G_i^*(\omega)\right)+\\ \left(2G_i^*(\omega)+3G_m^*(\omega)\right)\left(16G_m^*(\omega)+19G_i^*(\omega)\right)\end{array}} \tag{9}$$

and $G^*_i(\omega)$, $G^*_m(\omega)$ and $G^*(\omega)$ are, respectively, the complex moduli of the dispersed phase, matrix and blend at frequency ω, α is the interfacial tension, ϕ is the volume fraction of the dispersed phase and $\overline{R}_V$ is the volume average drop radius of the included phase:

$$\overline{R}_V = \frac{\int_0^\infty R^4\psi(R)dr}{\int_0^\infty R^3\psi(R)dr} \tag{10}$$

where $\psi(R)$ represents the drop size distribution.
An example is reported in figure 3, where the elastic modulus for a blend composed by Poly-DiMethylSiloxane (PDMS) in Poly-IsoButylene (PIB) is reported with respect to the oscillation frequency, together with the elastic modulus of the neat constituents.

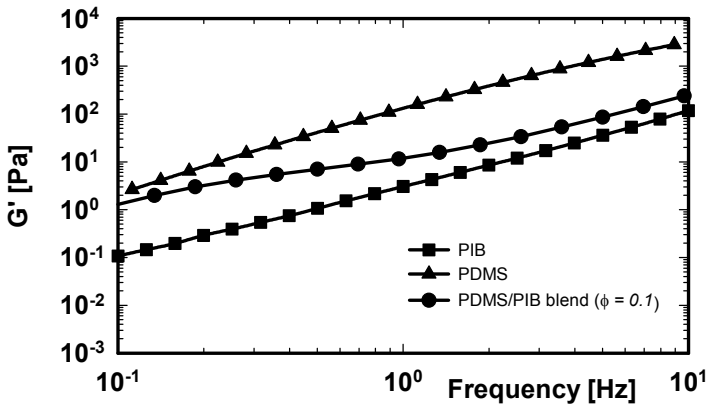


Fig. 3. Elastic modulus as a function of frequency for the pure components (PIB: square points; PDMS: triangles) and the polymer blend (full circles). The temperature is 30°C. The strain amplitude is 50%

The Palierne model is widely used to extract morphological or dynamical properties from oscillatory data, for example the average drop radius can be estimated (e.g. Das et al., 2005) or alternatively the surface tension (e.g. Huitric et al., 1998; Vincze-Minya and Schausberger, 2007). As already remarked in the literature (Graebling et al. 1993a), however, it is difficult to achieve a more detailed description of the morphology of the included phase . Indeed, a complete description of the drop size distribution cannot be reliably obtained with this technique. To our knowledge, the size distribution inference for a polymer blend based on the Palierne method has been carried out only by Friedrich et al. (1995). The methodology there proposed is based on a Tikhonv regularization, and gave satisfactory results only for very dilute blends with unimodal drop radius distributions.

2.2 Large Amplitude Oscillations

The above mentioned limits of the traditional linear oscillatory experiences motivated the study of alternative experimental techniques that might be more sensitive to the material morphology. In this regard, Large Amplitude Oscillatory Shear flows (LAOS) proved to be a possible candidate for the morphological characterization. The capabilities of this technique to pinpoint nonlinear material characteristics have been already analyzed in other contexts (Neidhofer et al., 2004; Schlatter et al., 2005), proving to be quite effective.

In the following we will briefly review some basic issues on LAOS (see Wilhelm et al. 1999 for details).When dealing with LAOS flows, the nonlinear dependences of the viscosity on the applied shear rate $\eta(\dot{\gamma})$ can be expected to be important, and equation 3 is no longer valid.

Due to the symmetry properties, the viscosity is independent of the shear direction and therefore it can only depend on the absolute shear rate (Wilhelm et al., 1999):

$$\eta(\dot{\gamma}) = \eta(-\dot{\gamma}) = \eta(|\dot{\gamma}|) \tag{11}$$

The Taylor expansion for the viscosity at small shear rates is given in equation (12):

$$\eta(\dot{\gamma}) = \eta_0 + \eta_1 |\dot{\gamma}| + \eta_2 |\dot{\gamma}|^2 + \ldots \tag{12}$$

If the applied shear deformation is a harmonic oscillation with a given frequency ω_1, strain and strain rate are:

$$\gamma = \gamma_0 \sin \omega_1 t \qquad \Rightarrow \qquad |\dot{\gamma}| = \omega_1 \gamma_0 |\cos \omega_1 t| \tag{13}$$

Therefore, the absolute value of the shear rate signal $|\dot{\gamma}|$ can be represented in terms of a proper Fourier series (Ramirez, 1985):

$$|\dot{\gamma}| = \omega_1 \gamma_0 \left(\frac{2}{\pi} + \frac{4}{\pi} \left(\frac{\cos 2\omega_1 t}{1 \cdot 3} - \frac{\cos 4\omega_1 t}{3 \cdot 5} + \ldots \right) \right) \tag{14}$$

$$|\dot{\gamma}| = a' + b' \cos 2\omega_1 t + c' \cos 4\omega_1 t + \ldots \tag{15}$$

By substituting Equation 15 into Newton's equation for the viscosity leads to equation 16 (Wilhelm et al., 1999)

$$\sigma = \eta(\dot{\gamma})\dot{\gamma} = \left(\eta_0 + \eta_1|\dot{\gamma}| + \eta_2|\dot{\gamma}|^2 + ...\right)\cos\omega_1 t \tag{16}$$

$$\sigma = \eta(\dot{\gamma})\dot{\gamma} = \begin{pmatrix} \eta_0 + \eta_1\left(a' + b'\cos 2\omega_1 t + ...\right) + \\ \eta_2\left(a' + b'\cos 2\omega_1 t + ...\right)^2 + ... \end{pmatrix}\cos\omega_1 t \tag{17}$$

The terms in brackets in Equation 17 can be thus simplified and written as a sum of even harmonics:

$$\sigma = \left(a'' + b''\cos 2\omega_1 t + c''\cos 4\omega_1 t + ...\right)\cos\omega_1 t \tag{18}$$

By multiplying the terms in the brackets by cosω_1t one ends up with the shear stress expression depending only on odd harmonics:

$$\sigma = A\cos\omega_1 t + B\cos 3\omega_1 t + C\cos 5\omega_1 t + ... \tag{19}$$

where A, B, C are complex numbers. In a last step the non-linear torque signal is analyzed towards frequency components by Fourier transformation. Eventually, the signal can be described in terms of an odd function of the sinusoidal deformation

$$\sigma(t) = \sum_{\substack{k=1 \\ odd\,k}}^{\infty} I_{Rk}\cos j\omega_1 t + I_{Ik}\sin j\omega_1 t = \sum_{\substack{k=1 \\ odd\,k}}^{\infty} I_{Ak}\cos\left(k\omega_1 t + \phi_k\right) \tag{20}$$

In equation (20) I_{Rk}, I_{Ik} and I_{Ak} are real coefficients. Straightforwardly, one can easily express the measured shear stress signal in the Fourier domain as:

$$\sigma(t) \overset{FT}{\Leftrightarrow} \tilde{\sigma}(\omega) = \int_{-\infty}^{+\infty} \sigma(t)e^{-i\omega t} = \sum_{\substack{k=-\infty \\ odd\,k}}^{k=\infty} I_k\delta(\omega - k\omega_1) \tag{21}$$

In equation 21, $\delta(\omega - k\omega_1)$ is the Dirac delta located at $\omega = k\,\omega_1$ ($k \in Z$), i is the imaginary unit, and I_k is the (complex) coefficient of the k-th harmonic. As the stress time series $\sigma(t)$ is real valued, the condition $I_k = I^*_{-k}$ (with * denoting the complex conjugate) holds. As a consequence of the assumption made in equation 19, only odd terms of the Fourier series could be in principle accounted for in equation 21 (Wilhelm et al., 1998). It is easy to show that the following relationship among the coefficients holds:

$$I_{Aj} = |I_k| = \sqrt{I_{Rk}^2 + I_{Ik}^2} \tag{22}$$

When dealing with SAOS flows only the first term of the summation in equation 21 is significant. Incidentally, one can notice that, as γ_0 tends to zero, the linear behaviour is recovered, thus $I_{I1} = G'\gamma_0$ and $I_{R1} = G''\gamma_0$, and $I_{Rk} \approx I_{Ik} \approx 0$ for any $k > 1$. The appearance of

significant values for I_k ($k > 1$) marks the onset of nonlinearities in the stress response. As a further remark, it has been shown that, at vanishing amplitudes the following scaling for intensity of nth harmonic with strain has been observed for the constitutive equations so far investigated (Nam et al. 2008; Yu et al. 2008):

$$I_{An} \propto \gamma_0^n \tag{23}$$

This allows the definition of new scalars, based on the ratios of intensities of higher harmonics and first harmonic of the stress response. For example, coefficient Q is defined as (Hyun & Wilhelm, 2009)

$$Q = \frac{I_{A3}}{I_{A1}} \frac{1}{\gamma_0^2} \tag{24}$$

This coefficient has been claimed to be helpful in distinguishing molecular architecture of polymers based on LAOS (Hyun & Wilhelm, 2009).

From a practical point of view, the measurement of the stress, through the torque sensor of the rheometer, is usually performed discretely at finite sample intervals (Δt). Based on the sampling frequency, ($r = 1/\Delta t$, number of data points collected per second), we obtain a time series $\sigma(n)$ of discrete measurements collected at N_P instants.

Discrete Fourier transform of this time domain series will be a series of N_{ST} complex numbers evaluated through well consolidated FT techniques (Bracewell, 1986):

$$\Sigma(k) = \sum_{n=1}^{\infty} \sigma(n) e^{-\frac{i2\pi(k-1)(n-1)}{N}}, \quad 1 \le k \le N \tag{25}$$

The maximum frequency in the Fourier domain will correspond to the Nyquist frequency = $2\pi/\Delta t$. With the property of the Fourier transform leading to meaningful $N/2$ (symmetric) terms, the resolution in the frequency domain is $2\pi/T$. Therefore, sampling interval determines maximum frequency to which information can be obtained, while the duration of measurements determines the resolution of frequency. Larger T values also lead to higher signal to noise ratio. It should be remarked that some techniques are introduced in the literature in order to improve the sensitivity to the signal of the measurement (Wilhelm et al., 1999).

3. FTR on polymer blends

3.1 Theory

In this section, we will focus on the theoretical aspects concerning the characterization through FTR of immiscible blends with low fraction of the dispersed phase. In this case the morphology of the included phase is globular: the basic element of such a dilute blend is thus a single drop dispersed in a matrix. Therefore, the study of single droplet behaviour is regarded as a reasonable starting point to model the complex behaviour of immiscible polymer blends. The overall rheological response, in fact, could be determined just on such a basis.

The dynamic behaviour of dilute polymer blends subjected to LAOS flows can be modeled as recently proposed in the literature by considering the single droplet dynamics together

with a proper stress expression (Rallison, 1984; Stone, 1994; Almusallam et al., 2000; Yu et al., 2002; Jackson and Tucker, 2003; Yu and Bousmina, 2003). For what matters the dynamics of the drop, a handy though effective phenomenological model has been proposed by Maffettone and Minale (1998) and applies to generic flow fields. The model is formulated in terms of at most six first-order, ordinary, differential equations, and is capable of describing drop deformation up to the nonlinear range. This model is known to be quite accurate for small-to-medium droplet deformation, but loses some quantitative accuracy as droplet deformation becomes large. We use here this model for its simplicity, even though significant distortion of drop shape is expected under LAOS. Still, the Maffettone and Minale model provides a useful basis for analyzing and interpreting the experimental results also when significant strain deformations occur (Guido et al. 2004). The drop is described as an ellipsoid by a second rank symmetric, positive definite, and time dependent tensor. The shape dynamics can be thus described by the evolution of tensor **S** which follows the equation:

$$\frac{d\mathbf{S}}{dt} - \left(\mathbf{\Omega}\cdot\mathbf{S} - \mathbf{S}\cdot\mathbf{\Omega}\right) = -\frac{f_1}{\tau}\left(\mathbf{S} - \frac{3}{I_2}\mathbf{I}\right) + f_2\left(\mathbf{S}\cdot\mathbf{D} + \mathbf{D}\cdot\mathbf{S}\right) \tag{26}$$

In Equation 26, τ is the emulsion time ($\tau=\eta R/\Gamma$) where η is the matrix viscosity, R the undistorted drop radius and Γ is the interfacial tension; I is the second rank unit tensor, **D** and **Ω** are the deformation rate and the vorticity tensors respectively, and I_2 is the second scalar invariant of tensor **S**. The shear flows here considered give the following forms for the deformation and vorticity tensors:

$$\mathbf{D} = \frac{1}{2}Ca\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad D = \frac{1}{2}Ca\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{27}$$

In equation 27 the Capillary number is introduced:

$$Ca = \frac{\text{viscous stress}}{\text{interfacial stress}} = \frac{\eta\dot{\gamma}}{\Gamma/R} \tag{28}$$

This gives the ratio between the two competing forces affecting the drop shape in shear flow experiences: the driving force of deformation (i.e. the shear stress), and the resistance force supporting the shape of the drop, that is the interfacial tension. The dependence of the capillary number on time is understood. The functions f_1 and f_2 appearing in Eq. 26 are given by (Maffettone & Minale, 1998):

$$\begin{aligned} f_1(\lambda) &= \frac{40(\lambda+1)}{(2\lambda+3)(19\lambda+16)} \\ f_2(\lambda) &= \frac{5}{2\lambda+3} + \frac{3Ca^2}{2+6Ca^2} \end{aligned} \tag{29}$$

In equation 29 the ratio $\lambda = \eta_d/\eta_m$ is defined, where η_d is the viscosity of the dispersed phase and η_m is the matrix viscosity. At rest the drop is spherical (S=I). Notice finally that within this description drop break-up is absent under shear flow for $\lambda \geq 3$.

Once the state of the drop deformation is known, one can calculate the stress **σ** of a dilute polymer blend according to Batchelor (1970):

$$\boldsymbol{\sigma} = \underbrace{-p\mathbf{I}}_{\text{isotropic term}} + \underbrace{\eta_m\left(\nabla\mathbf{v}+\nabla\mathbf{v}^T\right)}_{\text{Newtonian contribution}} - \underbrace{\frac{\eta}{V}\int_S(\mathbf{nu}+\mathbf{un})dA}_{\text{viscous term}} + \underbrace{\frac{\Gamma}{V}\int_S\left(\mathbf{nn}-\frac{1}{3}\mathbf{I}\right)dA}_{\text{elastic term}} \tag{30}$$

In Eq. 30, p is the pressure, **∇v** is the velocity gradient tensor and $\nabla\mathbf{v}^T$ its transpose, η is the viscosity of the continuous phase, V is the total volume of the system, **n** is the unit vector normal to the ellipsoid surface, representing the interface between the two phases, **u** is the velocity at the interface, *dA* is the area of an interfacial element, and the integrals are calculated over the whole interface of the system, *S*. Equation 26 can be used to predict the stresses if **n** and **u** are known. Predictions are obtained by integrating equations (26) and (29) for the drop morphology. The elastic interfacial term in equation (30) is calculated as suggested by Almusallam et al. (2004). The viscous term in the interface stress is neglected.

Equation 30 is the sum of two conceptually different terms: the first one is due to the Newtonian matrix contribution and depends linearly on the velocity gradient, whereas the second term (the viscous and the elastic term) corresponds to the sum of interfacial contributions related to the entire drop population. The first part depends linearly on the applied shear rate, whereas the interfacial contribution is the only nonlinear term appearing in Equation 29. Under LAOS, the first term will not contribute to higher harmonics in the shear stress for its linear nature. Conversely, the interface contribution will give rise to higher harmonics in the power spectrum of σ(t) (Grosso and Maffettone, 2007). Consequently, the contribution to the higher harmonics in the Fourier spectrum of each drop with radius *R* is directly related to the interfacial contribution. In the frequency domain this can be written as:

$$\begin{aligned} I_k(R) &= \frac{2\pi}{T}\int_T \sigma(t)e^{-ik\omega_1 t}dt \\ &= -\frac{2\pi}{T}\int_T\left(\frac{\eta}{V}\int_A(\mathbf{nu}+\mathbf{un})dA + \frac{\eta}{V}\int_A\left(\mathbf{nn}-\frac{1}{3}\mathbf{I}\right)dA\right)e^{-ik\omega_1 t}dt \end{aligned} \tag{31}$$

with k odd and > 1.

Simulations are performed by mimicking realistic conditions for PDMS in PIB samples with the relevant parameters reported in the Table 1. These parameter values are consistent with experiments carried out at a temperature T = 35 °C. The volume fraction is always assumed to be $\phi = 0.1$.

Figure 4 shows the tangential stress σ of two simulated polymer blends both in the time and in the Fourier domains. The imposed deformation amplitude and oscillation frequency are set respectively equal to $\gamma_0 = 800\%$ and $\omega = 0.1$ s^{-1}. The two blends differ for drop radius. Figure (4.a) and (4.b) show the time evolution and the Fourier transform (namely the absolute values of power spectrum rescaled with respect the fundamental harmonic, thus $I(\omega_1)=1$) of a polymer blend consisting of equal drops with radius $R_1 = 1$ μm, respectively,

whereas figure (4.c) and (4.d) refer to a blend with drop radius R_2=5 μm. It is apparent that no significant difference can be appreciated in the time domain, the signals looking very close to a sinusoidal waveform in both cases. This result is not unexpected since, as already mentioned, the linear Newtonian matrix contribution dominates the response when observed in the time domain.

On the contrary, the nonlinear features appear more evidently in the Fourier domain. The principal harmonic (corresponding to the forcing frequency ω = 0.1 s^{-1}) is not reported entirely in order to magnify the harmonics appearing at higher frequencies. It is evident that Fourier analysis allows a clear detection of the nonlinearities that are otherwise not appreciable in the time domain. By comparing Fig. (4.b) with Fig. (4.d), it can also be noted the significant dependence of higher harmonics of the shear stress on drop size.

Polymer	Formula	Molecular Weight [Da]	Density [Kg/m^3]	Viscosity [Pa·s]	Interfacial tension [mN/m]
PDMS	$[-Si(CH_3)_2O-]_n$	200000	971	175	3
PIB	$[-CH_2C(CH_3)_2-]_n$	1300	894	57	

Table 1. Main physical properties of PDMS/PIB system

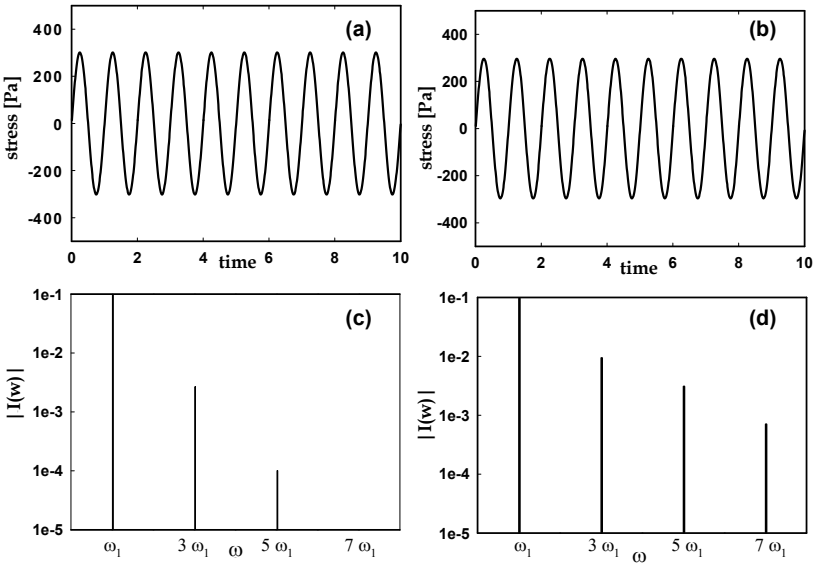


Fig. 4. Tangential stress of a simulated polymer blend in a LAOS experience in the time domain (Figures 4.a and 4.b) and in the Fourier domain (Figures 4.c and 4.d) for two different drop radii: R_1 = 1 μm (Figures 4.a and 4.c) and R_2 = 5 μm (Figures 4.b and 4.d). The deformation amplitude is γ_0=800% and ω_1 = 0.1 ·2π s^{-1}. The physical parameters are in Tab. 1

Figure 5 reports the absolute values I_{3A}/γ_0^3 and I_{5A}/γ_0^5 for a fixed value of the radius R=10 μm versus the strain amplitude γ_0. As the strain deformation tends to zero, both quantities approach a constant value thus confirming the asymptotic behaviour previously observed for other constitutive models (Nam et al. 2008; Ewoldt et al. 2008). It should be remarked that the limiting values depend on the blend properties (i.e. the phase viscosities, the surface tension and the drop radii).

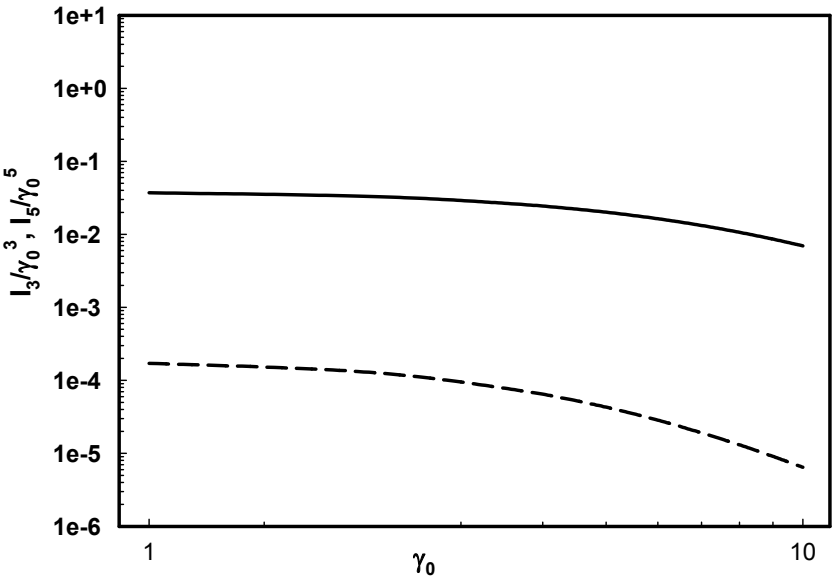


Fig. 5. Scalars I_3/γ_0^3 and I_5/γ_0^5 *vs* the strain deformation γ_0 for a simulated monodisperse polymer blend with radius R = 10 μm

3.2 Experiments

In this section we will show some experimental results that demonstrate the sensitivity of the FTR methodology when analyzing blend or emulsion morphology. The details of the experimental part can be found elsewhere (Carotenuto et al. 2008).

The polymer blend is prepared with PDMS and PIB that are immiscible at room temperature. PIB/PDMS emulsion is a widely used model system largely studied in the literature by means of both rheological and optical techniques (Jansseune et al., 2000; Guido et al. 2004; Wannaborworn et al. 2002). All the experiments were performed at constant temperature T = 30°C. The main physical properties of the polymers are reported in Table 1. The value of the interfacial tension for the very same polymers is found in the literature (Sigillo et al., 1997). PIB is the continuous phase and PDMS is the dispersed phase. All the experiments were carried out with a volumetric fraction, ϕ, of the dispersed phase fixed to 0.1 thus leading to a globular morphology. This value is small enough to consider coalescence negligible. The viscosity ratio is equal to 3, and it is large enough to avoid

significant break-up phenomena under pure shear flow. Thus, the blend can be assumed to be stable, and its microstructure should not significantly vary in time during the experiments (negligible breakup and negligible coalescence). The blend morphology, i. e., the drop size distribution $\psi(R)$, is then assumed to remain unchanged during LAOS experiments.

The experiments were conducted on three different blend samples, which hereafter will be indicated with a capital letter A, B and C. The morphologies of the samples are supposed to have a similar (but not equal) morphology since their preparation followed the same protocol.

Oscillatory shear measurements (both SAOS and LAOS) were performed in a conventional strain controlled rheometer (ARES, TA Instruments). Linear viscoelastic measurements were analyzed using the software provided by the rheometer manufacturer. LAOS experiments required a modification and improvement of the traditional rheometer data acquisition system. The raw data coming from both motor and transducer were acquired and digitized with a 16-bit analog-to-digital converter (National Instrument, PCI_6251). The motor signal was correlated to the imposed strain deformation, γ, while the transducer signal was associated with the measured torque. In order to maximize S/N, the rheometer was equipped with a very sensitive torque transducer (2KFRTN1) that could detect a torque ranging from 0.002 to 200 mN·m.

Before starting the acquisition, two main parameters were set: the scan rate, r [=] pts/s, and the number of data points, N_p. They were the same for both the channels (motor signal and transducer signal). The ratio between N_p and r gives the time required for the entire acquisition, $t_{acq} = N_p/r$. The oscillation cycles collected during t_{acq} depend on the imposed deformation frequency (ω_1). Typical values of r and N_p are 1000 pts/s and 80000 pts, respectively, thus t_{acq} = 80 *s*. Thus, for an imposed deformation frequency ω_1=0.1 Hz, 8 complete cycles were acquired. It should be noted that the higher values of r and N_p, the higher the S/N ratio (Wilhelm et al., 1999). It was however checked that acquisitions with larger amount of data (r = 5,000 pts/*s* and N_p = 400,000 pts) did not show any significant increase in the quality of our data.

Raw data coming from transducer were collected and subsequently transformed into the corresponding Fourier spectra. Odd multiples of the fundamental harmonic appear in the nonlinear regime (LAOS). For the polymer blend under investigation, the 3rd and the 5th overtones could be clearly detected in the shear stress Fourier spectrum for deformation amplitudes $\gamma 0 > 100\%$. The electric signal measured by the torque transducer is supplied in terms of potential difference units.

LAOS data were analyzed according to the FTR protocols. The imposed sinusoidal deformation is $\gamma(t)= \gamma_0 \sin(\omega_1 t)$, where $\omega_1 = 2\pi\Omega_1=2\pi/T$ is the characteristic angular frequency with T the oscillation period.

Linear viscoelastic measurements were carried out for a preliminary characterization of the microstructure of the samples. Oscillatory measurements were performed with frequency ranging from 0.1 to 10 Hz. Strain amplitudes up to 50% gave shear stress responses well within the linear region. It was found that traditional SAOS measurements did not give a clear discrimination between different blends, and the G' curves for the three emulsions are almost overlapping, thus indicating that SAOS suggest that the three blends have similar morphologies.

Blend	$\overline{R}_V$ [μm]
A	5.6
B	6.7
C	8.0

Table 2. Average drop radii for the three blends estimated with the Palierne method

From SAOS measurements one can obtain an estimation of the average dimension of the dispersed phase, namely the volume-average drop radius, $\overline{R}_V$. According to Palierne (1990), one can estimate the volume average drop radius for the emulsions. Table 2 contains the values of the estimated average drop radii for the blends A, B, C. As expected, the volume averaged drop radius, $\overline{R}_V$, for the three samples is very similar. A more detailed description of blend morphology cannot be attained with the linear rheological measurements.
LAOS measurements were performed with $\gamma_0 \geq 200\%$, where nonlinearities in the response become clearly appreciable. A typical experimental result is shown in Fig. 6, where the tangential stress response is reported both in the time (Fig. 6.a) and in the frequency domain (Fig. 6.b) for γ_0 = 800% and Ω_1 = 0.1 Hz (or, equivalently, $\omega_1=2\pi$ 0.1 rad/sec). Fourier spectra report the absolute value of the overtones, normalized with the first harmonic (I_k/I_1 or equivalently I_{k1}) as commonly done in the FTR literature (e.g. Wilhelm et al., 1998). The nonlinear shear stress response cannot be easily detected in the time domain, but the corresponding power spectrum clearly shows the occurrence of a third and a fifth peak.

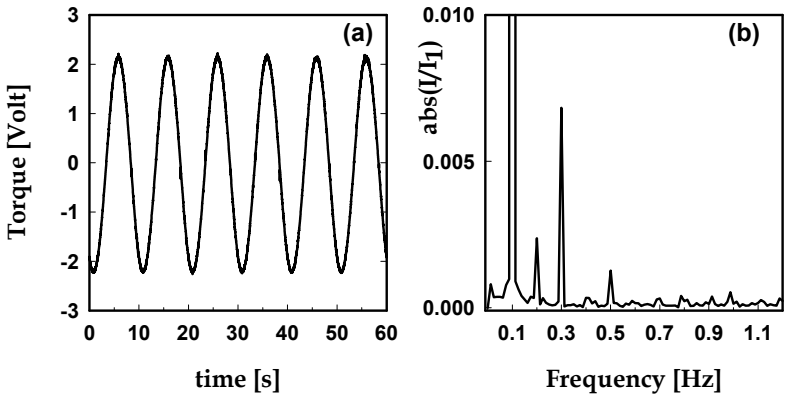


Fig. 6. Transducer signals in the time domain (a) and the corresponding Fourier spectra (b) for a polymer blend, at 30°C

In figure 6.b a peak $I(2\omega_1)$ at an even multiple of the fundamental harmonic is also observable. It should be reminded that this occurrence is unexpected since the stress signal is demonstrated to be an odd function of the time. Several explanations for the presence of even overtones in the spectra have been proposed in literature. Quite often, the occurrence

of even harmonics is attributed to some artefact in the experiments as e.g. wall-slip phenomena (Hatzikiriakos & Dealy, 1991). In the case here reported, the second harmonic seems to be material-independent, and it can be attributed to an imperfect alignment of the upper and lower plates of the rheometer (Carotenuto et al., 2008): it reasonably comes from the instrument itself and results unrelated to the measured sample, for this reason it is simply neglected.

Figure 7 shows the Q coefficient defined in equation 24 as a function of the strain amplitude for the blend C. The third harmonic is clearly detected for the polymer blend under investigation. The value of I_{31} is small but reproducible with an experimental error lower than 3%. For the sake of comparison, data of the neat PIB and PDMS are also reported in Fig. 7. The pure component I_{31} is weighted by the corresponding amount in the blend (i.e., 0.1 for the PDMS and 0.9 for the PIB). It is apparent that the I_{31} values of the pure components are extremely low, according to their quasi-Newtonian behavior, and negligible when compared with the I_{31} values of the blend. This experimental evidence unequivocally suggests that the observed nonlinear response of the blend does not derive from simple superposition of the nonlinear contribution of the neat polymers, but it seems essentially due to the interface stress contribution. Such behaviour confirms the validity of the assumptions made in Equation 29.

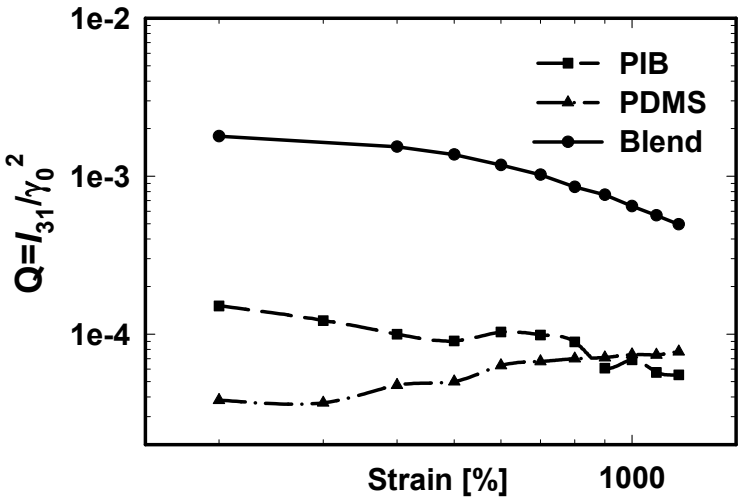


Fig. 7. The Q coefficient as a function of the imposed strain deformation for the pure components (PIB: dashed line with square; PDMS: dashed-dotted line with triangles) and for the blend C (solid line with circles). The temperature is 30°C. The oscillation frequency is 0.1 Hz

Figure 8.a shows the coefficients $Q = I_{31}/\gamma_0^2$ and $P = I_{51}/\gamma_0^4$ for the three blends A, B and C as a function of the strain amplitude. It is shown that, as the strain deformation decreases, the curves seem to tend to an asymptotic plateau value Q_0 and P_0, accordingly with the theoretical predictions.

The curves do not superimpose, thus suggesting that LAOS experiences could discriminate between different morphologies. Indeed, as reported from Palierne results, the upper curve refers to the blend A ($\overline{R}_V$ = 5.6 μm), the medium to the blend B ($\overline{R}_V$ = 6.7 μm), and the lower to the blend C ($\overline{R}_V$ =8 μm). Hence, the Q_0 coefficient of the blend seems to decrease with the mean size of the inclusions. In Figure 8.b, the ratio $P = I_{51}/\gamma_0^4$ for the blends A, B and C are also reported. Since the fifth overtones are significantly smaller than the third ones, they are more affected by experimental noise. Analogously to Q behaviour, the fifth peaks are larger for blend with smaller volume averaged drop radius.

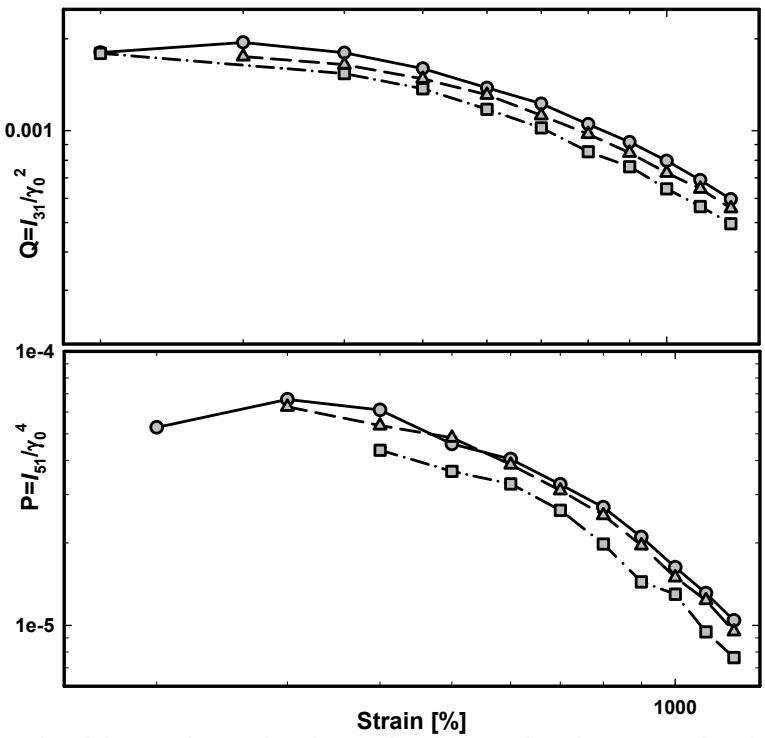


Fig. 8. The Q and P coefficients as a function of strain amplitude for the blend A (solid line with circles), B (dashed line with triangles) and C (dashed-dotted line with squares). The oscillation frequency is 0.1 Hz

4. Conclusion

Fourier Transform Rheology is a valuable tool to characterize the microstructure of dilute immiscible polymer blend as it was shown both theoretically and experimentally. We analyze the case of a blend with Newtonian constituents, and in such a case the nonlinearity of the response under LAOS comes exclusively from the presence of a polymer-polymer

interface. Indeed, distinct odd multiples of the fundamental harmonic are clearly evident in the power spectrum of the emulsion, while are only barely distinguishable in the spectra of the pure components (PIB and PDMS). FTR greatly enhances the sensitivity of the experiments to the blend morphology, thus allowing the evaluation of details that are otherwise difficult to be appreciated with time domain analysis.

5. Acknowledgments

Authors acknowledge Dr. Claudia Carotenuto for the helpful discussions on the experimental technique.

6. References

Almusallam A.S., Larson R. & Solomon M.J. (2004). Comprehensive constitutive model for immiscible blends of Newtonian polymers, *Journal of Rheology*, 48(2), pp 319-348 ISSN: 0148-6055

Batchelor G.K. (1970) The stress system in a suspension of force-free particles, *J. Fluid Mech.* 41, pp 545-570 ISSN: 0022-1120

Bracewell R.N. (1986) The Fourier-transform and its application. McGraw-Hill, New York, ISBN: 978-0073039381

Carotenuto C., Grosso M., Maffettone P.L. (2008) Fourier Transform Rheology of Dilute Immiscible Polymer Blends: A Novel Procedure To Probe Blend Morphology, *Macromolecules*, 41, pp 4492-4500 ISSN: 0024-9297

Das N.C., Wang H, Mewis J. & Moldenaers P. (2005), Rheology and Microstructures Formation of Immiscible Model Polymer Blends Under Steady State and Transient Flows, *J. Pol. Sci.: Part B: Polymer Physics*, 43, pp. 3519–3533 ISSN: 1099-0488

Ewoldt RH, Hosoi AE & McKinley GH (2008) New measures for characterizing nonlinear viscoelasticity in large amplitude oscillatory shear. *Journal of Rheology* 52(6), pp. 1427–1458 ISSN: 0148-6055

Friedrich C., Gleinser W., Korat E. Maier D. & Weese J. (1995). Comparison of sphere-size distributions obtained from rheology and transmission electron microscopy in PMMA/PS blend, *Journal of Rheology*, 39(6) pp. 1411-1425 ISSN: 0148-6055

Graebling D,, Muller R. & Palierne J.P. (1993). Linear viscoelastic behavior of some incompatible polymer blends in the melt - interpretation of data with a model of emulsion of viscoelastic liquids *Macromolecules* 26, pp. 320–329 ISSN: 0024-9297

Graebling D,, Muller R. & Palierne J.P. (1993). Linear viscoelasticity of incompatible polymer blends in the melt in relation with interfacial properties, *Journal de Physique IV*, 3 pp. 1525–1534, ISSN: 1155-4339

Grosso M. & Maffettone P.L. (2007) A new methodology for the estimation of drop size distributions of dilute polymer blends based on LAOS flows, *J. Non Newt. Fluid Mech.* 143, pp. 48-58 ISSN: 0377-0257

Guido S.; Grosso M.; Maffettone P.L. (2004). Newtonian drop in a Newtonian matrix subjected to large amplitude oscillatory shear flows, *Rheologica Acta*, 43, pp. 575-583 ISSN: 0035-4511

Hatzikiriakos, S. G., Dealy J. M. (1991) Wall slip of molten high density polyethylene. I. Sliding plate rheometer studies. *Journal of Rheology* 35(4), 497-523 ISSN: 0148-6055

Huitric J., Mederic P., Moan M., & Jarrin J. (1998) Infuence of Composition and morphology on rheological properties of polyethylene/polyamide blends, *Polymer* 39(20) pp. 4849-4856 ISSN: 0032-3861

Hyun K, Wilhelm M (2009) Establishing a New Mechanical Nonlinear Coefficient Q from FT-Rheology: First Investigation of Entangled Linear and Comb Polymer Model Systems. *Macromolecules* 42 pp. 411–422 ISSN: 0024-9297

Jansseune T.; Mewis J.; Moldenaers P.; Minale M. & Maffettone P.L. (2000) Rheology and rheological morphology determination in immiscible two-phase polymer blends, *J. Non-Newtonian Fluid Mech.* 93, pp. 153-165 ISSN: 0377-0257

Lacroix C., Bousmina M., Carreau P.J., Favis B.D. & Michel A.(1996) Properties of PETG/EVA blends .1. Viscoelastic, morphological and interfacial properties *Polymer* 37 (1996), pp. 2939 ISSN: 0032-3861

Macosko W. (1994) Rheology: Principles, Measurements and Applications, Wiley-VCH, New-York, ISBN: 978-0471185758

Maffettone P.L. and Minale M. (1998) Equation of change for ellipsoidal drops in viscous flows, *J. Non-Newtonian Fluid. Mech.*, 78, pp. 227-241 ISSN: 0377-0257

Minale M. (2010) Models for the deformation of a single ellipsoidal drop: a review, *Rheol. Acta*, 49(8), pp. 789-806, ISSN: 0035-4511

Nam JG, Hyun K, Ahn KH, Lee SJ (2008) Prediction of normal stresses under large amplitude oscillatory shear flow. *J Non-Newtonian Fluid Mech* 150:1–10 ISSN: 0377-0257

Neidhöfer T., Siuola S., Hadjichristidis N., Wilhelm M. (2004) Distinguish Linear from Star Branched Polystyrene Solutions with Fourier Transform Rheology, *Macromol. Rapid Comm.*, 25, pp. 1921-1926 ISSN: 1022-1336

Palierne J.F. (1990) Linear rheology of viscoelastic emulsions with interfacial tension, *Rheol. Acta* 29 pp. 204-214 ISSN: 0035-4511

Rallison J.M. (1984) The Deformation of Small Viscous Drops and Bubbles in Shear Flows, *Ann. Rev. Fluid Mech.* 16, pp. 45-66 ISSN: 0066-4189

Ramirez RW (1985) The FFT Fundamentals and Concepts. Prentice-Hall, Englewood Cliffs, NJ, ISBN: 978-0133143867

Schlatter G., Fleury G., Muller R. (2005) Fourier Transform Rheology of Branched Polyethylene: Experiments and Models for Assessing the Macromolecular Architecture, *Macromolecules*, 38, pp. 6492-6503 ISSN: 0024-9297

Sigillo I.; Di Santo L.; Guido S.; Grizzuti N. (1997) Comparative Measurements of Interfacial Tension in a Model Polymer Blend, *Polymer Eng. and Sci.* 37(9),pp. 1540-1549 ISSN: 0032-3888

H.A. Stone, Dynamics of Drop Deformation and Breakup in Viscous Fluids (1994) *Ann. Rev. Fluid Mech.* 26, pp. 65-102 ISSN: 0066-4189

Tucker C.L. and Moldenaers P. (2002) Microstructural Evolution in Polymer Blends, *Ann. Rev. Fluid. Mech.* 34, pp. 177-210 ISSN: 0066-4189

Utracki L.A. (2003) Polymer Blends Handbook, Kluwer Academic Publishers, Dordrecht-Boston-London ISBN: 978-1402011146

Vincze-Minya K.A., Schausberger A. (2007) Characterization of Phase Morphology of Polymer Melts(PP/PE Blends) Via Rheology, *Journal of Applied Polymer Science*, 105, pp. 2294–2298, ISSN: 0021-8995

Wannaborworn S; Mackley M.R; Renardy Y. (2002) Experimental observation and matching numerical simulation for the deformation and breakup of immiscible drops in oscillatory shear. *Journal of Rheology* 46(5), pp. 1279-1293 ISSN: 0148-6055

Wilhelm M, Maring D, Spiess HW (1998) Fourier-transform Rheology, *Rheol. Acta*, 37, pp. 399-405 ISSN: 0035-4511

Wilhelm M., Reinheimer P., Ortseifer M (1999) High sensitivity Fourier Transform Rheology. *Rheol. Acta* 38 pp. 349-356, ISSN: 0035-4511

Yu W, Wang P, Zhou C (2008) General stress decomposition in nonlinear oscillatory shear flow. *Journal of Rheology* 53(1):215–238 ISSN: 0148-6055

15

Application of Fast Fourier Transforms in Some Advanced Electroanalytical Methods

Parviz Norouzi[1], Morteza Pirali-Hamedani[2,3], Tayebeh Mirzaei Garakani[1] and Mohammad Reza Ganjali[1]

[1]*Center of Excellence in Electrochemistry, Faculty of Chemistry, University of Tehran, Tehran,*

[2]*Department of Medical Chemistry, Faculty of Pharmacy, Tehran University of Medical Sciences, Tehran,*

[3]*Pharmaceutical Sciences Research Center, Tehran, Iran*

1. Introduction

In the most of electrochemical (EC) experiments, measurements mostly are performed in the time domain. However, in some cases, we require more information for the obtained data such as knowledge about the frequency content and behavior of the electroanalytical signals and of complete systems. Fortunately, there exists a defined method for transforming data from the time domain into the frequency domain, where information exist about the spectral content of EC data. The method for this propos is Fourier Transform (FT), which has the ability to convert a time domain data to the complex frequency domain, meaning the spectral data contains information about both the amplitude and phase of the sinusoidal components that make up the signal. In addition, the inverse FT, converts the generated complex frequency-domain signal data back into the time-domain without losing wanted information. Accordingly, it can say that both the time- and frequency-domain data complement and the two domains can provide a different view of the same EC data.

Application of fast Fourier transformation (FFT) algorithm for numerical EC data provides the complex spectrum according to magnitude and phase, which can be used for real time analysis. In this direction, in modern electrochemistry, FFT has been used for digital signal processing and filtering. Also, the FFT process returns a vector of real and imaginary elements, which represent the various resolved harmonics in impedance spectroscopy, AC and square wave voltammetry (Popkirov, 1996).

On the other hand, it must be noted that in all EC data collection, to hold on the sampling theorem for FFT, the bandwidth of the input signal is limited by an analog low pass filter (cutoff frequency $f_c = f_{in,max}$) ahead of the Analog to Digital (A/D) converter. In fact, after collecting data in the computer memory, they are used for calculating the signal in the frequency domain.

This chapter serves as summary application of the FFT analysis techniques implemented in EC measurement platform. By reading through this document, you will receive a comprehension of the fundamental concepts in FFT-based measurements used throughout EC application, providing you insights to better understanding of the measurement

parameters, procedures, and resulting EC data. In addition, this chapter describes the general operation of the FFT analysis accompanied with modern EC methods.

2. Basic FFT theory

The FT, a pervasive and adaptable tool, is used in many fields of science as a mathematical technique to alter a problem into one that can more easily be solved. Scientists consider FT theory as a physical phenomenon, not simply as a mathematical tool. Based on the Fourier theory, any signal in periodic manner in the time domain can be derived from the sum of sine and cosine signals of different frequencies and amplitudes, which is called as a Fourier series (Weaver, 1983). Thereby, it is notable to calculate the frequency spectrum of a periodic signal according to Equation 1,

$$x(t) = \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \cdot \sin(n \cdot \omega_0 \cdot t) + \sum_{n=1}^{\infty} B_n \cdot \cos(n \cdot \omega_0 \cdot t) \quad (1)$$

where, *x(t)* is data in time domain, A_n and B_n are the amplitude, w_o is the frequency of the waveform and n is the harmonic number. Each of these elements leads to a discrete component in the frequency domain, and periodic signals exhibit discrete line spectra. However, signals with a non-periodic characteristic in the time domain cannot be described by FT, and those signals exhibit a continuous frequency spectrum with a frequency-dependency. Therefore, the frequency spectrum of such signals is not composed of discrete spectral components. The signal in the frequency domain is calculated by means of a FT (Equation 2).

$$X_f(f) = F(x(t)) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt \quad (2)$$

Also, for measuring the harmonic section of the EC data, it is more useful to examine the signal in the frequency domain (Rauscher et al., 2001). It has been shown that the signal in the frequency domain of the fundamental (1st order harmonic) is superimposed by several higher-order harmonics with the aid of a spectrum analyzer. As a matter of fact, this information cannot be simply obtained by examining the signal in the time domain. Practically, the higher order harmonics are not possible, and limited number of the data samples can be used for FFT calculation.

3. Fundamentals of electroanalytical signals

It is well known that many fundamental microscopic processes take place on the electrode surface, which can lead to the overall electrical signals. They may include the transport of electrons through the electronic conductors, the transfer of electrons at the electrode/electrolyte interfaces to form species which are originated from the cell materials, and also, the stream of charged atoms. Indeed, the current depends on the ohmic resistance of the electrodes and the electrolyte and also on the process rates at the electrode/electrolyte interfaces.

In practical point of view, there are three different types of electrical signals in EC measurements. Each basic electrical measurement of current (*i*), resistance (*R*), and potential (*V*) has been used alone or in combination for analytical measurements (Brett & Brett, 1993).

First, in transient measurements a waveform function of potential may be applied at electrode surface and then the resulting time varying current measured. The ratio voltage to current often called the time varying resistance, measures the impedance resulting from the voltage perturbation at the electrode/solution interfaces. If a FFT is used, a distortion arises because of the non-periodicity of excitation. Such transformation is only valid when the applied potential waveform is sufficiently small so that system response becomes linear .
The second type contains signals containing random noise, and measure the resulting current and voltage, and application of FFT to the results to obtain the frequency domain data. This process can be used in electrochemical noise analysis (ENA) method for determination of corrosion. This approach offers the advantage of fast data collection because only one signal is applied to the interface for a short time .
The last type, the most common and standard one, is to measure EC data by applying a single-frequency voltage or current to the electrode interface and measuring the phase shift and amplitude, which leads to measure the real and imaginary parts of the resulting current at a certain frequency. The most important advantage of such FFT analysis is in AC and Square Wave Voltammetry (SWV), in which combines the first and third techniques.

4. Application of FFT in electroanalytical methods

As mentioned above, most electrical signals in EC measurements may be examined in the time domain with the aid of a potentiostat and in the frequency domain with the aid of the computer digital spectrum analyzer. The two display modes are related to each other, where each signal variable in the time domain has a frequency spectrum characteristic (Gavaghan & Bond, 2000). This calculation would be obtained in a continuous data collection, so the frequency resolution would be unlimited. Noticeably such exact calculations are not possible in practice, where by given certain prerequisites, the spectrum can be determined with sufficient accuracy. In practice, the FT data is constructed with the aid of digital signal processing, as a result, the signal to be analyzed has to be sampled by an A/D converter and quantized in amplitude.
Most of the generated electrical signals in electroanalytical methods are continuous, in which for every time value there is a defined data. However, in order these continuous signals to be analyzed, it is needed a computer-based measurement system employed, such as Labveiw interfacing program or other interfacing system. Those systems can convert the EC electrical signal into a stream of digital data, which each data represents a numeric value that is proportional to the measured data at a specific time. This process is known as data sampling: converting the analog signals into a discrete-time signal (a process handled by A/D converter) to allow electroanalytical data in a wide level range to be simultaneously processed by FFT program and to be displayed on the computer screen (Norouzi et al., 2003).
In the computing process, the FFT method operates by decomposing N data point in time domain signal into N frequency domain signals each composed of a single point of the electroanalytical measurement. The next step is to calculate the N frequency spectra corresponding to these N time domain signals. Finally, the N spectra are synthesized into a single frequency spectrum. If N is an integer power of 2, thus $N= 2^p$ (p =1, 2, 3, ….), use the fastest FT function that uses the decimation-in-time algorithm. Given N samples of a periodic function $f(t)$ by a normalized sampling rate (T=1), { f(0), f(1), f(2),....f(N-1) }, the discrete Fourier transform (DFT) is defined by:

$$F(k) = F_r(k) + i \cdot F_i(k) = \sum_{n=0}^{N-1} f(n) \cdot \left[\cos(2\pi nk / N) - i \cdot \sin(2\pi nk / N) \right] \quad (3)$$

Normally, at first, an electrode response was recorded and then, DFT was applied on the collected data and the existing frequencies, phase angle, real and imaginary parts are calculated. Based on this calculation, the modern methods are established, such as ENA, FFT Cyclic voltammetry, FFT SW voltammetry and FFT impedance spectroscopy.

4.1 Application of FFT electrochemical measurements based on noise analysis

The frequencies of unwanted signals in EC data, with noise or random characteristics, cannot be found easily (Aballe et al., 1999; Sang et al., 2009; Dai, 2000). The analyst requires a plot of the intensity at each individual frequency in order to make identification and the EC signal to be interpreted. A means of "decoding" for calculating the individual frequencies is needed. This procedure can be accomplished via a well-known mathematical technique such as DFT. This transformation is performed by the computer based programs, which then presents the user the desired spectral information. When the frequency region of the noise is found, it can be used for two aims in the EC analysis; data filtering for enhancing signal-to-noise (S/N), and corrosion monitoring (Darowicki & Zieliski, 2001; Safizadeh& Ghali, 2010).

4.2 Application of FFT in cyclic potential sweep voltammetry based on filtering

In modern EC methods, the FFT analysis voltammetry were developed in order to overcome some existing limitations encountered with electronic instrumentations. In fact, the selectivity and percision characteristics of the classic electroanalytical methods depend on the number of filter circuits. Normally, the most potentiostat typically has many filter circuits that are arranged before and after the amplifier. The main problem here is to apply a fast excitation signal for fast EC measurement. Actually, application a fast excitation signal can produce a large charging current due to existing capacitor in the analog filters.

On the other hand, by using FFT filtering method, instead of the analog filters, the signal can be measured very quickly. Consequently, the time element per sample is reduced to a matter of less than second (10^{-9} s) rather than several minutes which happened in the classical analog measurements. As cyclic voltammetry finds its greatest use in the study of fast electrochemical process, and transient intermediates, we are supposed to choose this method for its description and the FFT application. Through the use of analog filters, the maximum permissible sweep speed is limited by the transient time of the applied filter. The maximum span that can be analyzed at a specific resolution by means of FFT is limited by the sampling rate of the A/D converter and by the memory available for saving the sampled data. In fact, to allow shorter sweep times, FFT digital filters are advantageous for narrow resolution bandwidths. Basically, the EC digital analyzer is designed for bandwidths from 100 Hz to 10 MHz. This digital filtration provides condition that can be used for very high potential scan rates on working electrode for electrochemical measurement in flowing solution.

Figure 1 shows an example for FFT filtration for fast cyclic voltammetric measurement (Norouzi et al. 2007). In this method at the beginning, a CV of the electrode was recorded (see Fig. 1a) and then by applying FFT on the collected data, the existing high frequency noises were indicated (see Fig. 1b). Finally, by using this information, the cutoff frequency of

the analog filter was set at a certain value (where the noises were removed from the CV). The resulted CV in Fig. 1c shows successfulness of the filtering procedure. This kind of filtration and also current integration significantly reduces the noise level in the obtained data.

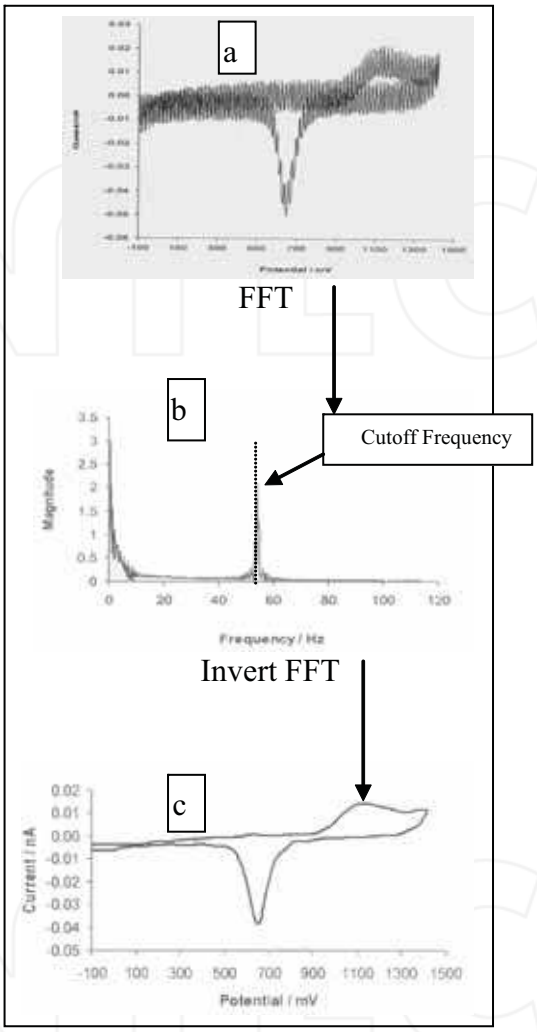


Fig. 1. Application of FFT filtration to smooth a noisy CV, a) original CV, b) FFT spectrum of the CV (the inset shows the cutoff frequency that is selected for filtration), c) the resulted CV after removing the noise frequencies

The EC measurement based on the FFTdigital filters is widely used by Norouzi group for determination of several organic compounds and drugs such as; Diphenhydramine (Norouzi et al. 2010 b), Lidocaine (Norouzi at al. 2007), Methyldopa (Norouzi et al. 2009), and Salbutamol(Ganjali et al. 2005). The EC method under reported condition was named FFT Continuous Cyclic Voltammetry (FFTCCV). In this method the potential waveform was

continuously applied during an experiment run, where the collected data was digitally filtered by the technique, before using them in calculation for the analyte response (Norouzi et al., 2010).

It must be noted that, this digital analyzing, also, offers the possibility of the frequency-dependent gain. Recorded spectra can automatically be displayed with the correct levels. In these cases a reduction of the displayed noise by decreasing the resolution bandwidth is not permitted. Due to this fact that, the sensitivity is also important for the fast measurement speeds the program of the FFT digital analyzer featuring a low noise figure leads to the use of greater resolution bandwidths, and also with manual setting of the resolution and bandwidths, the sweep time can be adapted automatically.

An example of application of that waveform is shown in Figure 2. This figure shows a sequence of CVs recorded during the flow injection of 50 μL of 1.0×10^{-6} M Cl^- (in 0.05 M H_3PO_4) into the eluent solution containing 0.05 M H_3PO_4. The potential axis on this graph represents potential applied to the working electrode during each sweep. The time axis represents the time passing between the beginning of the flow injection experiment and the beginning of a particular sweep (i.e. it represents a quantity proportional to the sweep number). The characteristic element of CVs at gold electrode is a set of peaks associated with the formation and dissolution of a surface oxide layer at about 1600 and 400 mV (when potential sweep rate is 20 Vs^{-1}), respectively. The process is also initiated by the electrosorption of the hydroxyl ion, which at more positive potentials undergoes deprotonation and structural rearrangement. The surface oxidation can be initiated by adsorption of water molecule and then at more positive potential AuOH forms leading to the formation of a two-dimensional phase of gold oxide;

$$Au\left(H_2O\right) \rightarrow AuOH + e + H^+$$

At more positive potentials we have AuO according to the following reaction;

$$AuOH \rightarrow AuO + e + H^+$$

Figure 2b shows the absolute current changes in the CVs curves after subtracting the average background of 10 CVs (in the absence of the analyte). As can be seen, this way of presentation of the electrode response gives more details about the effect of adsorbed ion on currents of the CV. The curves show that current changes mainly take place at the potential regions of the oxidation and reduction of gold. When the electrode-solution interface is exposed to Cl^-, which can adsorb on the electrode surface, the oxide formation process is inhibited.

All CV curves observed during the entire experiment (typically 100 to 1000 curves) were always stored in computer memory and could be saved onto a hard drive for future analysis (See Fig. 2a). The important point in FFTCCV method is that parts of the adsorbed analyte still remain on the electrode surface that can inhibit the oxidation process of the electrode surface. In this method, ΔQ is calculated based on the all-current changes at the CV (See Fig. 2c). A total absolute difference function can be calculated by using the following equation:

$$\Delta \boldsymbol{Q}_{TA}\left(s\tau\right) = \Delta t\left[\sum_{E=E_i}^{E=E_v}\left|i\left(s,\boldsymbol{E}\right) - i\left(s_r,\boldsymbol{E}\right)\right| + \sum_{E=E_v}^{E=E_i}\left|i\left(s,\boldsymbol{E}\right) - i\left(s_r,\boldsymbol{E}\right)\right|\right] \quad (4)$$

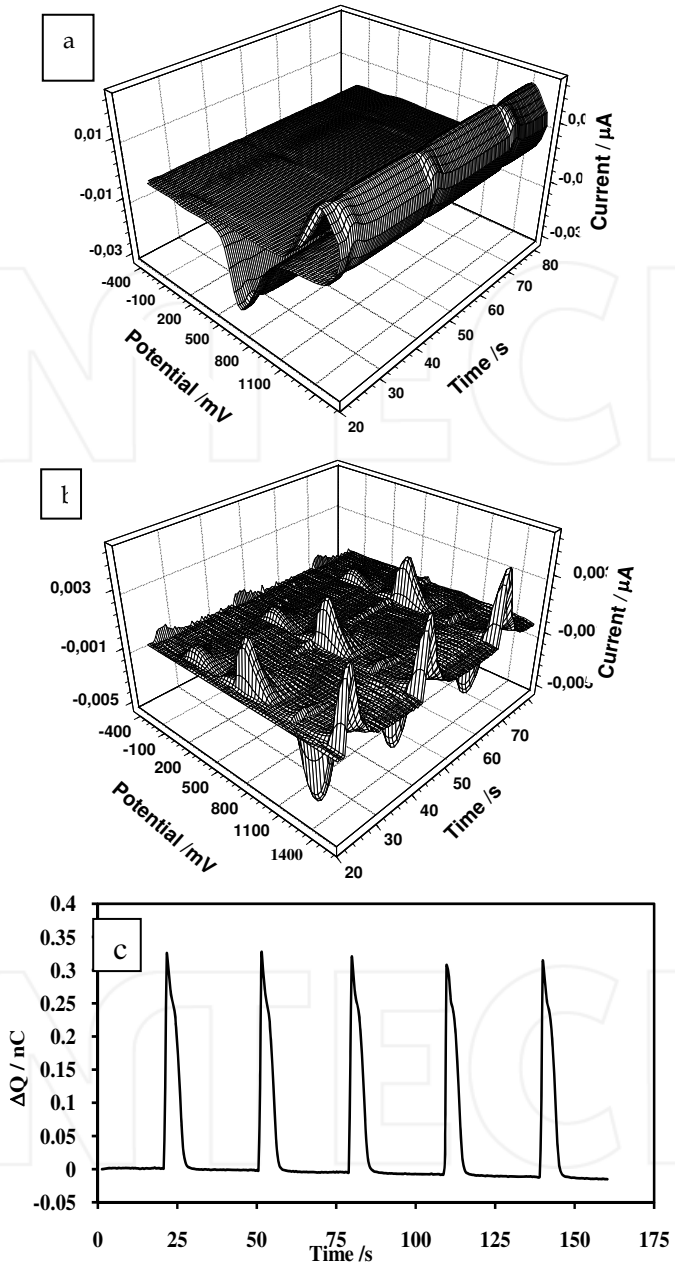


Fig. 2. a) Cyclic voltammograms at Au ultra-microelectrode recorded during the flow injection. The eluent was 0.05 M H_3PO_4 with the flow rate of 0.5mL/min., b) Curves resulted from subtracting the CVs in fig. a, from the average of 10 CVs (in the absence of analyte), c) Response of Au ultramicroelectrode to 5 consecutive injections of analyte

Where, s is the sweep number, τ is the time period between subsequent sweeps, Δt is the time difference between two subsequent points on the CV curves, i (s, E) represents the current of the CV curve recorded during the s-th sweep and i (s_r, E) is the reference current of the CV curve. E_i and E_υ are the initial and the vertex potential, respectively. The reference CV curve was obtained by averaging a few CV curves recorded at the beginning of the experiment (i.e. before injection of the analyte). These equations show that for the same flow injection experiment the analyte response can be obtained using different integration limits.

It should be noted that in this method, all studied processes involve adsorption of analyte; hence both charging and faradic currents may potentially carry useful analytical information. To get such information, it was important to sample current at a frequency at least two times higher than the current transducer bandwidth. In order to fulfill this requirement the sampling frequency was always adjusted at 100 kHz. In addition FFT digital low pass filters with 0.5-30 kHz cutoff frequencies lead to remove noise from the data. If the main contribution to the baseline noise is from the "white" noise generated by the potentiostat, the integration procedure usually provides a 3 to 60 fold improvement in S/N compared to the simple monitoring of the current at a fixed potential. However, in the case of severe environmental noise (e.g. power line noise) the improvement may be much larger. Signals with weak level are thus shown more distinctly in the voltammogram and the measured level values are thereby stabilized and reproducible. In the case of a sinusoidal signal, the displayed level is not influenced by a reduction of the bandwidth. To obtain stable and reproducible results of noise measurements, a narrow bandwidth should be selected. The noise bandwidth is thus reduced and high noise peaks are better to be averaged.

5. Application of FFT in electrochemical noise analysis

As mentioned above the identification of noise frequencies in the electrochemical data by FFT method, can be used for study of some processes that occur on the electrode surface. This type of noises, which are created by EC process, called electrochemical noise (EN). For many years, EN has been observed during corrosion and other electrochemical reactions, and the phenomenon is well established.

The theoretical behavior of EN is not, completely, discussed, but there have been many useful applications, both in scientific surface and in corrosion monitoring studies. In an EC system, noise of potential and of current can be made independently or together during corrosion reaction that can be take placed on the electrode surface. The important point is that EN measurement technique does not involve any external perturbation of the corroding system. In fact, the instruments required to perform EN measurements are reasonably simple, particularly with FFT digital analyzer and modern computer-based data acquisition techniques. More commonly, the EN curve for potential and current has an appearance similar to that shown in Figure 3.

Generally, localized corrosion processes tend to give particularly strong EN signals, which consist of low frequency (< 1 Hz) and small amplitude signals. Those signals are spontaneously generated by electrochemical reactions occurring at corroding or other surfaces (Zaveri et al., 2007). However, In this process, both the noise of potential and current are of the $1/f$ type. The range of frequencies depends on the sampling interval Δt (typically 0.5 s) and on the number of readings M of a time record. The maximum and

minimum frequencies that can be analyzed are: $f_{max} = 1/2\Delta t$ *and* $f_{min} = 1/Mt$). The limiting value at a frequency approaching zero is defined as the spectral noise resistance.

In practice, spectral noise plots can be obtained only in a frequency range that is more limited than in EIS. On the high frequency side, the limit is imposed by the instrumental noise, whereas in the low frequency region, the time of acquisition becomes very long (Lafront et al., 2010).

In a different approach, the time record of potential or current is converted into a power spectral density (PSD), which is the distribution of the power in the frequency domain. This transformation is usually made by means of the FFT method. Figure 4 shows the calculated power spectral density (PSD $= 1/f^n$, where n is a constant). The figure shows that PSD falls with increasing frequency, giving a straight line on a log-log plot, implying a relationship. In addition to the sloping $1/f^n$ region of the spectrum, there are often indications of plateaus at the high and/or low frequency ends of the spectrum. Potential noise frequency relation changes have been noticed in the form of a component of character $1/f^2$, this being attributed to pit initiation.

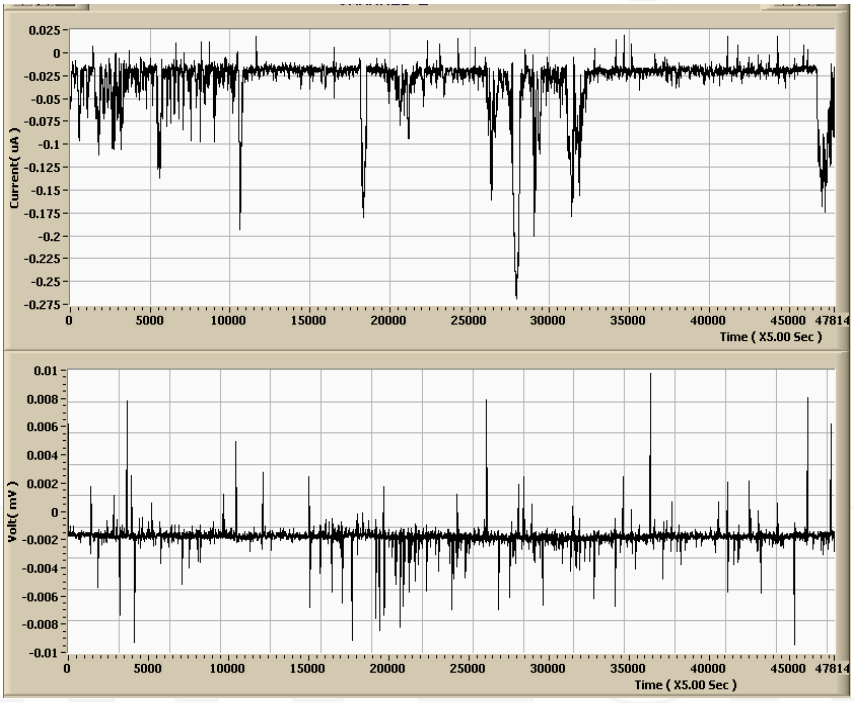


Fig. 3. The sample graph for EN measurements of an Al palate in NaCl 0.01 M, top) Current noise, bottom) Potential noise

The character of PSD changes as a function of frequency depend on the type of corrosion and has also been the subject of investigations. The area under the curve is the total power in the signal, and is identical to the standard deviation calculated from the time record. Thus, as the frequency spectrum moves to higher PSDs, so the rate of reaction may be expected to increase .The details of the calculation are beyond the scope of this book.

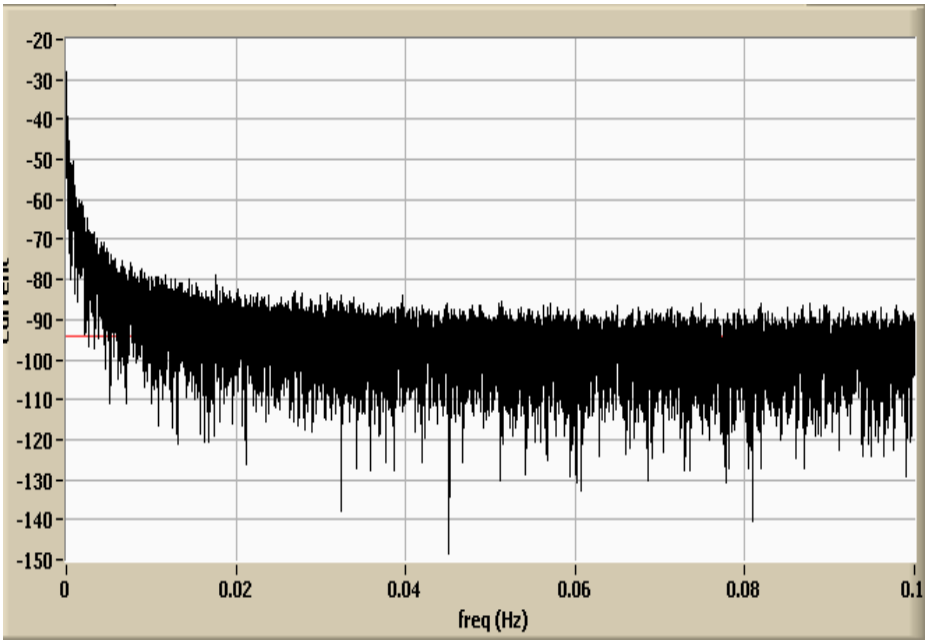


Fig. 4. The sample of PSD graph for EN measurements of an Al plate in NaCl 0.01 M

6. Application of FFT in AC voltammetry

The determination of the characteristics of the EC data system by AC techniques requires measuring the impedance at various frequencies, to result a frequency spectrum (Bond et al., 1997). The most dominant application of FFT methods in AC voltammetry is in calculation of the impedance of the electrode response, which is known as electrochemical impedance spectroscopy (EIS).

By definition, AC voltammetry (ACV) utilizes a small-amplitude sine wave which is added to a potential ramp to modulate the current output. In fact, ACV is an extension of classical linear sweep techniques such as cyclic voltammetry. A DC ramp with a comparatively slow sweep rate and an AC signal are superimposed and applied to a working electrode, and the response of the AC current and its phase angle are registered. In general, modulation potential amplitude up to 20 mV is used; higher amplitudes are not used to specifically avoid contributions from higher order harmonics. This may be considered as a limiting case for ACV. In fact, all of the information is intentionally contained in only the lowest order harmonics. Its main strength lies in the quantitative characterization of electrode processes, and it can also be used for analytical purposes.

A second general approach has employed AC voltammetry and the mathematics of the FFT. This technique was introduced and used extensively by (Smith, 1976). In it, the perturbation signal is composed of a sum of selected sinusoids. This approach, although powerful, is quite expensive and complex as well, and has not been widely employed. Both experimental implementation and data analysis employ a small-amplitude excitation waveform.

The applied excitation waveform consists of a fundamental harmonic frequency f_0 and a number of odd harmonics $(2n + 1)\ f_0$. This arrangement is superior to other excitation waveforms. All these frequencies are applied at the same time and the data to each frequency is found by the FFT. Depending on the method used for the data acquisition, different methods are used for data validation. In case of a sinusoidal signal were applied to a nonlinear system, the data function would contain multiples harmonics of the excitation waveforms.

The techniques commonly employed for AC-impedance measurements in modern equipment can be subdivided into two main groups: single-sine and multiple-sine methods. The lock-in technique and frequency-response analysis will be described as representatives of the single sine techniques and FFTs will be introduced as an example for multiple-sine techniques. In single-sine methods, a small-amplitude sinusoidal signal with a fixed frequency is applied to the test cell. The response signal is then analyzed to extract the two components of the impedance (real and imaginary parts or magnitude and phase).

Here, invalid impedance data cannot only be caused by nonlinearity but very frequently by a lack of stability of the system under investigation. Impedance test equipment usually comprises an AC measurement unit and a potentiostat or galvanostat. For many applications, such as biomedical investigations or the characterization of thin films in which it is not essential to maintain a DC-voltage level during the impedance measurement, a potentiostat is not required (Házì et al., 1997).

EIS is frequently used to characterize systems that are changed during the time. Another way to reduce the effects of a system changing during the measurement is to reduce the total measurement time by using a multi-sine technique, which is FFT method. Figure 6 shows typical curve obtained by this technique for electrodeposition of Cd on a gold electrode in 0.1 M Cd SO_4.

In the case of multi-sine techniques, a measurement is carried out at several frequencies simultaneously. The phases of the superimposed signals are randomized to minimize the amplitude of the composite signal (see Fig. 5b). In contrast to single-sine techniques, multi-sine techniques do not require waiting for a full cycle to be completed for each of the frequencies used. The time-domain signals are digitized and transferred into the frequency domain by carrying out the FFT. The resulting data for each discrete frequency can be treated the same way as the impedance data obtained with a single sine technique. Repeated application of the waveform and averaging of the signal before FFT is applied can improve the S/N ratio of the multi-sine technique, although it increases the measurement time required. The impedance data obtained in this manner can be presented in several different formats.

Impedance spectra Z_{real} versus Z_{img} measured at a number of fixed DC potentials are suitable for quantitative studies of redox reaction kinetics whereas potential-dependent admittance values $(1/Z)$ obtained at fixed frequencies can provide AC voltammograms that are more readily interpreted in electroanalysis. The ratio of the two is the impedance of the test electrode. The measurement is then repeated at another frequency. The impedance spectra measured in this work were usually shaped as shown in the following Nyquist- and Bode-plot. Once real and imaginary parts of the input signals have been determined by correlation, the complex impedance of the test object can be calculated. It can be mathematically proven that all the spurious components are rejected by this technique of correlation provided that a sufficiently large number of cycles have been used for the integration (Garland et al., 2002).

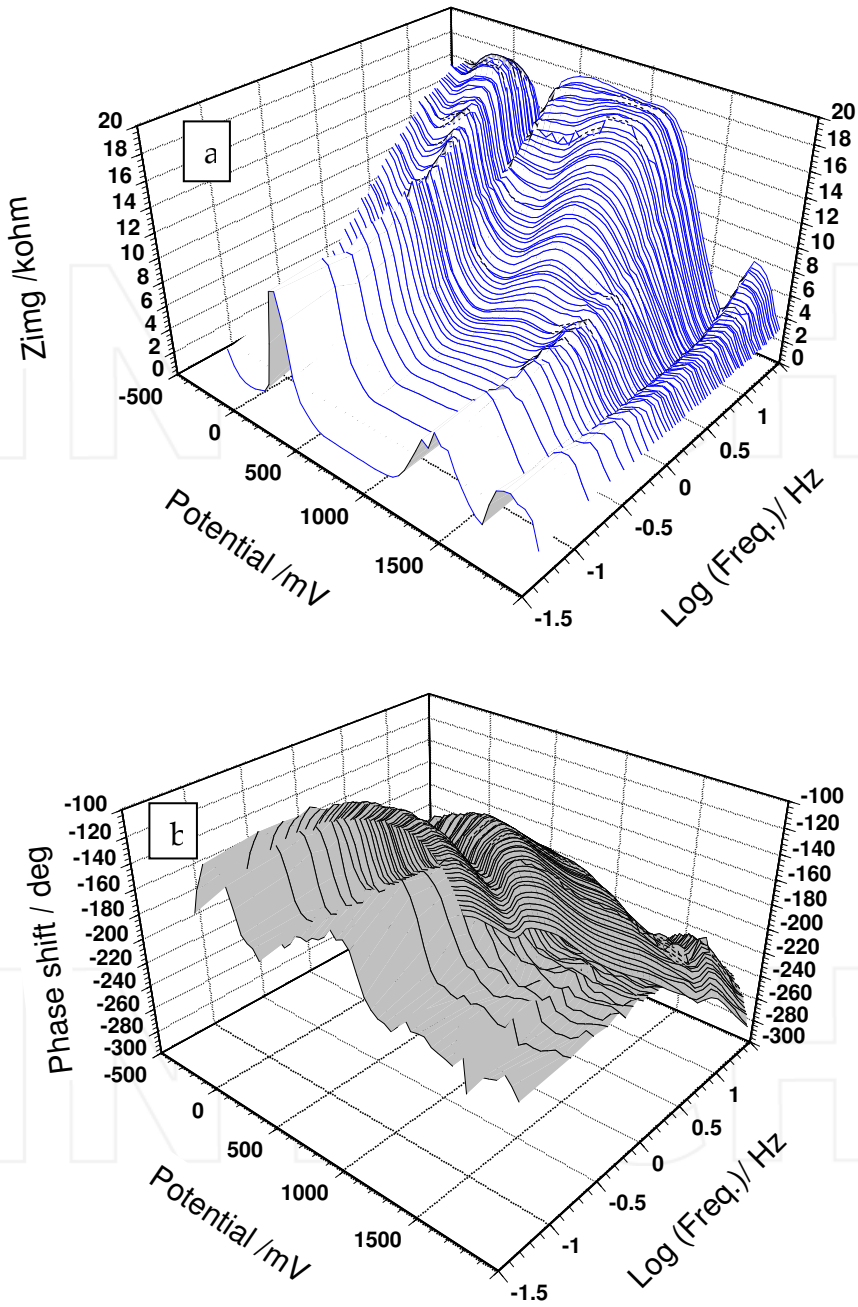


Fig. 5. Typical FFT EIS graph for electrodeposition of Cd on gold electrode, a) Z imaginary, b) phase shift, changes in deterrent frequency and potential

The advantage of modern FFT technique is that the information is obtained quickly; therefore it may be used to study impedances evolving with time. The limitation of the FFT technique is that the response to individual frequencies is usually weaker than that when only one frequency is used. It should be added that other types of analysis of system responses were also used, for example, Laplace transform of the applied perturbation and the response to determine the impedance spectra (Carstensen et al., 2008). The time-domain signals are digitized and transferred into the frequency domain by carrying out a FFT. The resulting data for each discrete frequency can be treated the same way as the impedance data obtained with a single sine technique.

In modern EIS analysis, lower frequency data are usually measured in the time domain. The current response is then measured using an A/D computer. In this case the FFT is used to convert the current signal into the frequency domain as carried out for other techniques. The FFT capabilities have subsequently been incorporated into several commercial instruments, primarily to speed up the acquisition of impedance data at low frequencies by exploiting the multiplex character of the technique. Such determinations are normally carried out at a single, fixed DC potential. In order to obtain potential-dependent impedance data, repeated experiments at different applied DC potentials are therefore required (Arundell et al., 2004).

The use of the FFT in combination with a controlled sequence of potential steps or pulses has been shown to offer an approach by which these time-consuming repetitions can be avoided and impedance measurements can be collected over a wide range of frequencies and DC potentials in a single experiment. However, here, the voltammetric waveforms composed of a sequence of potential steps are ideally suited in mathematical modeling based on the techniques of numerical integration. This approach is elegant in its generality, can be made arbitrarily precise, and is extremely efficient (Baranski et al., 1996). Baranski developed a technique in which a small amplitude square-wave potential perturbation is superimposed upon a potential staircase and in which the FFT is employed to convert current measurements taken as a function of time during several cycles of the square wave at each step potential to the frequency domain. Higher harmonics can be detected by FRAs or lock-in amplifiers, which can be tuned to detect a multiple of the excitation frequency. An alternative is to extract the harmonic signals from the response using FFT.

7. Application of FFT in SWV

The idea of obtaining electrode admittance from transient current time curves was investigated previously by Pilla (Pilla, 1972). The speed and sensitivity of SWV is its main advantage (Osteryoung & O'Dea, 1986). In the last years, data acquisition boards in the electroanalytical instrumentation have improved significantly for carrying out SW voltammetric analysis. The SWV has been used for study of the kinetics of electrode process (Winston et al., 1988). Nevertheless such applications of this method are relatively rare, which may be as a result of the rather complicated equations relating the current response of the electrode to the kinetic parameters of the electrode processes. Also, the theory of SWV does not take into account the effect of uncompensated solution resistance and a distortion of the EC signal by a slow response of a current transducer. These problems are not easy to be solved in any time domain voltammetric techniques. Because the electrode response under AC voltammetric conditions is represented as an admittance (i.e. in the frequency

domain), all data manipulations needed for obtaining kinetic information are relatively simple.

The principles of this technique are simple. Normally, FFTSWV experiments are done under conditions identical to the traditional SWV, but the electrode response at each DC potential is converted into the frequency domain via FFT. Therefore, FFTSWV measures the admittance of the electrode as a function of potential (Baranski & Szulborska, 1994). The resultant data are almost similar to those obtained under AC voltammetric conditions.

Indeed, in comparison with traditional AC voltammetry the equipment is much simpler and less expensive, measurements are carried out much faster and it is possible to obtain information about the admittance of the electrode at different frequencies from a single run.

The potential waveform used in the FFTSW voltammetric measurement consists of many SW pulses were superimposed on a staircase potential function, which was changed by a small potential step of ΔE (Norouzi et al., 2008). The values of potential pulse of SW (E_{SW}) and ΔE were in a range of few mV (10 to 50 mV). In the computer program, the number of SW cycles, Nc, in each staircase potential step was calculated based on the SW frequency as follows, Nc= f_0/1400Hz, for f_0 >1400Hz, and Nc=1 for f_0 ≤1400Hz. The values of Nc, f_o, E_{sw}, $E_{initial}$ and E_{vertex} were the variable parameters of the technique, which were optimized for achieving to best detector performance. It should be noted that in this method all processes studied involve adsorption of analytes hence both charging and faradic currents may potentially carry useful analytical information.

To get such information, it was important to sample data current at a frequency at least two times higher than the current transducer bandwidth. In order to fulfill this requirement the data sampling frequency was always adjusted between 50 and 100 kHz (depending on scan rate). In addition a second order low pass filter with a 20 kHz cutoff frequency was placed between the current output of the potentiostat and the data acquisition board. In the computer program, the discrete FFT analysis was used for data processing. If one SW cycle per potential step is applied, the time domain response resulted with this method is similar to that which obtained using Osteryoung SW voltammetry. Here either four data points per SW cycle were collected. If there is more than one cycle at one potential step, the current recorded in different cycles at the same DC potential is averaged (i.e. the first data points in every cycle are added together and divided by the number of cycles then the second and subsequent data points are treated in the same way) (Baranski & Szulborska, 1994).

The first component in Eq. 5 gives the imaginary part (Z_{img}) of the impedance and the second part gives its real component (Z_{rel}). A full discussion for the determination of Z_{rel} and Z_{im} based on the sampled currents (I_s) will be given in the next section. Theoretically, the detector impedance,

$$|Z| = \sqrt{Z_{img}^2 + Z_{rel}^2} \tag{5}$$

where $|Z|$ in a specific frequency is equal to E/I.

Application of discrete FFT analysis on the sampled current requires a specific method in current sampling. The admittance of the electrode is calculated at each potential step by the DFT method. High frequency components are removed by placing an analog low pass filter

between the current transducer and the A/D converter (Van Valkenburg, 1982). It required the number of sampled currents at each pulse cycle which must be represented by 2^n (where n is an integer and greater than 1). Therefore the currents, I_s, were sampled at even time intervals, t_s,

$$t_s = 1 + \frac{s}{4f_0} \tag{6}$$

where s is an integer number and changes from 0 to 7. Therefore if currents are sampled at even time intervals, t_s, $t_{s+1}/4f_0$, $t_{s+2}/4f_0$ and $t_{s+3}/4f_0$, then the values of the sampled currents will be,

$$i_0 = \sum_{n=1} A_n \sin(2n\pi f_0 t_c - \phi_n) \tag{7}$$

$$i_1 = \sum_{n=1} A_n \sin(n\pi / 2 + 2n\pi f_0 t_s + \phi_n) \tag{8}$$

$$i_2 = \sum_{n=1} A_n \sin(n\pi + 2n\pi f_0 t_s + \phi_n) \tag{9}$$

$$i_3 = \sum_{n=1} A_n \sin(n3\pi / 2 + 2n\pi f_0 t_s + \phi_n) \tag{10}$$

The equations show that in the first harmonic (n=1) the current components i_0 and i_2 (as well as, i_1 and i_3) have a phase shift equal to π. However, their absolute values are the same with an opposite sign. As mentioned above, the currents were sampled four times per SW cycle, i_0, i_1, i_2 and i_3. In each step, ΔE, of staircase potential ramp, the total sampled currents were 4Nc (Nci_0, Nci_1, Nci_2 and Nci_3), which were reduced to four by averaging each Nci. Because of dependence of Nc on frequency, at SW frequencies lower than 1400 Hz lager number of currents were averaged, which could be helpful for reducing the noise level. At the end of each potential ramp, the data were stored in an array matrix as follows,

$$\text{Data array} = \begin{Bmatrix} i_0^1 & i_1^1 & i_2^1 & i_3^1 & E_0 & E_1 & E_2 & E_3 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ i_0^n & i_1^n & i_2^n & i_3^n & E_0 & E_1 & E_2 & E_3 \end{Bmatrix} \tag{11}$$

where n is the number of the potential step and E_0 to E_4 are the electrode potentials at which the current is sampled.

To calculate the admittance of the detector response, first the real and imaginary components of the alternating current need to be calculated. The real component of I' and E′ are given by,

$$I' = i_2 - i_0 \tag{12}$$

$$E' = E_2 - E_0 = -2E_s \tag{13}$$

and the equation for the imaginary components are,

$$I'' = i_1 - i_3 \tag{14}$$

$$E'' = E_1 - E_3 = 2E_s \tag{15}$$

Now, the real, Y', and imaginary, Y'', components of the detecting admittance can be calculated as follows,

$$Y' - jY'' = \frac{I' - jI''}{E' - jE''} = \frac{I'' - I' - j(I' + I'')}{4E_S} \tag{16}$$

and

$$Y' = \frac{i_0 + i_1 - i_2 - i_3}{e_0 + e_1 - e_2 - e_3} \tag{17}$$

$$Y'' = \frac{i_0 - i_1 - i_2 + i_3}{e_0 + e_1 - e_2 - e_3} \tag{18}$$

Also, the average current, I_s at each potential step is given by

$$I_s = \frac{i_0 + i_1 - i_2 - i_3}{4} \tag{19}$$

The results from application of FFT analysis showed that measurement based on the first harmonic component offered better detection limits. Therefore, the components at higher frequencies than fundamental frequency were removed from the current response. The filtration was initially done by utilizing analog filters. A series of low pass filters were located before the A/D converter board. However, such filter may cause a small distortion in the magnitude and phase of the fundamental harmonic (which can be determined by calibrating). Also, a digital filtration was occurred during data acquisition. If the excitation potential and the electrode response can be represented by periodic functions, the electrode admittance can be calculated. Briefly in order to avoid problems with the interpretation of the electrode admittance, it is necessary to remove all components of the electrode response at frequencies higher than half the data acquisition frequency. This condition is known as the Nyquist sampling theorem (Weaver, 1983).

The examples of the SW voltammetric responses on the Au UME in the FIA measurement are shown in Figure 6. The analyte signal appears as a current decline in certain potential at the SW voltammogram. It results the inhibition of the electrode surface processes by the analyte adsorption process (Norouzi et al., 2009). To visualize the dependence of the analyte signal to the electrode potential in Figure 6a, the differential form of the SW voltammograms are shown (Figure 6b). In the differential graphs, also, it can be noted that the analyte signal extends over a potential range of the SW voltammogram.

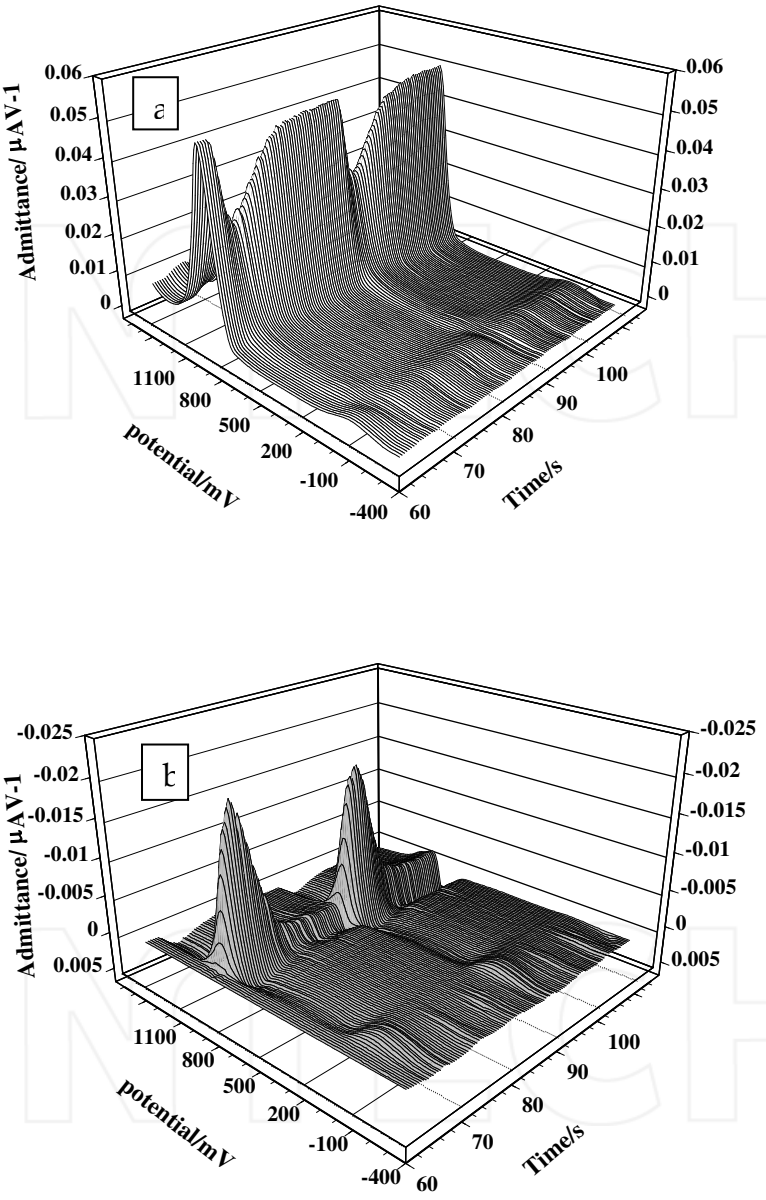


Fig. 6. a) FFT SW voltammograms at Au ultra-microelectrode recorded during the flow injection. The eluent was 0.05 M H_3PO_4 with the flow rate of 0.5mL/min., b) Curves resulted from subtracting the SWs in fig. a, from the average of 10 SWs (in the absence of analyte)

8. Conclusion

Electroanalytical techniques can offer rapid and low cost analysis of electroactive compounds and heavy metal ions in aqueous systems with a parts-per-billion sensitivity range. Electrical signals may be examined in both the time and in the frequency domain. The two display modes are related to each other by FTT, so each signal variable in the time domain has a characteristic frequency spectrum and vice versa. In the FFT based EC method (such as ENA, FFT Cyclic voltammetry, FFT SW voltammetry and FFT impedance spectroscopy), initially, an electrode response was recorded. Then, FFT was applied on the collected data and the existing high frequency noises were indicated. Based on this information, the cutoff frequency of the analog filter was set at a certain value (where the noises were removed from the electrode response). The smoothing function is, effectively, a moving average filter which is applied to the transfer function data before it is displayed in order to minimize the presence of jagged edges and discontinuities in the displayed data. Finally with the aid of this function a displayed trace can be smoothed by averaging over several electroanalytical measurements. Therefore the signal at the spectrum analyzer input may give rise to unwanted components which do not show any relationship to the input signal.

Some of the major advantages of FFT-voltammetry over other electrochemical techniques include:

- Speed: Because all of the frequencies are measured simultaneously, most measurements by FFT-voltammetry are made in a matter of nano seconds rather than several minutes.
- Sensitivity: Sensitivity is dramatically enhanced with FFT-voltammetry for many reasons. The detectors employed are much more sensitive, the electrical throughput is much higher which results in much lower noise levels, and the fast scans enable the co-addition of several scans in order to reduce the random measurement noise to any desired level (referred to as signal averaging).

Finally, the sensitivity and accuracy of electroanalytical methods based on FFT, along with a wide variety of software algorithms, have dramatically increased the practical use of voltammograms for quantitative analysis. Quantitative methods can be easily developed and calibrated and can also be incorporated into simple procedures for routine analysis. Thus, the FFT-electroanalytical techniques have brought significant practical advantages to other electroanalytical methods. It has made possible the development of many new sampling techniques which were designed to tackle challenging difficulties which were impossible by older technologies. It has made the application of electroanalytical analysis virtually limitless.

9. References

Aballe, A., Bethencourt, M., Botana, F.J. & Marcos, M. (1999). Using wavelets transform in the analysis of electrochemical noise data. *Electrochim. Acta*, 44, 26, 4805-4816.

Arundell, M., Patel, B.A., Yeoman, M.S., Parker, K.H. & O'Hare, D. (2004) Hilbert transform of voltammetric data . *Elecrochem. Commun.*, 6, 4, 366-372.

Baranski, A., & Szulborska, A. (1994) Fourier transform square-wave voltammetry. *J Electroanal. Chem.*, 373, 157-165.

Baranski, A.S., Norouzi, P. & Nelsson, L.J. (1996). *Proc. Electrochem. Soc.*, 9, 41.

Bond, A.M., Schwall, R.J. & Smith, D.E. (1977). On-line FFT faradaic admittance measurements application to A.C. cyclic voltammetry. *J Electroanal. Chem.*, 85, 2, 231-247.

Brett, C.M.A. & Brett, A.M.O. (1993). Electrochemistry: Principles, Methods, and Applications, Oxford University Press, Oxford.

Carstensen, J., Foca, E., Keipert, S., Foell, H., Leisner M. & Cojocaru, A. (2008) New modes of FFT impedance spectroscopy applied to semiconductor pore etching and materials characterization . *Phys. Status Solidi A*, 205, 11, 2485-2503.

Dai, Y. (2000). The time–frequency analysis approach of electric noise based on the wavelet transform. Solid-State Electron., 44, 12, 2147-2153.

Daneshgar, P., Norouzi, P., Ganjali, M.R. & Zamani, H.A. (2009). Ultrasensitive flow-injection electrochemical method for detection of anticancer drug tamoxifen. Talanta, 77, 3, 1075-1080.

Darowicki, K. & Zieliski, A. (2001). Joint time–frequency analysis of electrochemical noise. *J Electroanal. Chem.*, 504, 2, 201-207.

Ganjali, M.R., Norouzi, P., Ghorbani, M. & Sepehri, A. (2005). Fourier transform cyclic voltammetric technique for monitoring ultratrace amounts of salbutamol at gold ultra microelectrode in flowing solutions. Talanta, 66, 5, 1225-1233.

Garland, J.E., Assiongbon, K.A., Pettit, C.M., Emery, S.B. & Roy, D. (2002) Kinetic analysis of electrosorption using fast Fourier transform electrochemical impedance spectroscopy: underpotential deposition of Bi3+ in the presence of coadsorbing ClO4− on gold. *Electrochim. Acta*, 47, 25, 4113-4124.

Gavaghan, D.J. & Bond, A.M. (2000). A complete numerical simulation of the techniques of alternating current linear sweep and cyclic voltammetry: analysis of a reversible process by conventional and fast Fourier transform methods. *J Electroanal. Chem.*, 480, 1-2, 133-149.

Házì, J., Elton, D.M., Czerwinski, W.A., Schiewe, J., Vicente-Beckett, V.A. & Bond, A.M. (1997) Microcomputer-based instrumentation for multi-frequency Fourier transform alternating current (admittance and impedance) voltammetry. *J Electroanal. Chem.*, 437, 1-2, 1-15.

Lafront, A.M., Safizadeh, F., Ghali, E. & Houlachi, G. (2010) Study of the copper anode passivation by electrochemical noise analysis using spectral and wavelet transforms. *Electrochim. Acta*, 55, 7, 2505-2512.

Norouzi, P., Ganjali, M.R. & Meibodi, A.S.E. (2008) A novel adsorptive square wave voltammetric method for pico molar monitoring of lorazepam at gold ultra microelectrode in a flow injection system by application of fast Fourier transform analysis. *Anal. Lett.*, 41, 7, 1208-1224.

Norouzi, P., Ganjali, M.R., Daneshgar, P., Dinarvand, R., Moosavi-Movahedi, A.A. & Saboury, A.A. (2007). Development of fast Fourier transform continuous cyclic voltammetry at Au microelectrode in flowing solutions as a novel method for sub-nanomolar monitoring of lidocaine in injection and biological fluids. *Anal. Chim. Acta*, 590, 1, 74-80.

Norouzi, P., Ganjali, M.R., Nouryousefi, E., Shahtaheri, S.J. & Dinarvand, R. (2010). Trace Detection of Diphenhydramine by Adsorption on a Microelectrode at Flow Injection System by Fast Fourier Transform Continuous Cyclic Voltammetry . *Croatica Chem. Acta*, 83, 2, 135-142, b.

Norouzi, P., Ganjali, M.R., Shahtaheri, S.J., Dinarvand, R. & Hamzehpoor, A. (2009). Monitoring of Methyldopa by Fast Fourier Transform Continuous Cyclic Voltammetry at Gold Microelectrode .*Chinese J Chem.*, 27, 4, 732-738.

Norouzi, P., Garakani, T.M., Rashedi, H., Zamani, H.A. & Ganjali M.R. (2010) Ultrasensitive Flow-Injection Electrochemical Method Using Fast Fourier Transform Square-Wave Voltammetry for Detection of Vitamin B-1. *Int. J Electrochem. Sci.*, 5, 5, 639-652, c.

Norouzi, P., Rashedi, H., Mirzaei Garakani, T., Mirshafian, R., Ganjali, M. R. (2010). Fast Fourier transformation with continuous cyclic voltammetry at a Pt-Au dual microelectrode for the determination of Chloramphenicol in a flow injection system, *Int. J Electrochem. Sci.*, 5, 377 – 39, a.

Osteryoung, J. & O'Dea, J.J. (1986). Electroanalytical Chemistry, in Bard, A.J., (ed.) Vol. 14, Dekker, New York, p. 209.

Pilla, A.A. (1972). Computers in Chemistry and Instrumentation, in Mattson, J.S., Mark Jr., H.B. & MacDonald Jr., H.C. (eds.) Vol. 2, Dekker, New York.

Popkirov, G.S. (1996). Fast time-resolved electrochemical impedance spectroscopy for investigations under nonstationary conditions. *Electrochem. Acta*, 41, 7-8, 1023-1027.

Rauscher, C., Janssen V. & Minihold R. (2001). Fundamentals of Spectrum Analysis, First edition, Rohde & Schwarz GmbH & Co. KG, Mühldorfstrasse 15, 81671 München, Germany.

Safizadeh, F. & Ghali E. (2010). Monitoring Passivation of Cu-Sb and Cu-Pb Anodes During Electrorefining Employing Electrochemical Noise Analyses. *Electrochim. Acta*, In Press, Accepted Manuscript, Available online 21 September.

Sang, Y.F., Wang, D., Wu, J.C., Zhu, Q.P. & Wang, L. (2009). The relation between periods' identification and noises in hydrologic series data. *J Hydrol.*, 368, 1-4, 30 165-177.

Smith, D.E. (1976) *Anal. Chem.*, 48, 221A-517A.

Van Valkenburg, M.E. (1982). Analog Filter Design, Holt, Rinehart and Winston, New York.

Weaver, H.J. (1983). Applications of Discrete and Continuous Fourier Analysis, Wiley, New York.

Weaver, H.J. (1983). Applications of Discrete and Continuous Fourier Analysis, Wiley, New York.

Winston, S., O'Dea, J.J., & Osteryoung, J. (1988). Square wave voltammetry for the determination of kinetic parameters: The reduction of zinc(II) at mercury electrodes. *J Electroanal. Chem.*, 255, 21-44.

Zaveri, N., Sun, R., Zufelt, N., Zhou, A. & Chen, Y.Q. (2007). Evaluation of microbially influenced corrosion with electrochemical noise analysis and signal processing. *Electrochim. Acta*, 52, 19, 5795-5807.

16

Fourier Transforms Infrared Spectroscopy Applied in Selective Catalytic Reduction of NO by Acetylene

Xinping Wang
Dalian University of Technology
China

1. Introduction

Selective catalytic reduction of NO by hydrocarbons (HC-SCR) is believed to be one of the most promising ways to remove nitric oxide from the exhaust gas of diesel and lean-burn engines. Since HC-SCR was reported individually by Iwamoto and Held groups (Iwamoto et al., 1999; Held et al., 1990), many studies were carried out in this field both on new catalyst research and on reaction mechanism. In these studies, Fourier Transforms Infrared Spectroscopy (FTIR) was extensively used for interpreting the relationship between surface structure and the catalytic performance of the catalysts, especially, for disclosing the reaction mechanism over the catalysts. In this chapter, we summarize the FTIR studies used in the investigation of selective catalytic reduction of NO by acetylene (C_2H_2-SCR).

2. To characterize surface acidity of the catalyst

For HC-SCR, many zeolites have been investigated as catalysts. However, it seems that only limited types of zeolites including ZSM-5 (Li et al., 2004; Li et al., 2008; Wang et al., 2007a; Wang et al., 2007b; Wang et al., 2005; Niu et al., 2006), ferrierite (Seijger et al., 2003; Lee et al., 2003; Kubacka et al., 2005; Kubacka et al., 2006) and mordenite (Pieterse & Booneveld, 2007; Mosqueda-Jiménez et al., 2003a; Córdoba et al., 2005; Lónyi et al., 2007; Dorado et al., 2005), which are called as pentasil zeolites, displayed high selectivity to NO_x reduction in HC-SCR. For instance, Pieterse et al. (Pieterse et al., 2003) investigated the selective catalytic reduction of NO by methane (CH_4-SCR) over several zeolite-based catalysts and reported an activity order of Co-Pd-ZSM-5 > Co-Pd-MOR > Co-Pd-FER > Co-Pd-BEA. Abreu and his co-workers (Torre-Abreu et al., 1999) found that zeolites modified by Cu used for selective catalytic reduction of NO by propene (C_3H_8-SCR) are ordered in NO conversion at 400 °C with CuMFI (58%) > CuMOR (43%) > CuY (20%). Sultana et al. (Sultana et al., 2008) reported that zeolites are ordered with Pt-MOR (90%) > Pt-FER (77%) > Pt-ZSM-5 (74%) > Pt-BEA (70%) > Pt-HY (63%) in terms of NO conversion for selective catalytic reduction of NO at ~300 °C using diesel as reductant. Shibata et al. (Shibata et al., 2004) reported an order of Ag-MFI (56%) > Ag-BEA (38%) > Ag-MOR (11%) >> Ag-Y (2%) in NO conversion at ~300 °C for the C_3H_8-SCR assisted by H_2. It was also reported that Co-ZSM-5 (Ivanova et al., 2001) and Ni-ZSM-5 (Mihaylov et al., 2004) are more active than the corresponding Y zeolite promoted by

Co or Ni for the CH_4-SCR. All these results indicated that Y is inferior to the pentasil zeolites for the HC-SCR.

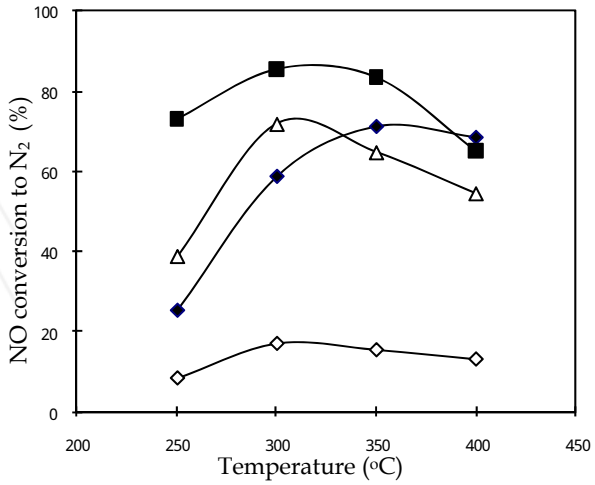


Fig. 1. Conversion of NO_x as a function of temperature over HZSM-5 (◆), HFER (■), HMOR (△) and HY (◇). Reaction conditions: 1600 ppm NO + 800 ppm C_2H_2 + 9.95% O_2 in He with a total flow rate of 50 ml/min over 0.2 g of catalyst

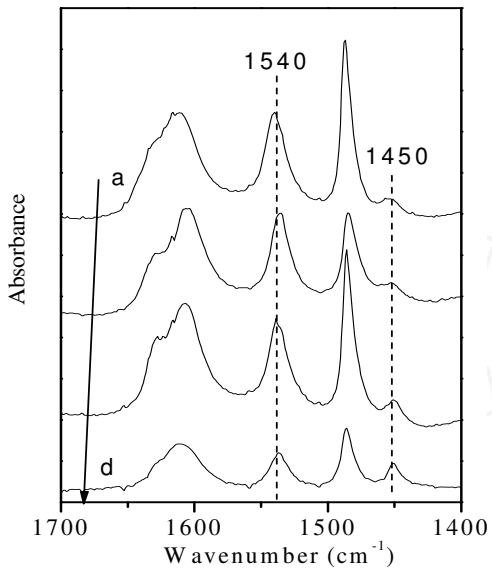


Fig. 2. FTIR spectra of pyridine adsorbed on HZSM-5 (a), HFER (b), HMOR (c) and HY (d) at 500 oC in evacuation

On the other hand, Shichi and co-workers (Shichi et al., 1998; Shichi et al., 2000; Shichi et al., 2001a; Shichi et al., 2001b; Shichi et al., 2004) have emphasized intracrystalline diffusion of reductants limited by the zeolites' channels, based on their investigation results. They found that NO conversion over MFI and MOR appeared to be significantly influenced by both hydrocarbon molecular size and zeolite particle size in some instances. In our investigation of HC-SCR using acetylene as reductant (C_2H_2-SCR), we found that the C_2H_2-SCR over HZSM-5 is greatly affected by the intracrystalline diffusion of NO_2 as well (Wang et al., 2007c). Thus, we suppose that the limited diffusion of the reactants in narrow channels of the pentasil zeolites is the main reason leading to the low NO conversion at high GHSV in HC-SCR. Due to wider channels compared to the pentasil zeolites (Elzey et al., 2008), Y zeolite may be favorable for intracrystalline diffusion of reactants. Besides of this, Y zeolite is stable in severe hydrothermal conditions and economical (Furusawa et al., 2002). Therefore, the zeolite could be expected to be a candidate for preparing a practical catalyst for HC-SCR working at high GHSV, if Y can be effectively modified, due to its larger pore size (0.74 nm × 0.74 nm) than HZSM-5 (0.53 nm × 0.56 nm), HFER(0.42 nm × 0.54 nm) and HMOR (0.65 nm × 0.70). It lead us to study the zeolite in C_2H_2-SCR investigation. However, a result in contrast with the expectation was obtained in the C_2H_2-SCR. HY displayed much low activity compared to HZSM-5, HFER, and HMOR, as shown in Fig 1. To answer the question why HY with the ideal larger pore size exhibited much inferior catalytic performance to the pentasil zeolites in the HC-SCR, we studied the surface acidity of HY in comparison with the pentasil zeolites (Ma et al., 2009). Fig.2 shows FTIR of pyridine adsorbed on different zeolites, obtained by pyridine adsorption over the zeolites and a degassing at 500 °C. Compared to HZSM-5, HFER and HMOR, HY gave considerable weaker IR band at 1540 cm^{-1} being associated with pyridine adsorption on protons, indicating that strong Brönsted acids over HY are much less than those of the pentasil zeolites in amount. In literature, zeolites with Na or H form were usually used in HC-SCR investigation as catalyst or support, and opposite results were obtained by the authors on the acidity or basicity of the zeolites favorable for HC-SCR activity, for different HC-SCR catalytic systems. For instance, it was reported that Ag-NaZSM-5 catalyst was more active than Ag-HZSM-5 for the selective catalytic reduction of NO by methane (CH_4-SCR) at 450 °C (Shi et al., 2002), and that Fe-ZSM-5 catalyst with precursor of NaZSM-5 was far more active than that with precursor of NH_4-ZSM-5 for selective catalytic reduction of NO by urea (Sullivan & Keane, 2005). Similarly, high activity for the selective catalytic reduction of NO by propene (C_3H_6-SCR) on Ce-NaZSM-5 (Niu et al., 2006; Seijger et al., 2003) and for CH_4-SCR on Pt-CoNaFER washcoated cordierite monolith (Lee et al., 2003) was obtained. Whereas, Brönsted acids have been suggested by the other authors to be essential for HC-SCR over many catalytic systems (e.g. ZSM-5 modified by Pd, Ga, In, Ce and Ag) (Shibata et al., 2004; Loughran & Resasco, 1995; Kikuchi & Yogo, 1994; Nishizaka & Misono, 1994; Li & Armor, 1994; Narbeshuber et al., 1997; Berndt et al., 2003; Gutierrez et al., 2005).

The authors found that Brönsted acids contributed to the aimed reaction in different steps. For instance, Stakheev and coworkers reported that exchange of partial protons by sodium with a level of 32% resulted in a nearly complete disappearance of the activity of the zeolite for oxidation of NO to NO_2, and a significant decrease of the activity for C_3H_6-SCR (Gutierrez et al., 2005). In the C_2H_2-SCR, we found that the proton form of ZSM-5 based catalyst were much active, whereas the sodium form of ZSM-5 based catalyst almost inactive for the reaction (Wang et al., 2007a; Wang et al., 2005), as shown in table 1. We also found that, with SiO_2/Al_2O_3 ratio of HZSM-5 increasing, C_2H_2-SCR activity of the

HZSM-5 based catalysts (e.g. Ce-HZSM-5, 2%Mo/HZSM-5 and 2%Zr/HZSM-5(Wang et al. 2008) as shown in table 2) decreased in same reaction conditions. These results indicated that protons presenting in the zeolites is indispensable for C_2H_2-SCR and activity of the catalyst strongly depends on the population of protons in the zeolites. Thus, one could believe that it is the weaker surface acidity of HY compared to those of pentasil zeolites which leads to the inferior catalytic performance of HY in comparison with HZSM-5, HFER and HMOR.

Catalysts	Reaction temperature (°C)	NO conv. to N_2 (100%)	Total C_2H_2 conv. (100%)
HZSM-5	300	58.9	86.6
NaZSM-5	325	~0	8.0
Ce-HZSM-5	300	83.0	98.9
Ce-NaZSM-5	300	~0	15.3
Mo-HZSM-5	350	82.7	100
Mo-NaZSM-5	350	~0	14.7

Table 1. Catalytic performance of Na-, and HZSM-5 based catalysts in C_2H_2-SCR*
*Note: All of the catalyst were prepared from ZSM-5 with SiO_2/Al_2O_3 of 25

Reaction conditions: 1600 ppm NO + 800 ppm C_2H_2 + 9.95% O_2 in He with a total flow rate of 50 ml/min over 0.2 g of catalyst.

Catalyst with SiO_2/Al_2O_3 ratio	Reaction temperature (°C)	NO conv. to N_2 (100%)	Total C_2H_2 conv. (100%)
Ce-25	300	83.0	98.9
Ce-38	300	64.2	100
Ce-50	300	~0	79
Mo-25	350	82.7	100
Mo-38	350	57.7	100
Mo-50	350	48.0	100
Zr-25	350	89	100
Zr-38	350	73	100
Zr-50	350	66	100

Table 2. Catalytic performance of Ce-HZSM-5, 2%Mo/HZSM-5 and 2%Zr/HZSM-5 prepared from the zeolite with different SiO_2/Al_2O_3 ratio*
*Note: Ce-HZSM-5 was prepared by ion exchange, and 2%Mo/HZSM-5 and 2%Zr/HZSM-5 were prepared by impregnation

3. To confirm rate-determining step

Over many kinds of catalysts, HC-SCR was believed to be initiated from NO oxidation to NO_2, and the step was generally accepted to be crucial step (Kubacka et al., 2006; Pieterse & Booneveld, 2007). To elucidate why HY displayed inferior catalytic performance in C_2H_2-SCR, activity of HY for NO oxidation to NO_2 was compared to those of HZSM-5, HFER and HMOR. As shown in Fig. 3, all of the pentasil zeolites exhibited considerably higher activity compared to HY, in particular, in the case of HFER at 250 °C. The results implies that the reaction of $NO + O_2 \longrightarrow NO_2$ is catalyzed by acid sites of the zeolite. It was strongly supported by experimental results that when sodium form of ZSM-5 was used as catalyst for the reaction of NO oxidation, almost no catalytic activity could be observed. Combined the catalytic activity of the zeolites both for NO oxidation and for C_2H_2-SCR of NO, it can be found that the curves of NO oxidation to NO_2
vs. reaction temperature resemble much to those of NO conversion to N_2 in C_2H_2-SCR in patterns, indicating that C_2H_2-SCR is significantly confined by the step of $NO + O_2 \longrightarrow NO_2$ over HY.

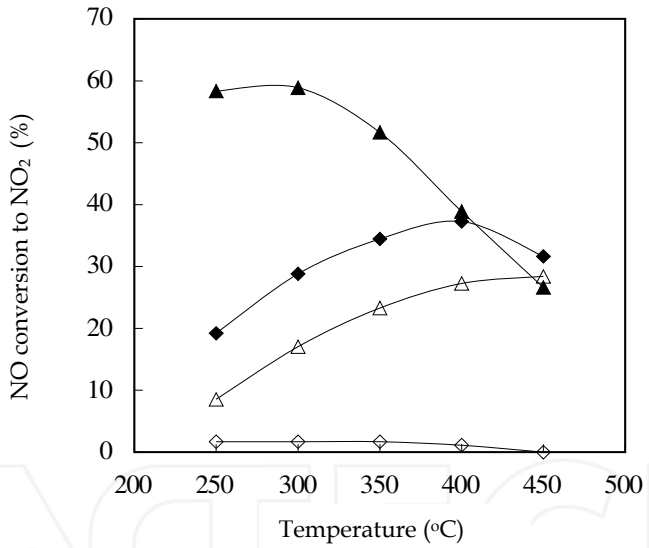


Fig. 3. Catalytic performance of HZSM-5 (◆), HFER (▲), HMOR (△) and HY (◇) in oxidation of NO with O_2 at different temperatures. Reaction conditions: 200 ppm NO + 10% O_2 in N_2 was fed to 0.200 g at a total flow rate of 100 ml/min

For C_2H_2-SCR over HZSM-5 and HFER, it can be speculated that the step of NO oxidation to NO_2 may not be the rate-determining step, as the two zeolites are rather active to NO oxidation, due to their lager amount of protons characterized by FTIR of pyridine adsorption (Fig. 2). The speculation was confirmed by the following experimental results. As shown in Fig. 4, although HZSM-5 itself is more active for NO oxidation with O_2 at the reaction temperature ranged from 200~400 °C than all of xY/HZSM-5 (x = 0 ~ 9), it gave a rather lower NO conversion to N_2 compared to 3%Y/HZSM-5, as shown in Fig. 5 (Wang et

al., 2008). To investigate the rate-determining step of C_2H_2-SCR over these catalysts, and effectively research more active catalyst for the reaction of C_2H_2-SCR, several possible step for the reaction over the catalysts, such as NO_x adsorption over the catalysts and the activity of the nitrous species toward reduction over the catalysts were investigated. At first, FTIR spectrum arising from nitrous species on HZSM-5 and 3%Y/HZSM-5 after saturated co-adsorption of NO+O_2 was measured at 250 °C, as shown in Fig. 6. Compared to HZSM-5, 3%Y/HZSM-5 gave significantly stronger band at 1585 cm^{-1} due to bidentate nitrates (Ivanova et al., 2001; Yu et al., 2004; Shimizu et al., 2001; He et al., 2004; Poignant et al., 2001) and a new band at 1609 cm^{-1} caused by NO+O_2 saturated co-adsorption at 250 °C. The results indicated that the nitrous species adsorption capacity of 3%Y/HZSM-5 is much stronger in comparison with that of HZSM-5. NO- and NO_2-TPD on the catalyst samples gave also the same conclusion. As shown in Fig. 7, substantially larger amounts of NO and NO_2 were measured in the temperature range of 250-550 °C on 3%Y/HZSM-5 in the NO_x-TPD compared to that on HZSM-5. It should be noted that the corresponding nitrous species which desorbed from the catalyst surface in the temperature range quite seems to be those reacting with C_2H_2 and contributing C_2H_2-SCR of NO, as large parts of them could be speculated to stay on the catalyst surface at the C_2H_2-SCR reaction conditions. By combining the C_2H_2-SCR activity with the nitrous species adsorption capacity of the catalysts, a consistent relationship was obtained for Mo-(Wang et al., 2007a), W-(Wang et al., 2007b), Zr-promoted HZSM-5 as well as Zr-promoted HFER, i.e. by these metals incorporating into HZSM-5 and HFER, nitrous species adsorption capacity of the catalyst was significantly enhanced, and at the same time, C_2H_2-SCR activity of the catalyst was largely improved.

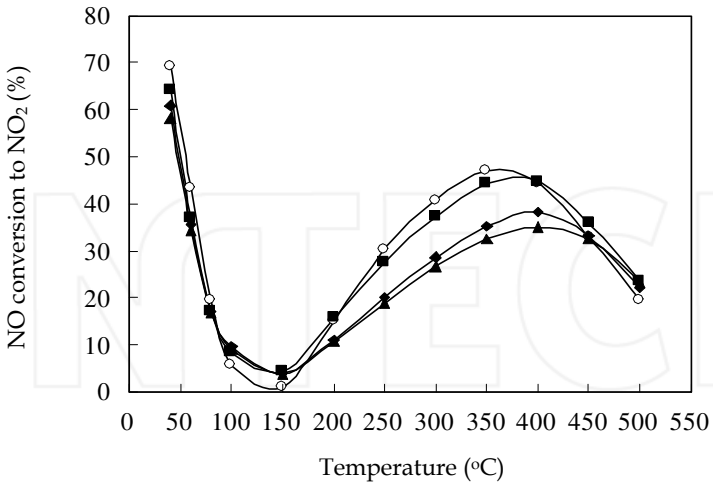


Fig. 4. Catalytic activity of xY/HZSM-5 for NO oxidation with O_2: x=0 (○), x=0.5 (■), x=3 (♦) and x=9 (▲). Reaction conditions: 200 ppm NO, 10 % O_2 in N_2 with a total flow rate of 100 ml/min over 0.200 g catalysts

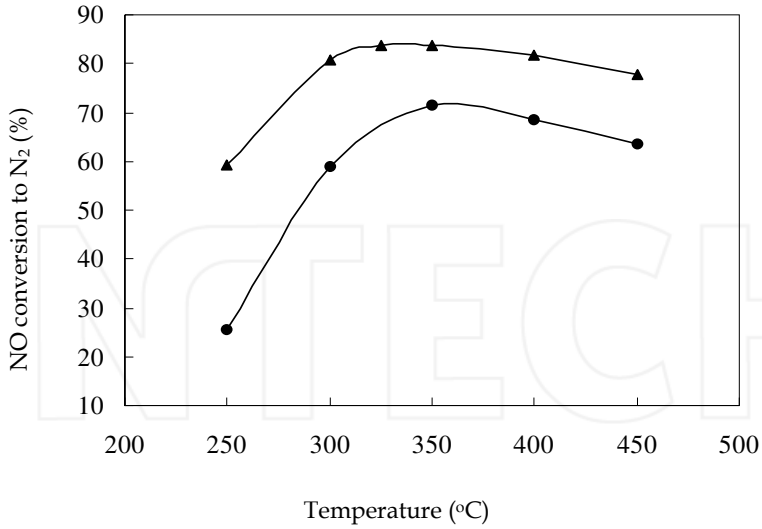


Fig. 5. NO conversion to N_2 as a function of reaction temperature over HZSM-5 (●), and 3%Y/HZSM-5 (▲). Reaction condition: 1600 ppm NO, 800 ppm C_2H_2, 9.95% O_2 in He with a total flow rate of 50 ml·min^{-1} over 0.200 g of catalyst

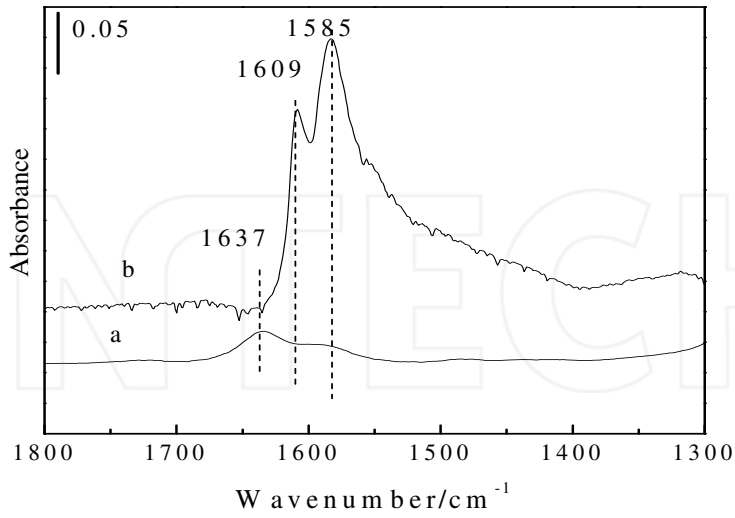


Fig. 6. FTIR spectra arising from nitrous species on HZSM-5 (a) and 3% Y/HZSM-5 (b) after saturated co-adsorption of NO+O_2 at 250 °C

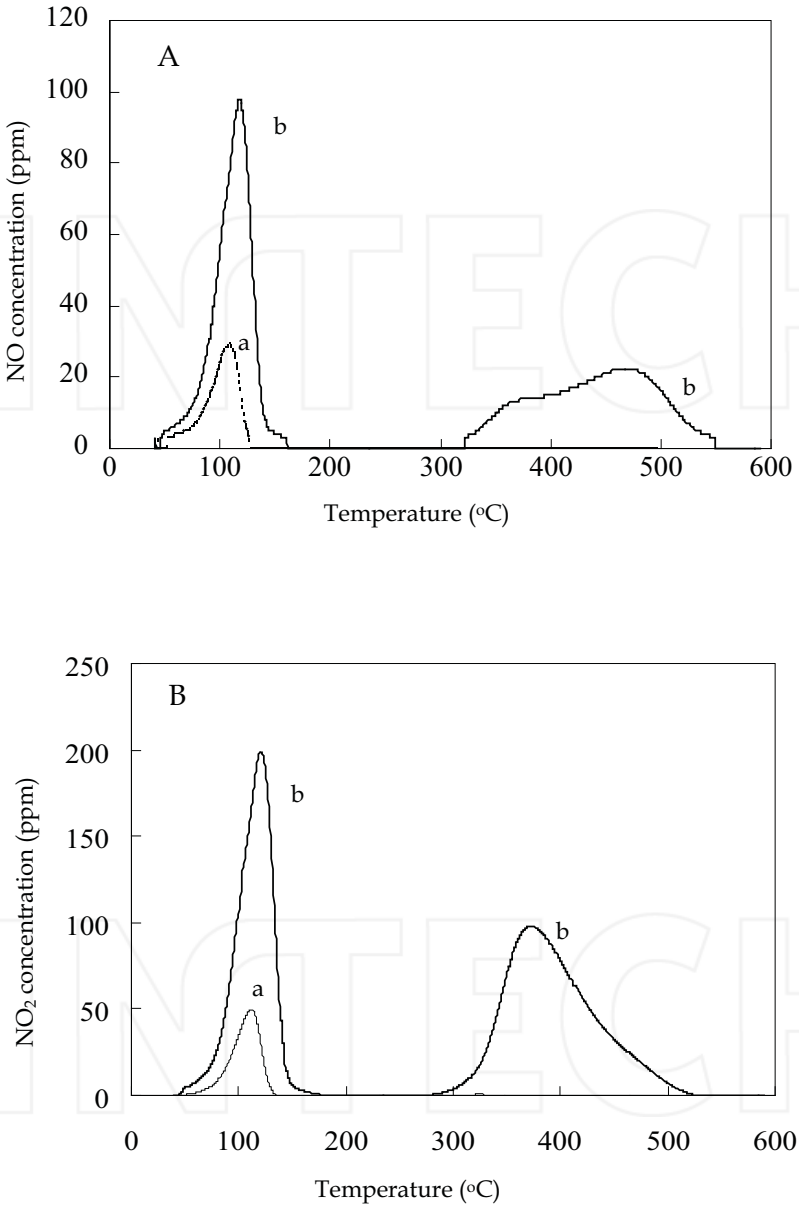


Fig. 7. TPD profiles of NO (A) and NO_2 (B) over HZSM-5 (a) and 3% Y/HZSM-5 (b) after saturated co-adsorption of $NO+O_2$

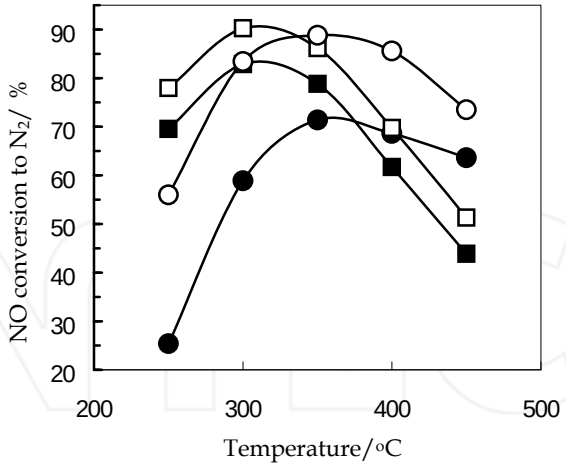


Fig. 8. The NO conversion to N_2 as a function of temperature over HZSM-5 (●), 2%Zr/HZSM-5 (○), HFER (■) and 2%Zr/HFER (□). The reaction conditions are the same as that in Fig. 5

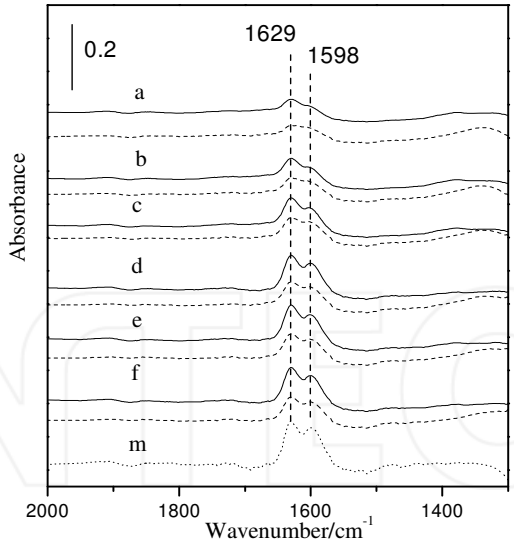


Fig. 9. Transient FTIR spectra at 250 °C upon HFER (solid curve) and HZSM5 (dashed curve) exposing to 1000 ppm NO + 10 % O_2 in N_2 (A) for 1 min (a), 2 min (b), 5 min (c), 10 min (d), 15 min (e) and 30 min (f). For comparison, the transient FTIR spectrum recorded by exposing 2%Zr/HZSM-5 to 1000 ppm NO + 10 % O_2 in N_2 at 250 °C for 30 min (dotted curve) was given as **m**

These results imply that the rate-determining step for the C_2H_2-SCR over HZSM-5 and HFER is the nitrous species formation from NO_x adsorption. Due to the restriction in length for the chapter, here, we will interpret only the rate-determining step of C_2H_2-SCR over HZSM-5 and HFER deduced from FTIR study on Zr and Y promoted zeolites.

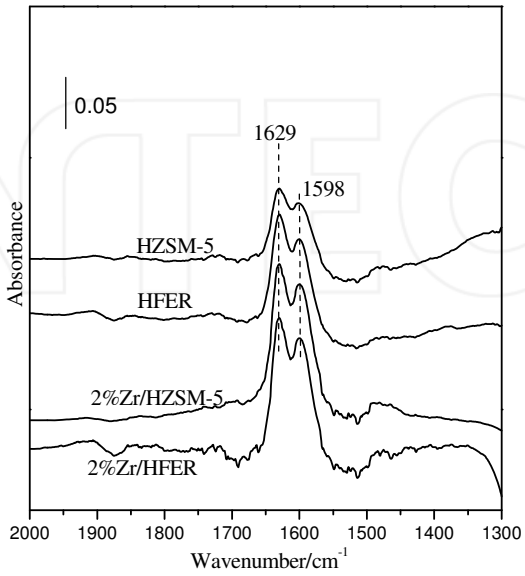


Fig. 10. FTIR spectra arising from surface nitrate species on HZSM-5, 2%Zr/HZSM-5, HFER, and 2%Zr/HFER upon the catalyst samples exposing to gas mixture of 1000 ppm NO + 10 % O_2 in N_2 for 30 min at 250 ^oC

The catalytic performance of HFER, HZSM-5, 2%Zr/HFER and 2%Zr/HZSM-5 in C_2H_2-SCR was shown in Fig. 8. The conversion of NO into N_2 was strongly influenced by the type of zeolites. The parent zeolite of HFER displayed much higher activity than HZSM-5 for C_2H_2-SCR in the temperature range of 250-375 ^oC. For instance, NO conversion to N_2 over HFER at 250 ^oC was 69.6%, which is much higher than that over HZSM-5 under the same reaction conditions. On the other hand, significant doping effect of zirconium incorporation into the zeolites on C_2H_2-SCR was observed, especially in the case of HZSM-5. For instance, NO conversion to N_2 was increased to 78.0% from 69.6% at 250 ^oC by 2% of zirconium incorporation into HFER, whereas it sharply increased to 56.0% from 25.4% in the case of HZSM-5. Why HFER was so much active at the lower reaction temperature compared to HZSM-5? Why the doping effect of zirconium on C_2H_2-SCR was so much larger for HZSM-5 than HFER? To answer these questions, we tentatively combined their catalytic performance in C_2H_2-SCR with their NO_x adsorption capacity. As shown in Fig. 9, Two bands at 1598 and 1629 cm^{-1} due to bidenate and bridging nitrates (Li et al., 2007; Tsyntsarski et al., 2003; Li et al., 2005a) respectively appeared on HZSM-5 and HFER when the zeolites were exposed to NO+O_2 at 250 ^oC. By comparing the band increasing in intensity at 1629 and 1598 cm^{-1} over the zeolites, it can be known that the nitrate species formation was faster over HFER than that over HZSM-5. Obviously, the relative nitrate species formation rate coincides well with

the relative C_2H_2-SCR activity for the two parent zeolites. Similar relationship of C_2H_2-SCR activity versus nitrate species formation rate can be also obtained on HZSM-5 and 2%Zr/HZSM-5. The bands due to nitrate species on 2%Zr/HZSM-5 (spectrum **m** in Fig. 9) were obviously strong in intensity compared to that obtained at the same exposing time over HZSM-5. Figure 10 gave the stable FTIR spectrum at 250 °C obtained by exposing HZSM-5, HFER, 2%Zr/HZSM-5 and 2%Zr/HFER respectively to gas mixture of 1000 ppm NO + 10 % O_2 in N_2 until the corresponding spectra no further change. By combining the results given in this figure with those in Fig. 8, the relative activity of the zeolites and those promoted by zirconium can be well understood. The lager nitrate species formation capacity of HFER compared to HZSM-5 may be the reason leading to the higher C_2H_2-SCR activity of HFER in comparison with HZSM-5. Accordingly, as 2%Zr/HFER have the further large nitrate species formation capacity, due to 2% of zirconium incorporation, it displayed more active than the zeolite itself. Also, the drastically larger C_2H_2-SCR activity of 2%Zr/HZSM-5 than that of the zeolite itself can be attributed to the substantially increased nitrate species formation capacity of the resulting material prepared from zirconium impregnation on the zeolite. To confirm the supposition that nitrate species formation on the catalyst surface is the rate-determining step for the C_2H_2-SCR over HZSM-5 and HFER, it must be validated that the nitrate species are important intermediate of the total reaction, i.e. they are active toward reduction, which will be discussed in the section 4. It should be also validated that, the step of nitrate species formation on the catalyst surface is most slow among all of the steps in the route of C_2H_2-SCR over the zeolites.

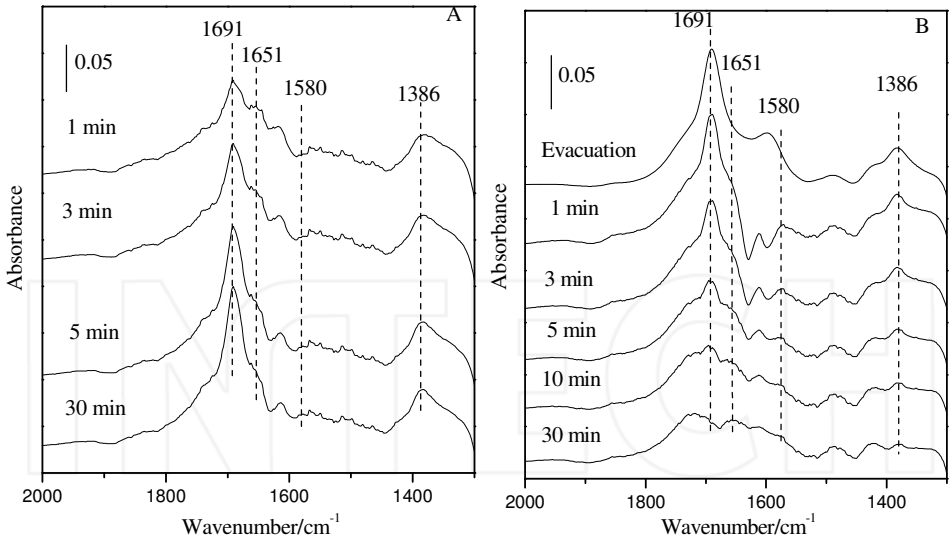


Fig. 11. FTIR spectra of surface species on HFER at 250 °C upon the fresh zeolite being exposed to 500 ppm C_2H_2 + 1000 ppm NO + 10 % O_2 in N_2 (A) and then to 1000 ppm NO + 10 % O_2 in N_2 (B)
Note: The cell was evacuated briefly before the zeolite was exposed to NO+O_2 in N_2

Figure 11 gave the transient in situ FTIR spectrum of surface species on HFER at 250 ^{o}C recorded when the zeolite was exposed to 500 ppm C_2H_2 + 1000 ppm NO + 10 % O_2 in N_2 and then reductant gas of acetylene was cut off thereafter. Although bands at 1598 and 1629 cm^{-1} due to nitrate species with strong intensity appeared in the spectra recorded when the zeolites were exposed to 1000 ppm NO + 10 % O_2 in N_2 (Fig.10), no one of them appeared in the transient in situ FTIR spectra as well as in the steady spectrum when 500 ppm C_2H_2 was introduced into the gas mixture, i.e. in the C_2H_2-SCR reaction condition. Furthermore, the nitrate species even could not be detected in the in situ FTIR in the followed period of 30 min (Fig 11B), even though C_2H_2 was completely removed out from the gas mixture. These experimental results indicate that no nitrate species (produced by NO+O_2 co-adsorption over the zeolite) can cumulate on the zeolite surface under the C_2H_2-SCR reaction condition. Instead, the species produced by adsorption of acetylene or its derivates obviously cumulated on the zeolite surface, being characterized by the bands at 1691, 1651 and 1380 cm^{-1}. It means that formation of the reductant species is faster by far than that required by nitrate species reduction. In other words, the rate of nitrate species reduction is limited by that of nitrate species formation. Thus, it can be concluded that the rate-determining step of C_2H_2-SCR over the HFER is the nitrate species formation step. Then, it can be further deduced that the C_2H_2-SCR reaction over HZSM-5 must be controlled also by this step, as discussed above, nitrate species formation capacity of HZSM-5 is much lower than that of HFER, and C_2H_2-SCR activity enhanced by zirconium impregnation was more significant on HZSM-5 than on HFER. Finally, it should be noted that, the quantitive evidence obtained by FTIR must be based on a strict experiment, e.g., the surface species must be measured over the self-supporting wafer with the same weight. Otherwise, accurate FTIR result may be hardly obtained. To compare the amount of surface species formed on the catalysts in FTIR, catalyst wafers with 14 ± 0.7 mg were selected and certain parameters of the IR spectrophotometer were set in our experiments.

4. To identify reaction intermediates

The identification of reaction intermediates is much important not only for investigating the reaction route over a catalyst, but also for understanding the effect of doping material being introduced to the catalyst. As motioned above, in the C_2H_2-SCR investigation, we found that Mo, W, Zr, Y impregnated on HZSM-5 and HFER significantly enhanced the formation of nitrate species over the catalyst and correspondingly increased the C_2H_2-SCR activity. These results could be well understood by the rate-determining step discussed in the section 3. On the other hand, we found that, NaZSM-5 is almost inactive for C_2H_2-SCR, though it can be considered that NaZSM-5 has much stronger nitrate species formation capacity compared to HZSM-5. Certainly, one can considered that due to lack of protons and inert for catalyzing NO oxidation, it can not initiate the aimed reaction. We also found that when NO in the gas mixture was completely changed to NO_2, the NO_2 conversion to N_2 over NaZSM-5 was still less much than that over HZSM-5. These results imply that NaZSM-5 hardly catalyzes C_2H_2-SCR due to not only inactive for NO oxidation, but also the other inferior action displayed in the subsequent steps. Thus, the role of protons in the NO reduction by acetylene over HZSM-5 was studied.

4.1 Characterization of nitrous species on HZSM-5 and NaZSM-5

The nitrous species formation and variation with temperature, as well as their reactivity towards to reduction by acetylene were investigated by FTIR. The spectra obtained upon

exposing HZSM-5 and NaZSM-5 to NO and O_2 at 40 °C were depicted in Fig. 12. On HZSM-5 (Fig.12 A), the bands at 1622 cm^{-1} due to weakly adsorbed NO_2 (Shimizu et al., 2001;

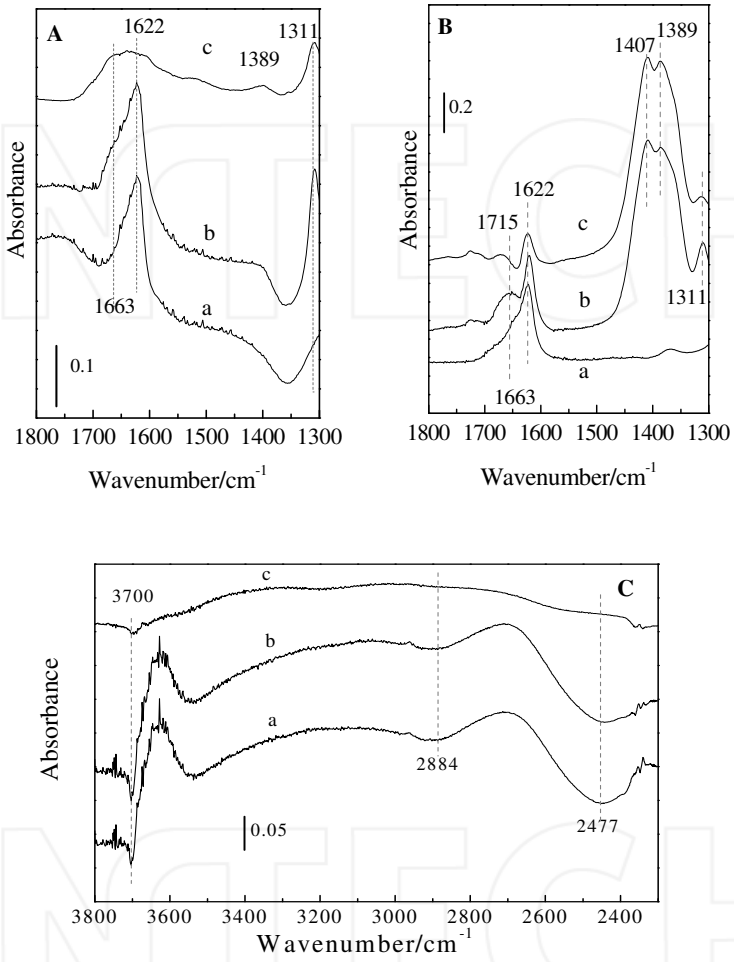


Fig. 12. FTIR spectra of surface nitrous species formed upon exposing HZSM-5 (A,C) and NaZSM-5 (B) to 1000 ppm NO and 10 % O_2 in N_2 at 40 °C for 1min (a), reaching saturated adsorption (b), and then a brief evacuation (c)

Pirngruber & Pieterse, 2006) or bridging nitrates (Ivanova et al., 2001; Yu et al., 2004; Sedlmair et al., 2003b; Li et al., 2005b) and that at 1311 cm^{-1} due to unidentate nitrates(Mihaylov et al., 2004; Shimizu et al., 2001; He et al., 2004) with three negative bands at 3700, 2884 and 2477 cm^{-1} were observed upon exposing the zeolite to $NO+O_2$ for 1 min. The band at 3700 cm^{-1} is due to asymmetric stretching vibration of molecular adsorbed water (Nakamoto, 1996;

Hadajiivanov et al., 1998), and the bands at 2884 and 2477 cm^{-1} are due to the well-known A-B-C structure produced by hydrogen-bonded hydroxyls (Mihaylov et al., 2004; Hadajiivanov et al., 1998). The results indicate that the formation of nitrates occurred at the expense of molecular adsorbed water through the reaction $2NO_2 + H_2O \rightarrow HNO_3 + HONO$ (Li et al., 2005b; Szanyi et al., 2004). Prolonged exposure of the sample to $NO+O_2$ led to an increase in intensity only at 1311 cm^{-1}, which can be explained by the reaction $M^{n+} - O^{2-} + 2NO_2 \rightarrow M^{n+} - NO_3^- + NO_2^-$ on cation defect sites such as extra-framework alumina, similar to the reaction pathway suggested by Larsen et al. (Li et al., 2005b). Part of the nitrous species (adsorbed NO_2 and some unidentate nitrates) corresponding to the bands at 1622 and 1311 cm^{-1} were so weakly adsorbed on the zeolite that they disappeared in the subsequent brief evacuation. At the same time, two bands at 1663 and 1389 cm^{-1} due to nitrites and/or nitrates associated with a very low concentration of Na^+ ions in HZSM-5 (Li. et al., 2005b; Satsuma et al., 1997; Szanyi & Paffett, 1996) were observed. Identical to HZSM-5, the band at 1622 cm^{-1} appeared rapidly with $NO+O_2$ co-adsorption on NaZSM-5 (Fig. 12B). However, different from HZSM-5, prolonged exposure of NaZSM-5 to $NO+O_2$ primarily resulted in the appearance of bands with strong intensity at 1407 and 1389 cm^{-1} due to nitrates banding to Na^+ (Li et al., 2005b; Satsuma et al., 1997; Szanyi & Paffett, 1999). It is clear by comparing Fig. 12 A and B that much more stable nitrous species could form on NaZSM-5 than on HZSM-5, which are in good accordance with the results obtained by NO_x-TPD as shown in Fig. 13. Thermal stability of the nitrous species on the zeolites was investigated by FTIR, as shown in Fig. 14. On HZSM-5, the band at 1311 cm^{-1} due to unidentate nitrates disappeared at 150 °C, and a band appeared at 1585 cm^{-1} due to bidentate nitrates (Ivanova et al., 2001; Yu et al., 2004; Shimizu et al., 2001; He et al., 2004; Poignant et al., 2001). It indicates that the bidentate nitrates were produced from unidentate nitrates at the elevated temperature. Compared with the bridging nitrates (1622 cm^{-1}), the bidentate nitrates (1585 cm^{-1}) were more thermally stable; the intensity at 1585 cm^{-1} slightly decreased with the temperature increasing. Similar to those on HZSM-5, unidentate nitrates

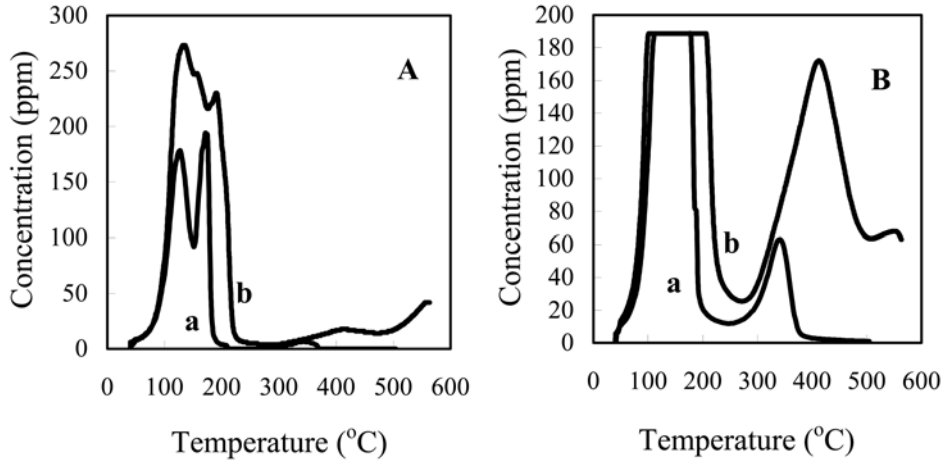


Fig. 13. TPD profiles of NO (A) and NO_2 (B) in N_2 flow after saturated co-adsorption of $NO+O_2$ on HZSM-5 (a) and on NaZSM-5 (b)

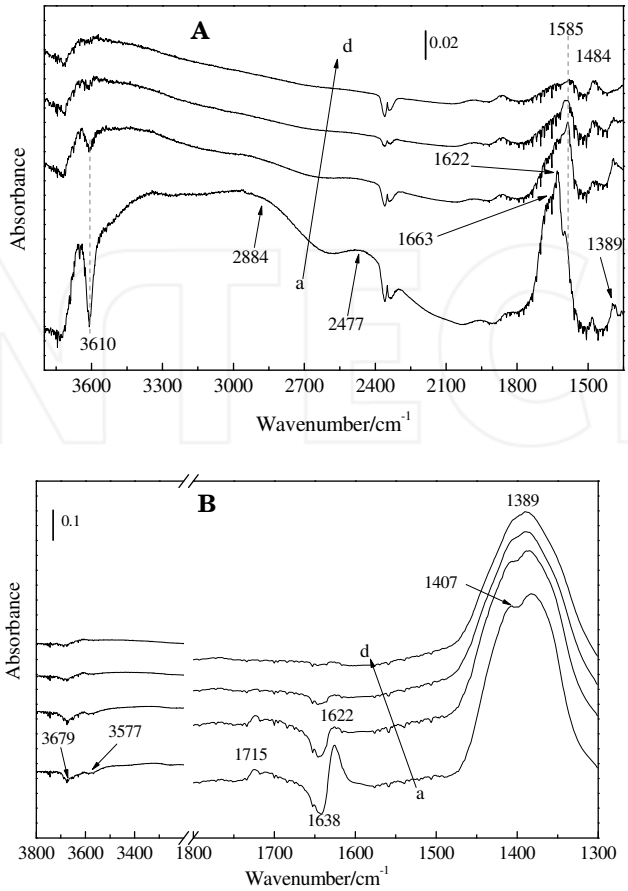


Fig. 14. The nitrous species on HZSM-5 (A) and on NaZSM-5 (B) at 150 °C (a), 250 °C (b), 350 °C (c) and 400 °C (d)

(1311 cm^{-1}) on NaZSM-5 disappeared at 150 °C, indicating that the unidentate nitrate species are less thermal stable than the other kinds of nitrates, irrespective of the adsorbents. It implies that the unidentate nitrates are not involved in the C_2H_2-SCR, as the reaction significantly occurred at the temperature above 250 °C. Quite different from the findings on HZSM-5, although the bridging nitrates (giving band at 1622 cm^{-1}) on NaZSM-5 changed little at 150 °C, most of them desorbed when the temperature increased to 250 °C. Band at 1715 cm^{-1}, observed exclusively on NaZSM-5, could be assigned to nitrates associated with Na^+ ions, because it appeared concomitantly with the band at 1407 cm^{-1} during the co-adsorption of NO+O_2 (Fig. 12 B) and disappeared at 350 °C together with the same band (Fig. 14). On NaZSM-5, the most stable nitrous species are the nitrate species associated with Na^+ (1389 cm^{-1}), which could remained on the zeolite at 400 °C. The FTIR results reveal that the adsorption of NO_x on ZSM-5 was strongly affected by the cations in the zeolite, and that protons were essential for the bidentate nitrates formation on the zeolites.

4.2 Characterization on the activity of nitrous species towards to reduction on the zeolites

The bands due to nitrate species at 1626 and 1585 cm^{-1} disappeared within 1 min upon exposure of HZSM-5 to C_2H_2+O_2 at 250 ◦C (Fig. 15A). Concomitantly, the bands at 1693 cm^{-1} due to carbonyl-containing compound and at 1605 cm^{-1} due to the C=O vibration of carboxylic groups (Li et al., 2005b) appeared. It indicates that the bridging and bidentate nitrates are very active with reductant. However, no significant change in the spectra was observed when the NaZSM-5 was exposed to the gas mixture of C_2H_2+O_2 at 250 ◦C, as shown Fig. 15 B), indicating that the nitrate species bonding to Na^+ on NaZSM-5 are inert to the reductant. The quite different nitrous species formed on HZSM-5 and on NaZSM-5 as well as their disparate reactivity with reductant could explain the distinct catalytic performance of the two zeolites in C_2H_2-SCR. Thus, it is rational to draw the conclusion that protons are responsible not only for catalyzing NO oxidation to NO_2, but also for the formation of active nitrate intermediates in the C_2H_2-SCR.

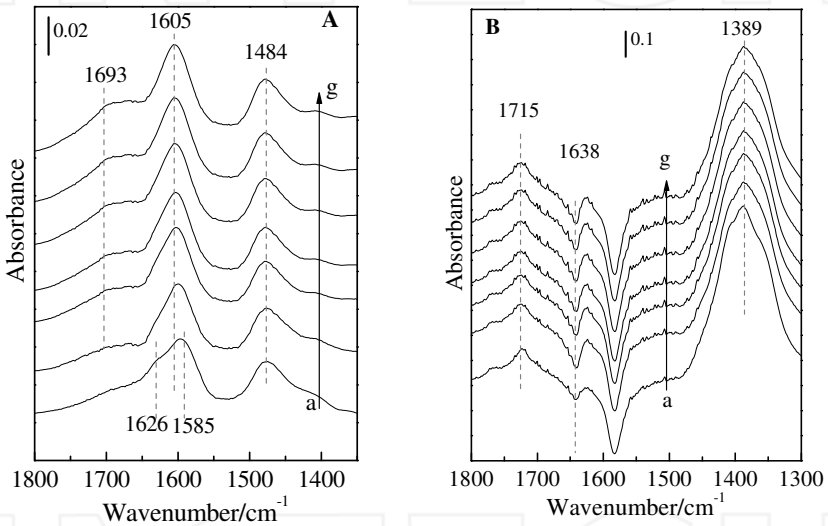


Fig. 15. FTIR spectra of surface species on HZSM-5 (A) and NaZSM-5 (B) at 250 ◦C when the zeolite was subjected to saturated coadsorption of NO+O_2 and a subsequent brief evacuation (a), and then an exposure to C_2H_2+O_2 for 1 min (b), 2 min (c), 4 min(d), 6 min (e), 8 min (f), and 10 min (g)

4.3 Characterization of carbonous species and the reactivity with nitrous species on the zeolites

The evolution of carbonous species and their reactivity toward NO_x were investigated by FTIR. The IR spectra of carbonous species due to saturated adsorption of acetylene on the HZSM-5 and NaZSM-5 at 80 ◦C are depicted in Fig. 16. The adsorption of acetylene on HZSM-5 resulted in the appearance of positive bands at 1673 and 1628 cm^{-1} and negative bands at 3700, 2884, and 2477 cm^{-1} (Fig. 16, spectrum c). As discussed in Section 4.1, the three negative bands are arisen from the consumption of adsorbed water on the Brønsted

acid site. No negative band at 3610 cm^{-1} due to consumption of acidic zeolite Al(OH)Si hydroxyls (Poignant et al., 2001; Brosius et al., 2005) was observed caused by acetylene adsorption. In an energy standpoint, acetylene would adsorb on the free Brønsted acid sites, which were characterized by band at 3610 cm^{-1} as shown in Fig. 17, rather than adsorb on the ones taken up by water. Therefore, it is reasonable to propose that acetylene reacted with the adsorbed water on the Brønsted acid site, but did not simply take up the Brønsted acid site by breaking the well-known A–B–C structure.

Based on above conclusion that acetylene reaction with water on Brønsted acid sites when adsorbed on HZSM-5, it can be further reasonably speculated that vinyl alcohol (CH_2=CH–OH) species were formed by acetylene adsorption on the zeolite as depicted by model 1 as shown in scheme 1: In principle, acetylene can be adsorbed on ZSM-5 in two ways. One way does by reacting with water to form vinyl alcohol and the later can be strongly bonded to Brønsted acid sites with hydrogen bond (model I), being characterized by a blue shift of ν(C=C) with respect to the general C=C double band. Due to hydrogen bond of the –OH group with the Brønsted acid sites, the blue shift of ν(C=C) may be further aggravated, and it is the primary case of acetylene adsorption on HZSM-5, which is strongly supported by the band of 1628 and 1723 cm^{-1}. Another way does by binding to the cations (e.g. H^+ and Na^+) in zeolite channels through weak static attraction to form a π-complex, which may decreases the electron density in the highest occupied molecular orbital, leading to a red shift of ν(C=C) with respect to the general C=C double band (model II). Obviously, it is the primary case of acetylene adsorption on NaZSM-5, which is strongly supported by the band at 1621 cm^{-1}.

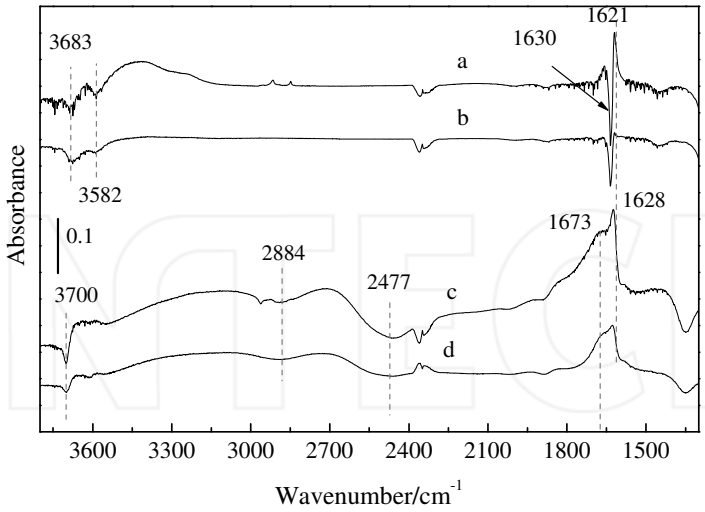


Fig. 16. FTIR spectra of carbonous species formed by saturated adsorption of C_2H_2 (500 ppm) in N_2 at 80 ^oC on NaZSM-5 (a), HZSM-5 (c), and those when the sample of NaZSM-5 (b) and HZSM-5 (d) was subsequently subjected to a brief evacuation at this temperature

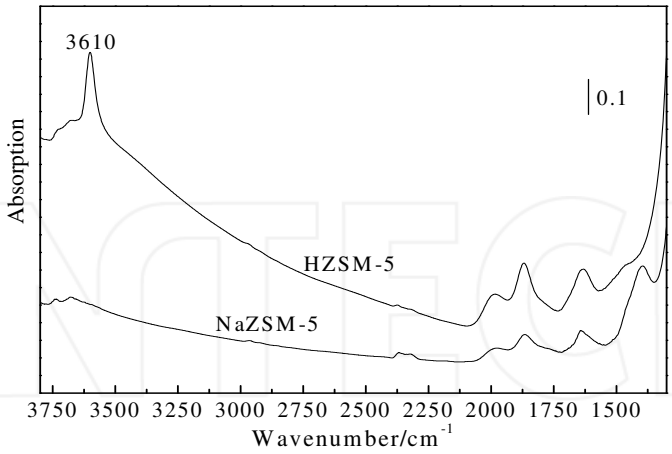
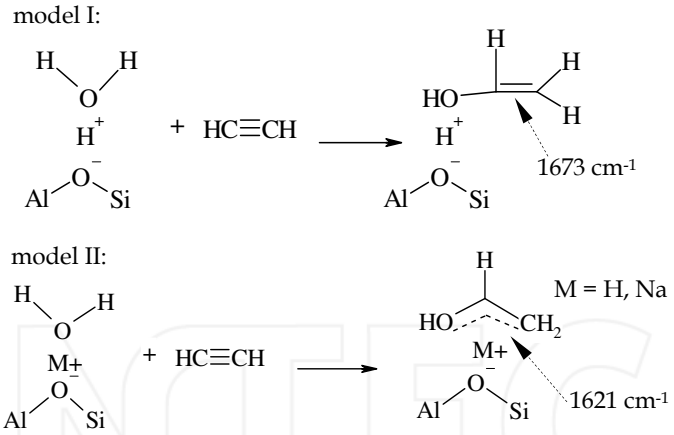


Fig. 17. FTIR spectra of fresh HZSM-5 and NaZSM-5



Scheme 1. C_2H_2 adsorption model on the HZSM-5 and NaZSM-5

Evolution of the carbonous species on HZSM-5 with temperature is shown in Fig. 18. It is obvious that the intensity of the band at 3659 cm^{-1} due to ν(–OH) of alcohol hydroxyl groups increased with the temperature increase from 80 to 150 °C. At the same time, the bands at 2884 and 2477 cm^{-1} due to the A–B–C structure, and the band at 1352 cm^{-1} due to δ(OH) of the H-bonded zeolite hydroxyls were recovered on the HZSM-5 (Fig. 18), accompanied by the appearance of the negative band at 3610 cm^{-1}. With the temperature further increase, the bands relating to adsorbed water on the Brønsted acid sites increased in

intensity. It indicates that water was formed and adsorbed on the Brønsted acid sites at elevated temperatures. Above 300 °C, the strong band centered at 1628 cm^{-1} disappeared, which correlates well with the end of desorption peak (around 215 °C) of acetylene in C_2H_2-TPD (not shown). As a result, a new band at 1646 cm^{-1} was clearly observed. Based on these results, a scheme 2 for the evolution of carbonous species with temperature on HZSM-5 can be proposed.

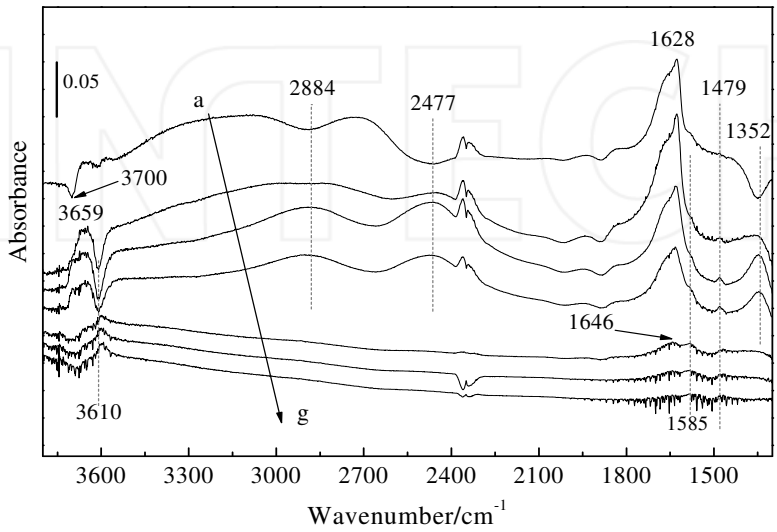
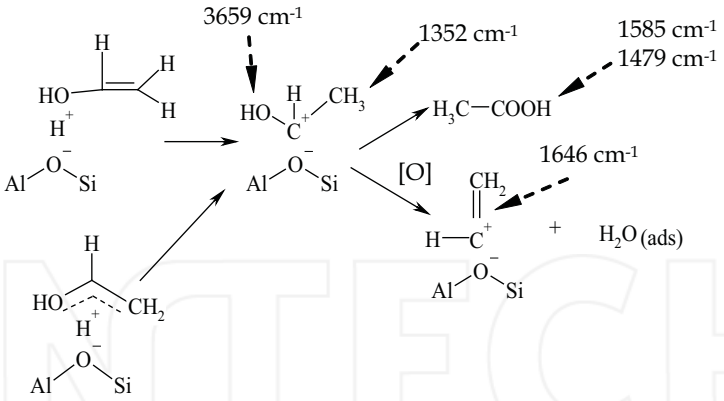


Fig. 18. Spectra of carbonous species on HZSM-5 when evacuated at 80 °C (a), at 150 °C (b), at 178 °C (c), (d) at 200 °C (d), at 300 °C (e), at 360 °C (f), and at 410 °C (g) after saturated adsorption of acetylene

The vinyl alcohol adsorbed on the Brønsted acid sites in the models both I and II was converted to a carbenium ion ($HO–HC^+–CH_3$), which gave bands at 3659 and 1352 cm^{-1} due to the OH stretching vibration and the CH_3 deformation vibration, respectively. The carbenium ion decomposed to water and another carbonium ion ($CH_2=HC^+$) that gave the band at 1646 cm^{-1}. The formed water adsorbed on Brønsted acid sites led to a recovery of intensity and an increase of the bands at 2884 and 2477 cm^{-1}, as well as the negative band at 3610 cm^{-1}. In addition, some of the carbenium ion ($HO–HC^+–CH_3$) was oxidized by oxygen (ca. 5000 ppm) in the pure N_2 (99.995%) to acetate species, giving the bands at 1585 and 1479 cm^{-1} due to the νas(COO) and νs(COO) of acetate species (Yu et al., 2004; Shimizu et al., 2001; He et al., 2004; Poignant et al., 2001; Wu et al., 2005).

Evolution of the carbonous species on HZSM-5 with temperature is shown in Fig. 19. Temperature increase from 80 to 200 °C resulted in recovery of the band at 1635 cm^{-1} as well as a shift of the band (from 1623 to 1628 cm^{-1}). Scheme 3 explains that acetylene was released by the decomposition of vinyl alcohol, leaving water on the cation sites in zeolite, leading to the disappearance of the negative band. Above 300 °C, almost no carbonous surface species were detected on Na-ZSM-5, which is in line with the results obtained in C_2H_2-TPD (not shown).



Scheme 2. The evolution of carbonous species on HZSM-5

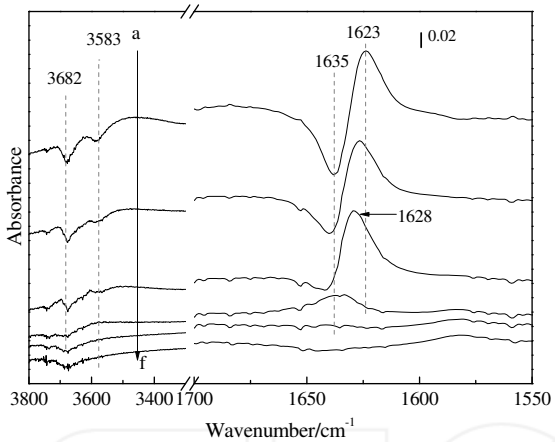
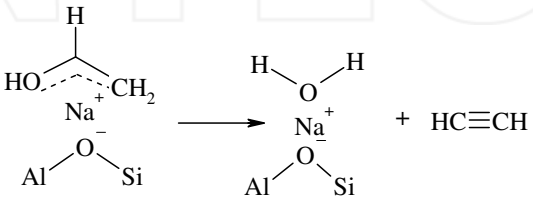


Fig. 19. Spectra of carbonous species on NaZSM-5 when evacuated at 150 °C (a), at 178 °C (b), at 200 °C (c), (d) at 300 °C (d), at 360 °C (e), and at 420 °C (f), after saturated adsorption of acetylene



Scheme 3. The decomposition of carbonous species on NaZSM-5

The reactivity of carbonous species on HZSM-5 and on NaZSM-5 with NO and NO+O_2 was studied at 200 °C. Acetate species (1585, 1479 cm^{-1}) were detected at steady state on HZSM-5 in NO/N_2 and N_2 (Fig. 20 A), indicating that the acetate species could exist in NO/N_2. This means that acetate species did not react with NO at the temperature. However, in NO+O_2/N_2, no acetate species could be detected, and instead, a new band at 2131 cm^{-1} appeared. This band can be assigned to ν(C≡N) of cyanide (Shimizu et al., 2001; Mosqueda-Jiménez et al., 2003b) produced by the reaction of acetate species with NO_x. The result indicates that the acetate species formed by carbenium ion (HO–HC^+–CH_3) oxidation on HZSM-5 are active with nitrous species arising from the co-adsorption of NO+O_2. Thus it can be reasonably considered that acetate species is one of the important intermediates for the C_2H_2-SCR. In contrast to the case of HZSM-5, although vinyl alcohol (1628 cm^{-1}) was also formed on NaZSM-5 (Fig. 20 B), no significant change of the band in intensity was observed when the gas mixture was switched form NO/N_2 to NO+O_2/N_2, indicating that the vinyl alcohol bonding to Na^+ in NaZSM-5 is inert under the reaction conditions.
Based on above discussion, it is clear that, the carbonous species formed by acetylene adsorption on HZSM-5 and on NaZSM-5 as well as their reactivity with nitrous species are also quite different.

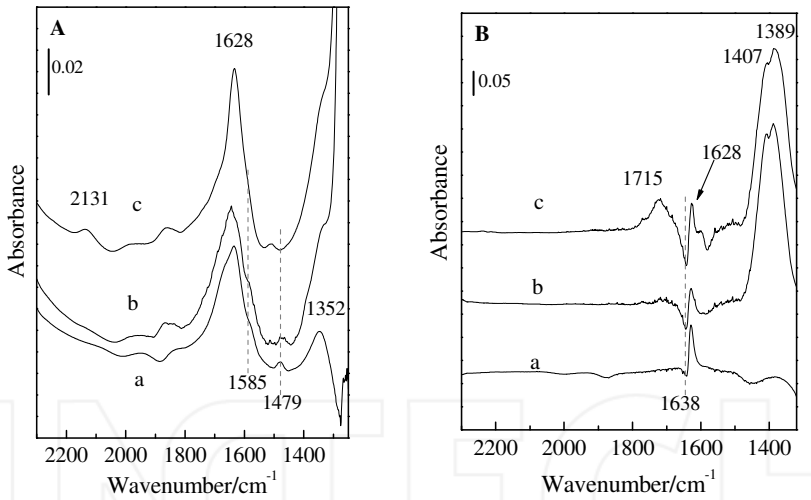


Fig. 20. FTIR spectra of surface species on HZSM-5 (A) and NaZSM-5 (B) at 200 °C taken in (a) N_2, (b) NO, (c) NO + O_2. Before the measurements, the catalysts were pretreated in C_2H_2/N_2 at 80 °C

5. To propose possible reaction mechanism

To design a catalyst more active for the HC-SCR in real lean-burn conditions, extensive studies were also carried out on mechanism of the reaction. There have been different opinions concerning the mechanism of HC-SCR in literature, which can be roughly classified as "dissociative" (Goula et al., 2007; Burch & Watling, 1996) and "reduction" ones

(Mihaylov et al., 2004; Mosqueda-Jiménez et al., 2003b; Hadjiivanov et al., 2003). The "dissociative" mechanism proposed by Burch and Watling (Burch & Watling, 1996) has been widely accepted by the authors being concerned with noble metal catalysts in the HC-SCR field, and it can be expressed as follows (Goula et al., 2007):

$$NO_{ads} \rightarrow N_{ads} + O_{ads} \quad (1)$$

$$N_{ads} + N_{ads} \rightarrow N_2(g) \quad (2)$$

$$N_{ads} + NO_{ads} \rightarrow N_2O(g) \quad (3)$$

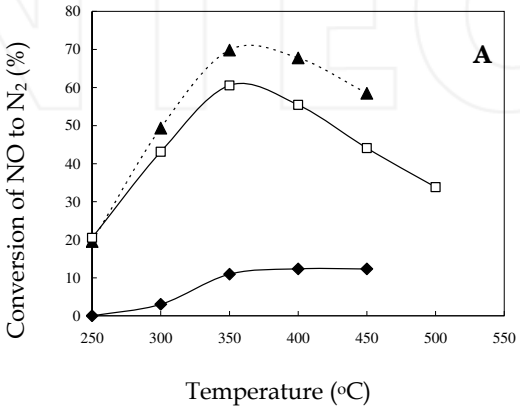
For the "reduction" mechanism of HC-SCR, some authors have claimed that activation of hydrocarbon occurs first, and some partially oxidized hydrocarbons ($C_xH_yO_z$) produced by the step then react with NO and/or NO_2 to form the secondary intermediates (Sasaki et al., 1992). For example, formate and acetate were proposed to be active species of HC-SCR over CoO_x/Al_2O_3 (He & KÖhler, 2006), Ga_2O_3/Al_2O_3 (He et al., 2005), Ag/Al_2O_3 (Shibata et al., 2003), SnO_2/Al_2O_3 (Liu et al., 2006), $Cu\text{-}Al_2O_3$ (Shimizu et al., 2000; Satsuma & Shimizu, 2003), In_2O_3/Al_2O_3 (Luo et al., 2007), Ag/Al_2O_3 (Zhan et al., 2007) and Pd/Al_2O_3 catalysts (Huuhtanen et al., 2002). Acetaldehyde deriving from propene was also proposed to be main active species of the HC-SCR over sulfated titania-supported rhodium catalyst (Flores-Moreno et al., 2005). However, some authors have claimed that activation of NO_x occurs first, forming nitrous surface species, such as nitro (Meunier et al., 2000), nitroso (Poignant et al., 2001, Gerlach et al., 1999), nitrosonium ions (Gerlach et al., 1999, Ingelsten et al., 2005), nitrate or nitrite (Luo et al.,2007, Anunziata et al., 2007) over the catalyst. For instance, bridging and bidentate nitrates were reported to be produced first by co-adsorption of $NO+O_2$ on $Co/SO_4^{2-}\text{-}ZrO_2$ (Tsyntsarski et al., 2003), BaY (Sedlmair et al., 2003a) and Ag/Al_2O_3 (Bentrup et al., 2005). Tsyntsarski et al. (Tsyntsarski et al., 2003) have suggested that both bridging and bidentate nitrates are active species of the HC-SCR. Mihaylov et al. (Mihaylov et al., 2004) have reported that monodentate nitrates on Ni-HZSM-5 are highly reactive towards methane. Lónyi et al. (Lónyi et al., 2007) have studied selective catalytic reduction of NO by CH_4 over Co-, Co,Pt-, and H-mordenite catalysts and suggested that nitrosonium ions are surface intermediates of the reaction. There are also some other suggestions about the reaction intermediate of HC-SCR in literature, including nitrile (Poignant et al., 2001; Delahay et al., 2007), isocyanate (Mihaylov et al., 2004; Mosqueda-Jiménez et al., 2004b; He & KÖhler, 2006), R-NO_x (Mosqueda-Jiménez et al.,2003; Cowan et al., 1998), amine (Poignant et al., 2001), acetonoxime (Shimizu et al.,2000; Resini et al., 2003) and ammonia (Lónyi et al., 2007).

Although most of the "reduction" mechanisms were supported by Fourier transform infrared (FTIR) identification of reaction intermediates (Joubert et al., 2006), none of them has been widely accepted because of the complexity of the process (Mosqueda-Jiménez et al., 2003b) being concerned with different catalysts, reductants and reaction conditions. More investigation on the reaction mechanism is required for understanding the real reaction route of HC-SCR over different catalysts and using different reductants.

5.1 Reaction mechanism of C_2H_2-SCR over Mo/HMOR

In the this section, a possible reaction mechanism of C_2H_2-SCR over Mo/ HMOR （Li et al., 2008） and Zr/HFER (Xing et al., 2008) investigated by in situ FTIR was summarized.

Fig 21 shows conversion of NO and C_2H_2 in C_2H_2-SCR over the mordenite-based catalysts (HMOR, 0.5%Mo/HMOR, NaMOR) as a function of temperature. The activity of 0.5%Mo/HMOR for C_2H_2-SCR was considerably high compared to HMOR and NaMOR. 70% of NO conversion to N_2 at 350 °C over 0.5% Mo/HMOR catalyst was obtained. It indicates that molybdenum has a significant promotional effect on C_2H_2-SCR. The peak of the "volcano" curve in NO conversion to N_2 versus reaction temperature seems to be in line with the temperature where C_2H_2 was nearly completely consumed. Hence, the drop of NO conversion above 350 °C can be considered to be caused by the lack of reductant.



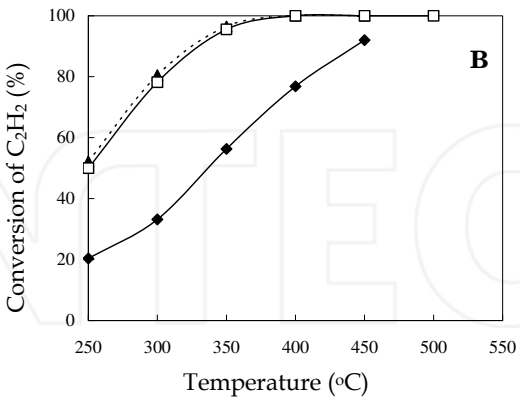


Fig. 21. Conversion of NO (A) and C_2H_2 (B) as a function of reaction temperature in C_2H_2-SCR. Reaction condition: 1600 ppm NO, 800 ppm C_2H_2, 9.95 % O_2 in He with a total flow rate of 50 ml/min over HMOR 0.100 g (□), 0.5 % Mo/HMOR 0.100 g (▲) or NaMOR 0.200 g (♦)

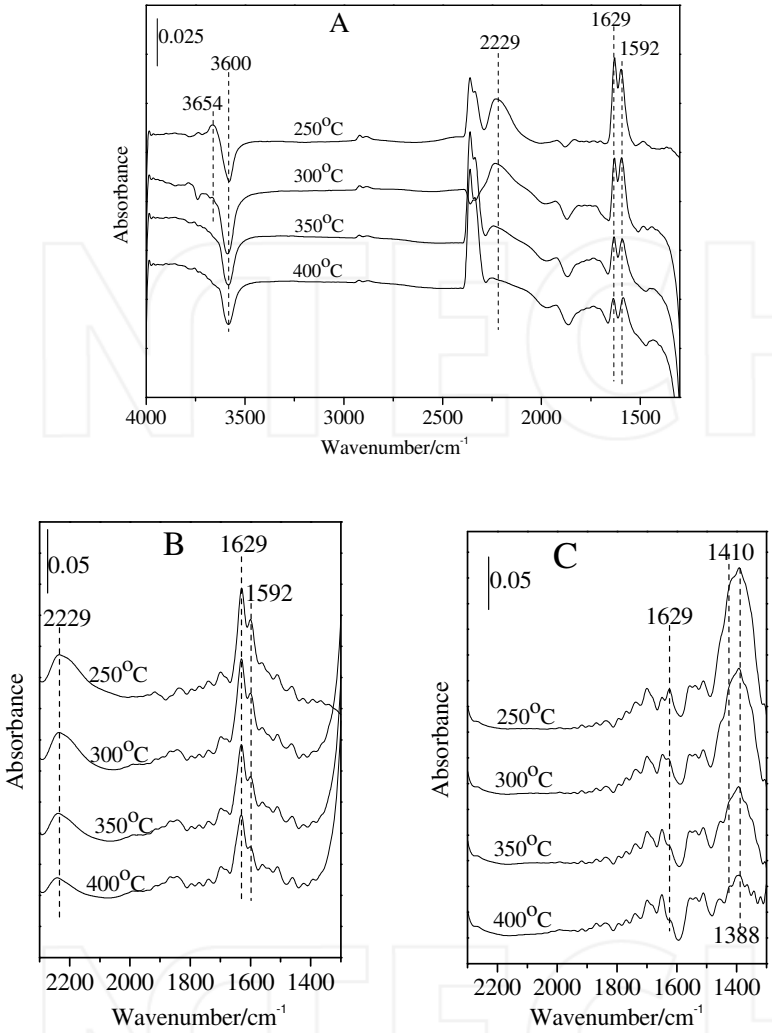


Fig. 22. Steady state in situ FTIR spectra of surface species on HMOR (A), 0.5%Mo/HMOR (B) and NaMOR (C) in 1000 ppm NO + 10% O_2 + N_2 at different temperature

Figure 22 shows steady state in situ FTIR spectra of nitric species formed by NO+O_2 co-adsorption on the mordenite-based catalysts at different temperatures. Three main bands at 2229, 1629 and 1592 cm^{-1} associated with nitric species were observed on HMOR (Fig. 22 A). The band at 2229 cm^{-1} is due to N-O stretching mode in NO^+ (Li et al., 2005a; Pirngruber & Pieterse, 2006; Gerlach et al., 1999), and the bands at 1629 and 1592 cm^{-1} can be assigned to bridging and bidentate nitrates (Poignant et al., 2001; Li et al., 2005a; Sedlmair et al., 2003; Yu et al., 2007), respectively. The NO^+ (2229 cm^{-1}) species was also detected by Gerlach et al. (Gerlach et al., 1999) at 120 °C when NO_x was adsorbed on the zeolite. As shown in Fig. 22

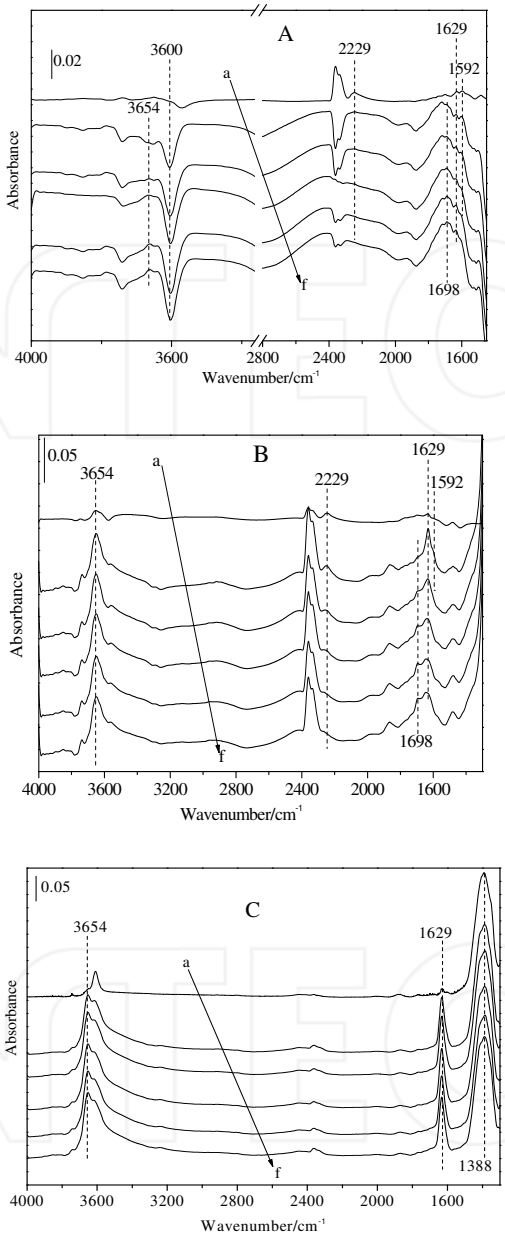


Fig. 23. In situ FTIR spectra on HMOR (A), 0.5%Mo/HMOR (B) and NaMOR (C) at 250 °C: a brief evacuation after saturated adsorption of $NO+O_2$ (a), and subsequently exposing to $C_2H_2+O_2$ for: 1 min (b), 3 min (c), 5 min (d), 8 min (e), 30 min (f)

A, a positive band at 3654 cm^{-1} due to adsorbed water (Mihaylov et al., 2004; Chafik et al., 1998) and a negative band at 3600 cm^{-1} due to Brønsted acid sites (Mihaylov et al., 2004; Gutierrez et al.,2007) were observed after co-adsorption of NO and O_2 on the zeolite at 250 ^oC. It can be well interpreted by the NO^+ formation pathway $NO + NO_2 + 2H^+ \rightarrow 2NO^+ + H_2O$, proposed by Hadjiivanov et al. (Hadajiivanov et al., 1998) and Gerlach et al. (Gerlach et al., 1999). Band at 1629 cm^{-1} due to bridging nitrate and band at 2229 cm^{-1} due to NO^+ on 0.5%Mo/HMOR are obviously greater in intensity respectively compared with those on HMOR, particularly above 300 ^oC (Fig. 22 B). It indicates that molybdenum loading on the HMOR zeolite have a promotional effect on the nitric species formation at higher temperature. On NaMOR (Fig. 22 C), band (1629 cm^{-1}) due to this type of bridging nitrate was rather weak, and bands due to bidentate nitrates (1592 cm^{-1}) and NO^+ species (2229 cm^{-1}) even could not be observed. Instead, a broad band at 1410-1388 cm^{-1} due to nitrate ions attached to Na^+ sites (Mihaylov et al., 2004; Li et al., 2005a; Yu et al., 2007) appeared after $NO+O_2$ co-adsorption on the sample under the same condition.

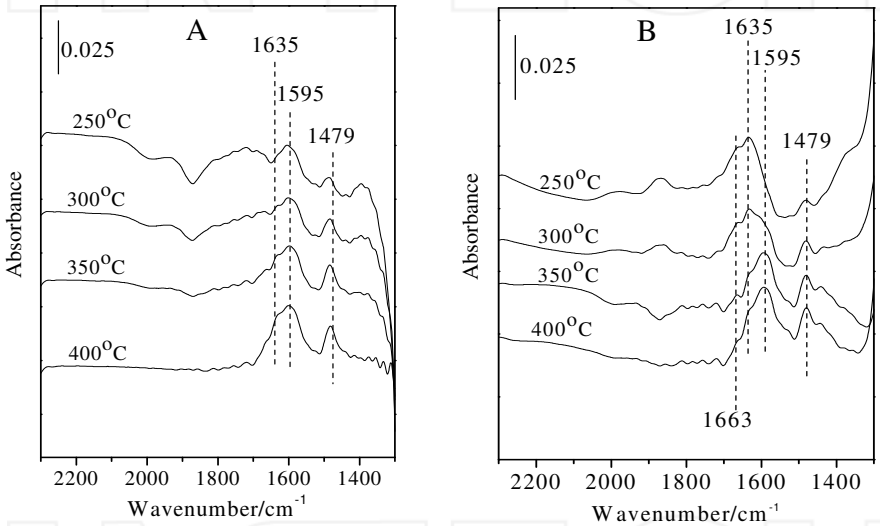


Fig. 24. Steady state in situ FTIR spectra of adsorbed species in 500 ppm C_2H_2 + 10 % O_2 + N_2 on HMOR (A), 0.5 % Mo/HMOR (B) at different temperature

Reactivity of the nitric species towards $C_2H_2+O_2$ over the mordenite-based catalysts was examined by in situ FTIR at 250 ^oC (Fig. 23). When $C_2H_2+O_2$ was introduced into the FTIR cell, bands due to bidentate nitrates (1592 cm^{-1}) and NO^+ species (2229 cm^{-1}) on HMOR arisen from $NO+O_2$ pre-adsorption (Fig. 23A) rapidly decreased. Concomitantly, a new band at 1698 cm^{-1} appeared and reached its maximum intensity within 3 min with disappearance of bands at 2229 and 1592 cm^{-1}. Similar result was obtained when C_2H_2 was used instead of $C_2H_2+O_2$ in the above experiment. The results indicate that NO^+ and bidentate nitrate species are fairly reactive towards acetylene at this temperature. It was evidenced by the following changes of bands associated with water formation during the process: A positive band at 3654 cm^{-1} due to water appeared, and at the same time, a negative band at 3600 cm^{-1} arisen from water adsorption on Brønsted acid sites

correspondingly increased in intensity. Unfortunately, the corresponding reactivity of bridging nitrate species (1629 cm^{-1}) could not be directly evaluated on the zeolite because of water formation. The band due to bending mode of water appears at the identical wave number with that of bridging nitrate species at 1629 cm^{-1}. Similar experimental results as that on HMOR was obtained on 0.5%Mo/HMOR (Fig. 23 B). However, quite different results were obtained on NaMOR. The nitrate species attached to Na^+ seem to be completely inert towards the reactant. No change in intensity of band at 1388 cm^{-1} due to the species could be observed on NaMOR (Fig. 23 C). Meanwhile, as expected, band at 1698 cm^{-1} did not appear on NaMOR. Instead, strong bands at 3654 and 1629 cm^{-1} due to water adsorbed on the Na-form zeolite were observed, which may be simply resulted from combustion of acetylene.

Figure 24 shows steady state in situ FTIR spectra of the surface species on HMOR and on 0.5%Mo/HMOR in gas mixture of 500 ppm C_2H_2 + 10% O_2/N_2 at different temperatures. No band at 1698 cm^{-1} could be observed on the catalysts. Instead, bands at 1635, 1595 and 1479 cm^{-1} due to ν(C=O), ν_{as}(COO) and ν_s (COO) of carboxylic groups (Hadajiivanov et al., 1998; Shimizu et al., 2007) appeared. Combined the result with that observed in Fig. 23, the band at 1698 cm^{-1} can be attributed to nitrogen containing organic species, because this band could appear only in the following conditions: both the nitric species (nitrogen oxides and/or nitric surface species) and the reductant were present together in the reaction system.

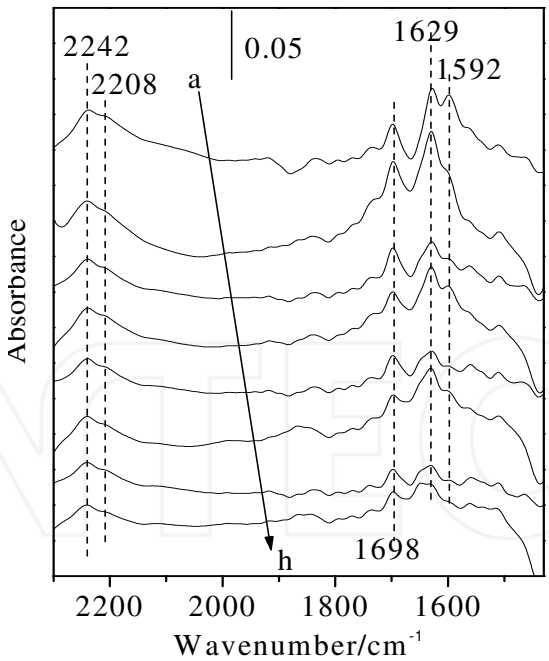


Fig. 25. Steady state in situ FTIR spectra of adsorbed species on 0.5%Mo/HMOR in gas mixture of 1000 ppm NO + 500 ppm C_2H_2 + 10 % O_2 + N_2 at 250 °C (a), 300 °C (c), 350 °C (e), 400 °C (g) and that on HMOR at 250 °C (b), 300 °C (d), 350 °C (f), 400 °C (h)

Figure 25 shows the steady state in situ FTIR reaction spectra of the surface species on HMOR and 0.5%Mo/HMOR in gas mixture of 1000 ppm NO + 500 ppm C_2H_2 + 10% O_2/N_2 at different temperatures. Two overlapped bands respectively centered at 2242 and 2208 cm^{-1} were appeared in the spectra, in addition to the bands at 1698, 1629 and 1592 cm^{-1}. The band at 2242 cm^{-1} can be assigned to -NCO vibration of isocyanate (2242 cm^{-1}) (Satsuma & Shimizu, 2003; Kameoka et al., 2000) and that at 2208 cm^{-1} can be assigned to N-O stretching of NO^+ (Pirngruber & Pieterse, 2006; Gerlach et al., 1999). It was reported that the stretching frequency of NO^+ on HMOR is influenced by an interaction between NO^+ and some other surface species formed on the zeolite (Gerlach et al., 1999). Herein, it should be noticed that although the band at 1698 cm^{-1} for 0.5%Mo/HMOR (spectrum a) was slightly weaker in intensity in comparison with that of HMOR (spectrum b) at 250 °C, it became much stronger than that of HMOR above 300 °C. The relative intensity of the band observed on HMOR and 0.5%Mo/HMOR at different temperatures could be correlated well with the relative activity of the catalysts for C_2H_2-SCR (Fig. 21). Good accordance of NO reduction with the band at 1698 cm^{-1} could also be obtained on NaMOR (Fig 26). The band at 1698 cm^{-1} just appeared at the temperature (spectrum e), where NO conversion to N_2 became significant over the zeolite in C_2H_2-SCR. Based on the above findings, we believe that the species with the band at 1698 cm^{-1} is a crucial intermediate for C_2H_2-SCR over the mordenite-based catalysts.

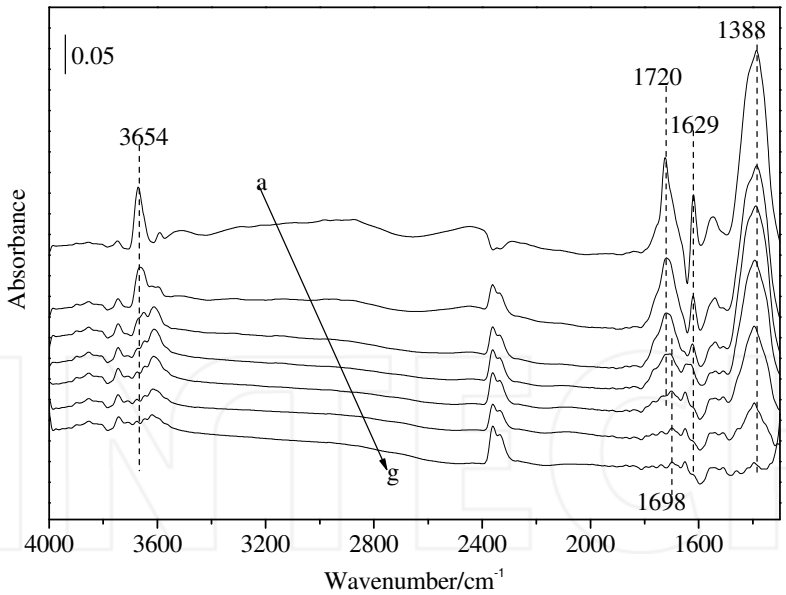


Fig. 26. Steady state in situ FTIR spectra of adsorbed species in gas mixture of 1000 ppm NO + 500 ppm C_2H_2 + 10% O_2 + N_2 on NaMOR at different temperature: 150 °C (a), 200 °C (b), 250 °C (c), 300 °C (d), 350 °C (e), 400 °C (f), 450 °C (g)

To further study the reaction route of C_2H_2-SCR over the catalysts, reactivity of the intermediate (1698 cm^{-1}) was investigated. After a pre-exposure of HMOR to 1000 ppm NO

+ 500 ppm C_2H_2 + 10% O_2/N_2 at 250 °C and a followed brief evacuation, the catalyst was exposed to gas mixture of 1000 ppm NO + 10% O_2/N_2. As a result, the intensity of the band at 1698 cm^{-1} rapidly decreased (Fig. 27, spectrum a). It indicates that the intermediate is rather reactive towards NO+O_2 at the temperature. On the other hand, no decrease in intensity of the band at 2242 cm^{-1} due to -NCO species could be observed during the period (Fig. 27), indicating that -NCO species is inert towards NO+O_2. When the sample was then exposed to gas mixture of 500 ppm C_2H_2 + 10% O_2 in N_2, as shown in Fig. 28 (c~f), the bands both at 2242 cm^{-1} due to -NCO and at 2229 cm^{-1} due to NO^+ disappeared within 1 min. Concomitantly, a band at 1698 cm^{-1} and bands at 3654 and 1629 cm^{-1} due to adsorbed water appeared. It should be noticed that the intensity of the three bands as well as that of the negative band at 3600 cm^{-1} continued to increase within three minutes. The result agrees well with the proposition in literature that isocyanate species can be easily hydrolyzed to amines (Poignant et al., 2001; Bion et al., 2003). Thus, the band at 1698 cm^{-1} can be assigned to acid amide species on the zeolite. It is in accordance with that reported by Poignant et al. (Poignant et al., 2001), who found the species with band at 1694 cm^{-1} in reaction of NO + C_3H_8 + O_2 over HZSM-5 at 350 °C.

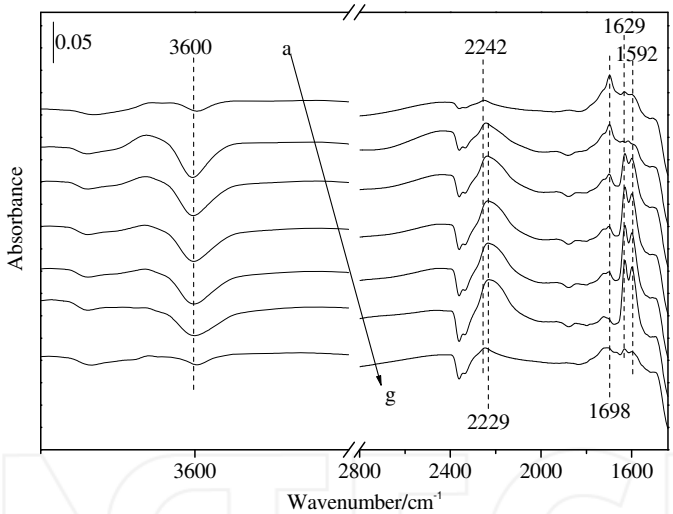


Fig. 27. In situ FTIR spectra of surface species on HMOR at 250 °C: a brief evacuation after a pre-exposure of the catalyst to NO+C_2H_2+O_2 for 30 min (a), and then when the catalyst was exposed to NO+O_2 for: 1 min (b), 3 min (c), 5 min (d), 8 min (e), 30 min (f), and finally evacuated briefly (g)

In Fig. 22, we showed that molybdenum loading on HMOR zeolite considerably promoted the formation of bridging nitrate species. However, neither higher NO conversion to N_2 in C_2H_2-SCR nor stronger band due to the acid amide species (1698 cm^{-1}) could be observed on 0.5%Mo/HMOR compared to HMOR at 250 °C. It leads us to speculate that bridging nitrate species (1629 cm^{-1}) may make no contribution to C_2H_2-SCR at the lower temperature (< 250 °C) over the mordenite-based catalysts. The speculation was validated by the following experimental results: Although the bridging nitrate species (1629 cm^{-1}) was detected by FTIR

after co-adsorption of $NO+O_2$ on NaMOR at 250 °C (Fig. 22C, spectrum a), no NO conversion to N_2 could be obtained at the temperature (Fig. 21).

Above 300 °C, both C_2H_2-SCR activity (Fig. 21) and the intensity of the band due to acid amide species detected by FTIR (1698 cm^{-1}, in Fig. 25) were larger on 0.5%Mo/HMOR compared to those on HMOR, which corresponds well with the larger population of bridging nitrate species (1629 cm^{-1}, in Fig. 22) given by 0.5%Mo/HMOR in comparison with those given by HMOR at the temperature. The results indicate that bridging nitrate species (1629 cm^{-1}) are also involved in C_2H_2-SCR at higher temperature. Same conclusion can be drawn on NaMOR. The band at 1629 cm^{-1} due to bridging nitrate species arisen from

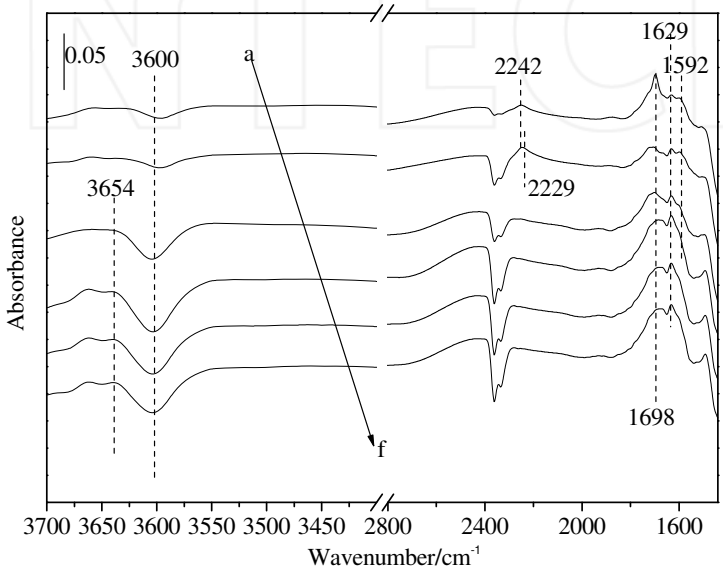


Fig. 28. FTIR spectra of the surface species on HMOR at 250 °C: a brief evacuation after exposing the catalyst to $NO+C_2H_2+O_2$ for 30 min (a), a brief evacuation after exposing the catalyst to $NO+O_2$ for 30 min (b), then exposed the catalyst to $C_2H_2+O_2$ for 1 min (c), 3 min (d), 5 min (e), 30 min (f)

$NO+O_2$ pre-adsorption on the zeolite at 350 °C (Fig. 29, spectrum a) rapidly disappeared when gas mixture of 500 ppm C_2H_2 + 10% O_2/N_2 was introduced to the FTIR cell at this temperature. Correspondingly, band at 1698 cm^{-1} due to acid amide species appeared (Fig. 29, spectrum b). No change of the band at 1388 cm^{-1} in intensity could be observed during this procedure. The results again indicate that bridging nitrate species are involved in the desired reaction at the higher temperature, whereas nitrate species attached to Na^+ have no contribution to C_2H_2-SCR under the reaction condition.

It explains why the C_2H_2-SCR activity of NaMOR became significant when the reaction temperature increased to 350 °C. Thus, the considerably larger activity of 0.5%Mo/HMOR in comparison with that of HMOR for C2H2-SCR above 300 oC (Fig. 21) can be rationally attributed to the larger bridging nitrate formation capacity of the catalyst.

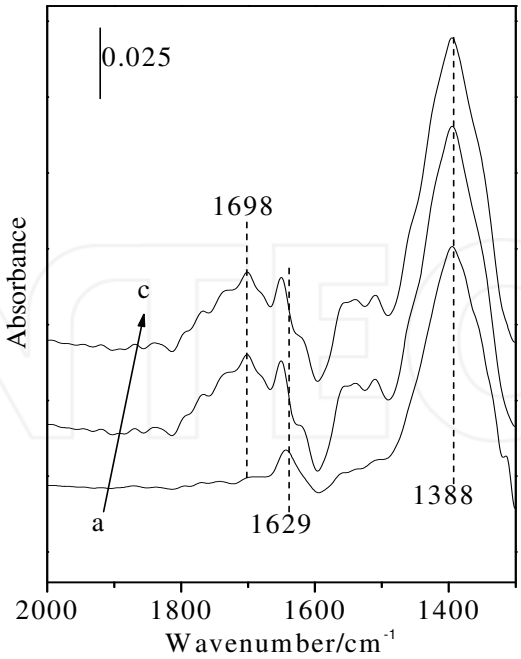
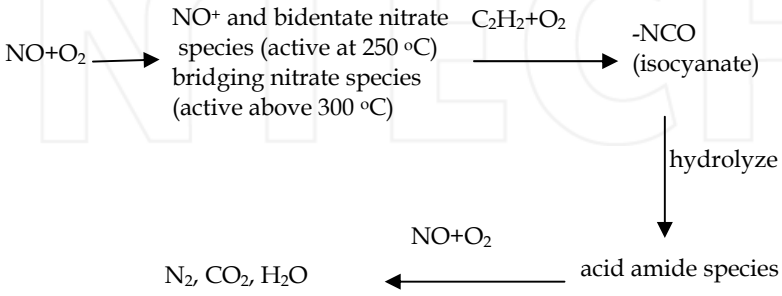


Fig. 29. In situ FTIR spectra of surface species on NaMOR at 350 °C: The catalyst was subjected a brief evacuation after saturation adsorption of 1000 ppm NO and 10 % O_2 in N_2 (a), and then exposed to 500 ppm C_2H_2 + 10 % O_2 / N_2 for 1min (b), 10min (c)

On the basis of above discussion, a possible reaction mechanism of C_2H_2-SCR over the mordenite-based catalysts can be outlined in Scheme 4. Nitrosonium ions (NO^+), bidentate and bridging nitrate species formed by NO+O_2 co-adsorption react first with C_2H_2, leading to isocyanate species (2242 cm^{-1}) formation. The isocyanate species is then rapidly hydrolyzed to the acid amide species (1698 cm^{-1}) that is a crucial intermediate for C_2H_2-SCR over the mordenite-based catalysts.



Scheme 4. A possible reaction pathway of the C_2H_2-SCR over the mordenite-based catalysts

As a summarization for C_2H_2-SCR over the mordenite-based catalysts, following reaction mechanism can be drawn: The nitric species, including nitrosonium ions (NO^+) and bidentate nitrate, are fairly active towards the desired reduction in the temperature range of 250-450 °C. However, bridging nitrate species begin to make its significant contribution to the reaction above 300 °C. Molybdenum incorporated into HMOR zeolite considerably improved the bridging nitrate formation capacity of the catalyst. It explains the promotional effect of molybdenum on C_2H_2-SCR at the higher temperatures. Isocyanate (-NCO), as an active species produced from the reaction of the nitric species with C_2H_2, can be rapidly hydrolyzed to the acid amide species (1698 cm^{-1}) that is a crucial intermediate for C_2H_2-SCR over the mordenite-based catalysts.

5.2 Reaction mechanism of C_2H_2-SCR over Zr/HFER

Surface species formed on 2%Zr/FER by exposing the catalyst to $NO+O_2/N_2$ at 250 °C for 30 min and subsequently to C_2H_2/N_2 characterized by FTIR spectra are shown in Fig. 30. NO^+, bridging and bidentate nitrate species, which give the bands at 2188, 1629 and 1598 cm^{-1} (Poignant et al., 2001; Li et al., 2007; Li et al., 2005a; Brosius et al., 2005; Yu et al., 2007) respectively, were presented by the co-adsorption of $NO+O_2$ in N_2 (spectrum a). The three

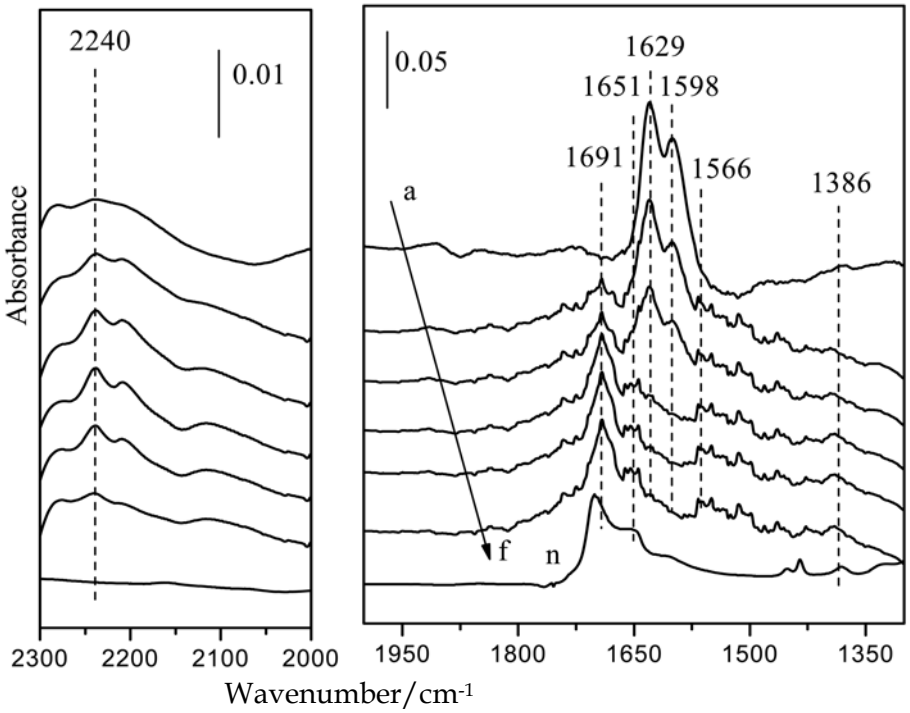
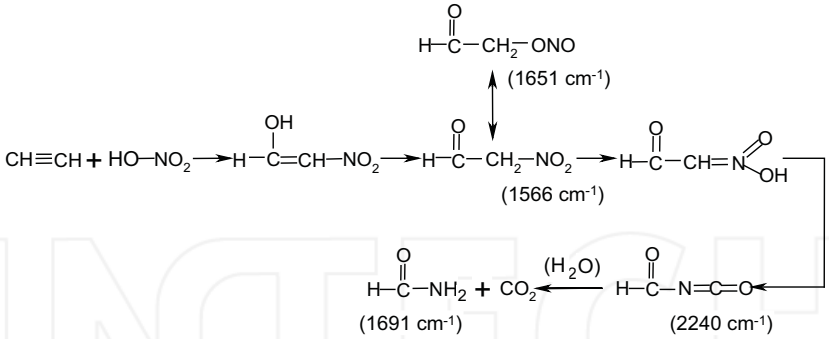


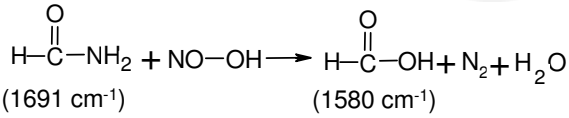
Fig. 30. FTIR spectra of surface species on 2%Zr/FER: the catalyst was exposed to 1000 ppm NO + 10 % O2 + N_2 at 250 °C for 30 min (a), and subsequently to 500 ppm C_2H_2 + N_2 for 1 min (b), 2 min (c), 3 min (d), 5 min (e) and 10 min (f); a adsorption steady state got by exposing the fresh catalyst to formamide vapor in N_2 at 150 °C (g)

bands rapidly decreased in intensity with an appearance of bands at 2240, 1691, 1651, 1580, 1566 and 1386 cm^{-1} when the gas mixture was switched to $C_2H_2+N_2$. Apparently, the bands at 1651 and 1566 cm^{-1} due to -ONO and -NO_2 vibrations [Yeom et al., 2006] increased with time, along with the band at 1691 cm^{-1}. The weak band at 2240 cm^{-1} due to -N=C=O vibration of isocyanate species (Liu et al., 2006) increased in intensity during the first several minutes, and decreased then after, and finally disappeared when the band at 1691 cm^{-1} reached its maximum in intensity. The intensity change of bands at 2240 and 1691 cm^{-1} quite seems that the species giving the band at 1691 cm^{-1} is produced by the isocyanate species.

In literature, Poignant et al. have found the band at 1694 cm^{-1} during the reaction of $NO+C_3H_8+O_2$ over HZSM-5 at 350 ^oC (Poignant et al., 2001) and assigned the band to acetamide species. Larrubia et al. also found a band at 1690 cm^{-1} in FTIR on Fe_2O_3-TiO_2 when they exposed the catalyst to acetamide vapor at 350 ^oC (Larrubia et al., 2001). However, no band at around 1690 cm^{-1}, but at 1717 cm^{-1} was found when we exposed the 2%Zr/HFER catalyst sample to acetamide vapor in N_2 in the temperature range of 30-300 ^oC. Instead, a spectrum that much close to spectrum f (in fig. 30) in shape, with the bands at 1691 and 1386 cm^{-1} along with 1651 cm^{-1} (-ONO vibration) was obtained when the catalyst was exposed to formamide vapor (spectrum g). In addition, the number of carbon in the amine species is also supported by a simultaneously formed formate species with the band at 1580 cm^{-1} (Haneda et al., 2002) (spectra c-f). Hence, we propose that formamide species, which gives the band at 1691 cm^{-1}, was formed during the reaction of C_2H_2-SCR over 2%Zr/FER catalyst by nitric species reacting with acetylene. Correspondingly, a possible formation route of the formamide species can be proposed as follows:



In the reaction of C_2H_2-SCR over 2%Zr/FER, formate species may be produced by the formamide species further reaction with nitric species:



In literature, it is widely accepted that amine species are curtail intermediates of HC-SCR (Poignant et al., 2001; Joubert et al., 2006; Larrubia et al., 2001; Arve et al., 2007). Fig. 31 shows FTIR spectrum of surface species on 2%Zr/FER at 300 ^oC recorded when a

saturated state was reached on the catalyst in the gas mixture of C_2H_2+NO+O_2 in N_2, and the transient spectra recorded when C_2H_2 was cut off from the gas mixture. Obviously, the bands at 1691 cm^{-1} due to formamide species instantly decreased in intensity by switching the gas mixture of C_2H_2+NO+O_2 to NO+O_2, indicating that formamide species is much reactive for reacting with NO+O_2 (and/or nitrate species) over the catalyst at the temperature. The result is in line with the assumption concerning the formate species formation.

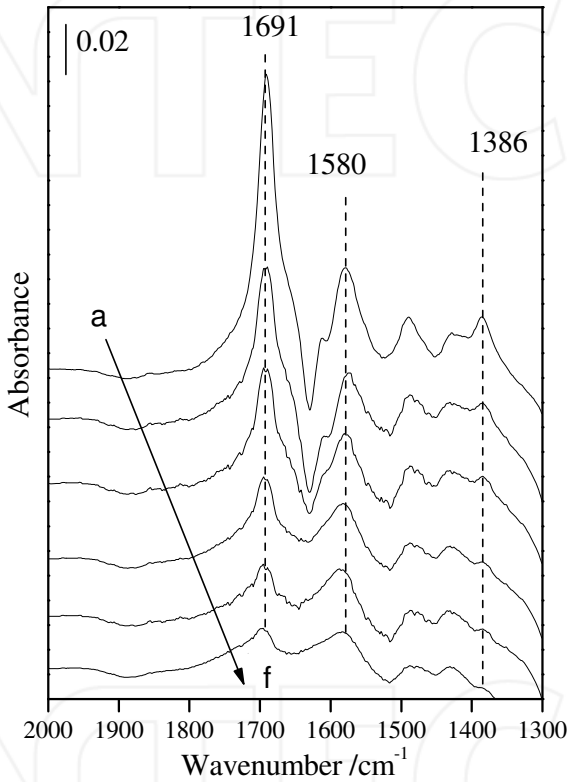


Fig. 31. FTIR spectra of surface species on 2%Zr/FER at 300 °C when the catalyst was exposed to 500 ppm C_2H_2 + 1000 ppm NO + 10 % O_2 + N_2 for 25 min (a) and subsequently exposed to 1000 ppm NO + 10 % O_2 + N_2 for 1 min (b), 5 min (c), 10 min (d), 15 min (e) and 30 min (f)

Steady state FTIR spectra of surface species formed on 2%Zr/HFER and HFER in gas mixture of C_2H_2+NO+O_2 in N_2 at some desired temperatures were compared in Fig. 32. The band at 1691 cm^{-1} due to formamide species on 2%Zr/HFER was higher in intensity than that on HFER at each temperature, which may be the result of the higher concentrated nitrate species on 2%Zr/HFER compared to HFER, as discussed in the section 3. It is in good accordance with their order in activity for C_2H_2-SCR (Fig. 8).

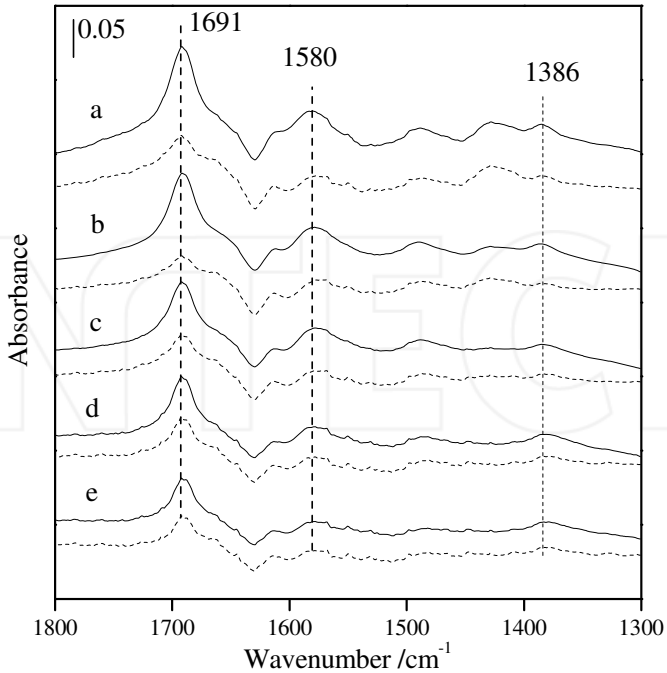
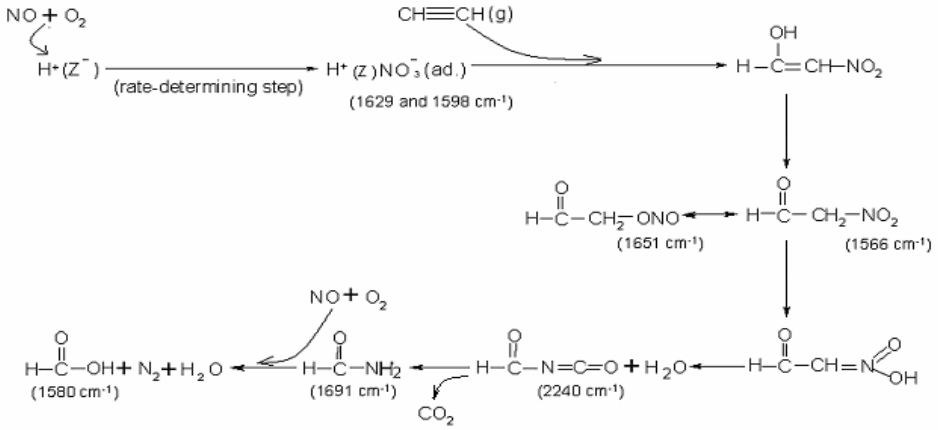


Fig. 32. Steady state in situ FTIR spectra of surface species on 2%Zr/FER (solid curve) and on HFER (dash curve) in gas mixture of 500 ppm C_2H_2 + 1000 ppm NO + 10 % O_2 + N_2 at the temperature: 250 °C (a), 300 °C (b), 350 °C (c), 400 °C (d) and 450 °C



Scheme 5. Possible reaction mechanism of C_2H_2-SCR over Zr/HFER

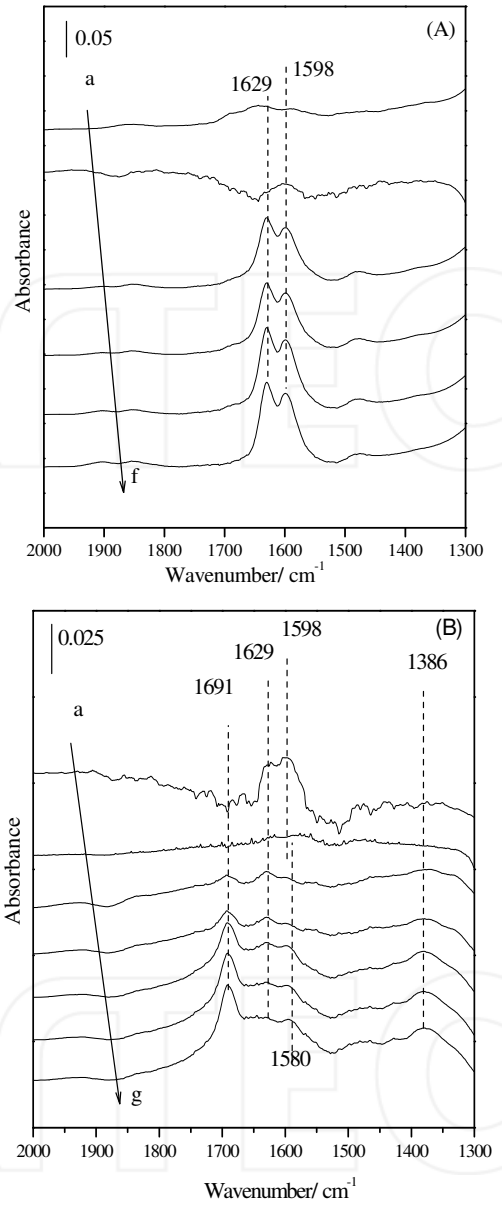


Fig. 33. FTIR spectra of the species recorded on two 2%Zr/FER catalyst wafers at 300 ºC: (A) In gas mixture of 500 ppm C_2H_2 + 10 % O_2 + N_2 (a), followed by a brief evacuation (b), then exposed to gas mixture of 1000 ppm NO + 10 % O_2 + N_2 for 1 min (c), 2 min (d), 5 min (e) and 30 min (f). (B) In gas mixture of 1000 ppm NO +10 % O_2 + N_2 (a), followed by a brief evacuation (b), then exposed to gas mixture of 500 ppm C_2H_2 +10 % O_2 + N_2 for 1 min (c), 2 min (d), 5 min (e), 10 min (f) and 30 min (g)

To have an insight into the formation route of formamide species, the reaction of acetylene in gas phase with the species formed on catalyst surface and that between the species were studied in different reaction conditions. No bands due to formamide species could be observed in the following conditions: (1) Gas mixture of C_2H_2+NO+O_2/N_2 was flowed through empty FTIR cell (without catalyst) at 300 ^oC (not shown). (2) Gas mixture of NO+O_2/N_2 was flowed through 2%Zr/HFER at 300 (Fig. 33 A) or 250 ^oC (not shown), before which the catalyst was exposed to C_2H_2+O_2/N_2 for 30 min and evacuated briefly. These results indicate that formamide species can be produced by NO+O_2 reacting neither with C_2H_2 in gas phase nor with the surface species arisen from C_2H_2+O_2 co-adsorption. However, the bands at 1691, 1580 and 1386 cm^{-1} due to formamide and formate species clearly appeared when the catalyst was exposed to gas mixture of C_2H_2+O_2/N_2 at 300 ^oC after a pretreatment in NO+O_2+N_2 for 30 min and briefly evacuated (Fig.33 B). It reveals that formamide species can only be produced by C_2H_2 in gas phase reacting with nitrate species.
Based on above discussion, the reaction mechanism of C_2H_2-SCR over Zr/HFER can be outlined in Scheme 5.

6. The end words of the chapter

For the convenience of reading, we would like to give a note describing the experimental results obtained by FTIR in the chapter to the readers.
The in situ FTIR studies were carried out in a quartz IR cell equipped with CaF_2 windows on a Nicolet 360 FTIR spectrophotometer. All of the spectra were obtained by accumulating 32 scans at a resolution of 2 cm^{-1}. The IR adsorption arising from gas phase on the pathway of infrared laser in the cell were recorded at desired temperature, which was coded as $S_g(T)$. For each experiment, the self-supporting wafer of the catalyst sample was first activated in the cell at 500 ^oC in N_2, and then its IR absorption was measured at each temperature, which was coded as $S_b(T)$. Thus the transient and steady states in situ IR spectra of surface species on the catalyst samples in gas mixtures given in this chapter are those in which the corresponding $S_g(T)$ and $S_b(T)$ were strictly subtracted.

7. Acknowledgments

This corresponding investigation was financially supported by the National Natural Science Foundation of China (Grant No. 20833011 and 20877015) and the State Hi-tech Research and Development Project of the Ministry of Science and Technology of China (Grant No. 2008AA06Z319).

8. References

Anunziata O. A.; Beltramone A. R.; Requejo F. G.. *J. Mol. Catal. A*, 267 (2007) 194.
Arve K.;Backman H.; Klingstedt F.; Eränen K.; Murzin Yu D.. *Appl. Catal. B*, 70 (2007) 65.
Bentrup U.; Brückner A.; Richter M.; Fricke R.. *Appl. Catal. B*, 32 (2001) 229.
Bentrup U.; Richter M.; Fricke R.. *Appl. Catal. B*, 55 (2005) 213.
Berndt H.; Schütze F.-W.; Richter M.; Sowade T.; Grünert W.. *Appl. Catal. B*, 40 (2003) 51
Bion N.; Saussey J.; Haneda M.; Daturi M. *J. Catal*., 217 (2003) 47.
Brosius R.; Bazin P.; Thibault-Starzyk F.; Martens J. A. *J. Catal*., 234 (2005) 191.

Burch R.; Watling T. C.. *Catal. Lett*., 37 (1996) 51.
Chafik T.; Kameoka S.; Ukisu Y.; Miyadera T. *J. Mol. Catal. A*, 136 (1998) 203.
Córdoba L.F.; Fuentes G.A.; Correa C.M.. *Micropor. Mesopor. Mater*., 77 (2005) 193.
Cowan A.D.; Cant N.W.; Haynes B.S.; Nelson P.F. *J. Catal*., 176 (1998) 329.
Delahay G.; Guzmán-Vargas A.; Coq B. *Appl. Catal. B*, 70 (2007) 45.
Dorado F.; de Lucas A.; García P.B.; Romero A.; Valverde J. L. *Ind. Eng. Chem. Res*., 44 (2005) 8988.
Elzey S.; Mubayi A.; Larsen S.C.; Grassian V.H.. *J. Mol. Catal. A*, 285 (2008) 48.
Flores-Moreno J. L.; Delahay G.; Figueras F.; Coq,B.. *J. Catal*., 236 (2005) 292.
Furusawa T.; Seshan K.; Lefferts L.; Aika K.. *Appl. Catal. B*, 39 (2002) 233.
Gerlach T.; Schütze F.W.; Baerns M.. *J. Catal*., 185 (1999) 131.
Goula G.; Katzourakis P.; Vakakis N.; Papadam T.; Konsolakis M. *Catal. Today*, 127 (2007) 199.
Gutierrez L.; Ulla M.A.; Lombardo E.A.; Kovács A.; Lónyi F., Valyon J. *Appl. Catal. A*, 292 (2005) 154.
Gutierrez L.B.; MiróE.E.; Ulla M.A.. *Appl. Catal. A*, 321 (2007) 7.
Hadajiivanov K.; Saussey J.; Freysz J.L.; Lavalley J.C.. *Catal. Lett*., 52 (1998) 103.
Hadjiivanov K.; Penkova A.; Daturi M.; Saussey J.; Lavalley J.C.. *Chem. Phy. Lett*., 377 (2003) 642.
Haneda M.; Bion N.; Daturi M.; Saussey J.; Lavalley J.; Duprez D.; Hamada H. J. *Catal*., 206 (2002) 114.
He C.H.; KÖhler K.. *Phys. Chem. Chem. Phys*., 8 (2006) 898.
He C.H.; Paulus M.; Find J.; Nickl J.A.; Eberle H.J.; Spengler J.. J. *Phys. Chem. B*, 109 (2005) 15906.
He H.; Zhang C.; Yu Y.. *Catal. Today*, 90 (2004) 191.
Held W.; Konig A.; Ritcher T.; Puppe L.. *SAE Paper*, 900496 (1990).
Huuhtanen M.; Kolli T.; Maunula T.; Keiski R.L.. *Catal. Today*, 75 (2002) 379.
Ingelsten H.H.; Zhao D.; Palmqvist A.; Skoglundh M.. J. *Catal*., 232 (2005) 68.
Ivanova E.; Hadjiivanov K.; Klissurski D.; Bevilacqua M.; Armaroli T.; Busca G.. *Micropor. Mesopor. Mater*,46 (2001) 299.
Iwamoto M.; Yahiro H.; Yu-u Y.; Shundo S.; Miauno N.; Shokubai, 32(1990) 430.
Joubert E.; Courtois X.; Marecot P.; Canaff C.; Duprez D.. J. *Catal*., 243 (2006) 252.
Kameoka S.; Ukisu Y.; Miyadera T.. *Phys. Chem. Chem. Phys*., 2 (2000) 367.
Kikuchi E.; Yogo K.. *Catal. Today*, 22 (1994) 73.
Kubacka A.; Janas J.;Włoch E.; Sulikowski B.. *Catal. Today*, 101 (2005) 139.
Kubacka A.; Janas J.; Sulikowski B.. *Appl. Catal. B*, 69 (2006) 43.
Larrubia M. A.; Ramis G.; Busca G.. *Appl. Catal. B*, 30 (2001) 101
Lee T.J.; Nam I.S.; Ham S.W.; Baek Y.S.; Shin K.H.. *Appl. Catal. B*, 41 (2003) 115.
Li G.; Larsen S.C.; Grassian V.H.. *Catal. Lett*., 103 (2005) 23.
Li G.; Larsen S.C.; Grassian V.H.. *J. Mol. Catal. A*, 227 (2005) 25.
Li G..; Wang X.; Jia C.; Liu Z.. *J. Catal*., 257 (2008) 291.
Li L.; Chen J.; Zhang S.; Guan N.; Richter M.; Eckelt R.; Fricke R.. *J. Catal*., 228 (2004) 12
Li L.; Zhang F.; Guan N.; Richter M.; Fricke R.. *Catal. Commun*., 8 (2007) 583.
Li L.; Zhang F.; Guan N.. *Catal. Commun*., 9 (2008) 409.
Li Y.; Armor J.N.. *J. Catal*., 145 (1994) 1.
Liu Z.; Woo S.I.; Lee W.S.. *J. Phys. Chem. B*, 110 (2006) 26019.

Lónyi F.; Valyon J.; GutierreL. Z; Ulla M.A.; Lombardo E.A.. *Appl. Catal. B*, 73 (2007) 1.
Loughran C.E.; Resasco D.E.. *Appl. Catal. B*, 7 (1995) 113.
Luo C.; Li J., Zhu Y.; Hao J.. *Catal. Today*, 119 (2007) 48.
Ma Xiaofei; Wang Xinping; Bi Ran; Zhao Zhen; He Hong. *J. Mol. Catal. A*, 303(2009) 90.
Meunier F.C.; Zuzaniuk V.; Breen J.P.; Olsson M.; Ross J.R.H.. *Catal. Today*, 59 (2000) 287.
Mihaylov M.; Hadjiivanov K.; Panayotov D.; *Appl. Catal. B*, 51 (2004) 33.
Mosqueda-Jiménez B.I.; Jentys A., Seshan K.; Lercher J.A.. *Appl. Catal. B*, 43 (2003) 105.
Mosqueda-Jiménez B.I.; Jentys A.; Seshan K.; Lercher J.A.. *Appl. Catal. B*, 46 (2003) 189.
Nakamoto K.. Infrared Spectra of Inorganic and Coordination Compounds (Mir, Moscow. 1996)
Narbeshuber T.F.; Brait A.; Seshan K.; Lercher J.A.. *J. Catal.*, 172 (1997) 127.
Nishizaka Y.; Misono M.. *Chem. Lett*., (1994) 2237.
Niu J.; Yang X.; Zhu A.; Shi L.; Sun Q.; Xu Y.; Shi C.. *Catal. Commun*., 7 (2006) 297.
Pieterse J.A.Z.; van den Brink R.W.; Booneveld S.; de Bruijn F.A.. *Appl. Catal. B*, 46 (2003) 239.
Pieterse J.A.Z.; Booneveld S.; *Appl. Catal. B*, 73 (2007) 327.
Pirngruber G.D.; Pieterse J.A.Z.. *J. Catal*., 237 (2006) 237.
Poignant F.; Freysz J.L.; Daturi M.; Sauaaey J.. *Catal. Today*, 70 (2001) 197.
Resini C.; Montanari T.; Nappi L.; Bagnasco G.; Turco M.; Busca G.; Bregani F.. *J. Catal*., 214 (2003) 179.
Sasaki M.; Hamada H.; Kintachi Y.; Ito T.. *Catal. Lett.,* 15 (1992) 297.
Satsuma A.; Yamada K.; Sato K.; Shimizu K.; Hattori T.; Murakami Y.. *Catal. Lett*., 45 (1997) 267.
Satsuma A.; Shimizu K.. *Prog. Ener. Com. Sci*., 29 (2003) 71.
Sedlmair C.; Gil B.; Seshan K.; Jentysa A.; Lercher J.A.. *Phys. Chem. Chem. Phys*., 5 (2003) 1897.
Sedlmair C.; Seshan K.; Jentys A.; Lercher J.A.. *J. Catal*., 214 (2003) 308.
Seijger G.B.F.; van Kooten Niekerk P.; Krishna K.; Calis H.P.A.; van Bekkum H.; van den Bleek C.M.. *Appl. Catal*. B; 40 (2003) 31.
Shi C.; Cheng M.; Qua Z.; Yang X.; Bao X.. *Appl. Catal*. B, 36 (2002) 173.
Shibata J.; Shimizu K.; Satokawa S.; Satsuma A.; Hattori T.. *Phys. Chem. Chem. Phys*., 5 (2003) 2154.
Shibata J.; Takada Y.; Shichi A.; Satokawa S.; Satsuma A.; Hattori T.. *Appl. Catal. B*, 54 (2004) 137.
Shichi A.; Satsuma A.; Iwase M.; Shimizu K.; Komai S.; Hattori T.. *Appl. Catal. B*, 17 (1998) 107.
Shichi A.; Katagi K.; Satsuma A.; Hattori T.. *Appl. Catal. B*, 24 (2000) 97.
Shichi A.; Satsuma A.; Hattori T.. *Appl. Catal. A*, 207 (2001) 315.
Shichi A.; Satsuma A.; Hattori T.. *Appl. Catal. B*, 30 (2001) 25.
Shichi A.; Satsuma A.; Hattori T.. *Catal. Today*, 93–95 (2004) 777.
Shimizu K.; Kawabata H.; Maeshima H.; Satsuma A.; Hattori T.. *J. Phys. Chem. B*, 104 (2000) 2885.
Shimizu K.; Okada F.; Nakamura Y.; Satsuma A.; Hattori T.. *J. Catal*., 195 (2000) 151.
Shimizu K.; Shibata J.; Yoshida H.; Satsuma A.; Hattori T.. *Appl. Catal. B*, 30 (2001) 151.
Shimizu K.; Sugino K.; Kato K.; Yokota S.; Okumura K.; Satsuma A.. *J. Phys. Chem*. C, 111 (2007) 6481.

Stakheev A.Y.; Lee C.W.; Park S.J.; Chong P.J.. *Catal. Lett*., 38 (1996) 271.
Sullivan J. A.; Keane O.. *Appl. Catal*. B, 61 (2005) 244.
Sultana A.; Haneda M.; Fujitani T.; Hamada H.. *Micropor. Mesopor. Mater*., 111 (2008) 488.
Szanyi J.; Paffett M.T.. *J. Catal*., 164 (1996) 232.
Szanyi J.; Kwak J.H.; Moline R.A.; Peden C.H.F.. *Phys. Chem. Chem. Phys*., 5 (2003) 4045.
Szanyi J.; Kwak J.H.; Peden C.H.F.. *J. Phys. Chem. B,* 108 (2004) 3746.
Torre-Abreu C.; Henriques C.; Ribeiro F.R.; Delahay G.; M Ribeiro.F.. *Catal. Today*, 54 (1999) 407.
Tsyntsarski B.; Avreyska V.; Kolev H.; Marinova Ts.; Klissurski D.. J. *Mol. Catal. A*, 193 (2003) 139.
Wang C.; Wang X.; Xing N.; Yu Q.; Wang Y. *Applied Catalysis A*. General 334 (2008) 137.
Wang X.; Xu Y.; Yu S.; Wang C.. *Catal. Lett*., 103 (2005) 101.
Wang X.,; Yu S.; Yang H.; Zhang S..*Appl. Catal. B*, 71 (2007) 246.
Wang X.; Zhang S.; Yu Q.; Yang H.. *Micropor. Mesopor. Mater*., 109 (2007) 298.
Wang X.; Yang H.; Yu Q.; Zhang S.. *Catal. Lett*., 113 (2007) 109.
Wang X.; Yu Q.; Li G.; Liu Z.; *Catal. Lett*., 123 (2008) 289.
Wu Q.; He H.; Yu Y.. *Appl. Catal. B*., 61 (2005) 107.
Xing N.; Wang X., Zhang A.; Liu Z., Guo X.. *Catal. Comm*., 9 (2008) 2117.
X Zhan.; He H.; Ma Z.. *Catal. Comm*., 8 (2007) 187.
Yeom Y. H.; Li M.; Sachtler W. M. H.; Weitz E.. *J.Catal*., 238 (2006) 100.
Yu Q.; Wang X.; Xing N.; Yang H.; Zhang S.. *J. Catal.,* 245 (2007) 124.
Yu Y.; He Hong; Feng Q.; Gao H.; Yang X.. *Appl. Catal*. B, 49 (2004) 159.

17

Description of Topography of Surfaces and Thin Films with the use Fourier Transformation, Obtained from Non-Standard Optical Measurements

Janusz Jaglarz
Institute of Physics, Cracow University of Technology
Poland

1. Introduction

This chapter explains how to interpret the results obtained from optical scattering study to characterize topography surfaces and films. In particular, it focuses on measurement of bidirectional reflectance distribution function (BRDF) and on optical profilometry (PO). These techniques allow to determine two main functions describing surface topography, namely: power spectral density (PSD) and autocorrelation function (ACF). They characterize any real surface quantitatively and qualitatively. The PSD and ACF functions are commonly calculated from Fourier transform (FT) of the surface profiles determined in AFM, SEM or PO) measurements.

The optical scattering is directly related to material and surface properties. We can distinguish three types of scatterings base on their features: 1) Topographic scattering resulting fom roughness causing phase fluctuations impressed on the reflected wavefront by the surface height variation [Beckmann and Spizinochino 1963] 2) material scattering created by fluctuations in the composition or density of the surface material [Elson 1984]. Defect scattering is resulting from a presence of sparse distribution of some surface features responsible for scattering, is distributed broadly and continuously over surfaces such as pits or bumps in case of topography and patches of different reflectivity in the case of material scattering [Stover 1995a].

One should emphasize that the roughness is the main source of scattering on surfaces at visible wavelengths range.

Topography of real surface may be described using some fundamental parameters described in national and international norms [ANSI/ASME B46.1, ISO 2517]

For macroscopic characterization of surface the following terms is applied [Whitehause 2003]): 1) Root mean square roughness σ (rms). It is calculated by the vertical deviations of a real surface from its ideal form. 2) The rms slope *s* and correlation length which is expressed by the ratio: $T = 2^{1/2}\ \sigma/s$. 3) The power coefficient the type of qualifying statistical distribution of surface heights. The distance between respective surface features is defined as spatial wavelengths. In topographic analysis the term spatial frequency which is a reverse of spatial wavelength is used.

The given statistical terms are calculated from the surface-profile data in a specified spatial bandwidths. The range of measured spatial wavelengths is determined by lateral resolution of measuring tool and sampling length in short and long wavelength limit respectively.

2. The power spectral density and autocorrelation function

The real surface and films are described by two statistical functions, namely autocorrelation (ACF) and power spectral density (PSD) functions [Bennett and Mattsson 1999]. PSD function is defined in spatial frequency domain and it expresses the roughness power per unit frequency over the sampling length. From mathematical point of view, PSD is evaluated from Fourier transform (FT) of surface profile *h(r)*. For one dimensional profile *h(r)* PSD is described as:

$$PSD(f_r) = \lim_{L\to\infty} \left\langle \left| \frac{2}{L} \int_{-L/2}^{L/2} h(r) e^{i2\pi f} \, dr \right|^2 \right\rangle \tag{1}$$

Where f – is spatial frequency and L is sampling length. The units one dimensional PSD are length to third power.

The roughness and average slopes are first and second statistical momentums of PSD function so they may well be directly calculated from PSD frequency spectra by the following integrals:

$$\sigma^2 = \int_{f_{min}}^{f_{max}} PSD(f)\, df \tag{2}$$

$$s^2 = \int_{f_{min}}^{f_{max}} (2\pi f)^2 PSD(f)\, df \tag{3}$$

If values of PSD are known, one can determine the statistical parameters, such as root-mean square (rms) roughness, slopes and correlation length by using ABC model which describe PSD in simple analytical formula [Elson and Benett 1995]:

$$PSD(f) = A[1 + (Bf)^2]^{-C/2} \tag{4}$$

where A, B, C are model parameters related with basic quantities characterizing a surface.

The A parameter is a *PSD(f)* value for low frequency formula, $B/2\pi$ is correlation length and C determine type of power law in high spatial frequency. The ABC model applies for single surface.

The other way of surface characterization is describing it by means of autocorrelation function *ACF(r)* defined in spatial wavelength domain. Generally the autocorrelation function of surface is used to compare data sets of heigths *h(r)* to a translated version of itself and the averaging.

$$ACF(r) = \lim_{L\to\infty} \frac{1}{L} \int_{-L/2}^{L/2} h(r) h(r+\tau) dr \tag{4a}$$

Where translation τ is so called lag distance. The ACF function is usually calculated from surface profiles obtained in AFM and SEM studies. For many random surfaces the *ACF(r)* may be expressed in the following form [Bendat and Piersol 1971, Whitehead 2003]:

$$ACF(r) = ACF(0)\exp(-\left|\frac{r}{T}\right|^{2\alpha}) \tag{5}$$

Where ACF(0) is the value of autocorrelation function in point $r = 0$. The value of coefficient specifies a type of statistical distribution of surface irregularities. So for $\alpha = ½$ and $\alpha = 1$ the ACF(r) describes Lorentzian and Gaussian distribution respectively. For distance $r = T$ the *ACF(r)* decreases e – fold. Thus the value of $r = T$ defines autocorrelation length. Such defined ACF is equal statistically averaged surface spatial wavelength.

In some surface investigations obtained by the use of optical methods, as in optical profilometry, the following useful relation between average slope *s* and ACF function could be applied [Stover *et. al* 1984]:

$$m^2 = \lim_{\tau \to \infty} \frac{d^2 ACF(\tau)}{d\tau^2} \tag{6}$$

According to Parseval's theorem PSD and ACF provide the same information about surface statistic expressed in space frequency and wavelength domain.

Therefore the PSD and ACF may express by using Wiener – Kninchin relations [Bendat and Piersol 1986] as Fourier transform and inverse FT:

$$ACF(r) = \int_{-\infty}^{\infty} PSD(f_x) e^{j2\pi f \tau} df \tag{7}$$

$$PSD(f) = \int_{-\infty}^{\infty} ACF(\tau) e^{-j2\pi f \tau} d\tau \tag{8}$$

The calculating procedure of macroscopic surface parameters is shown on Fig. 1.

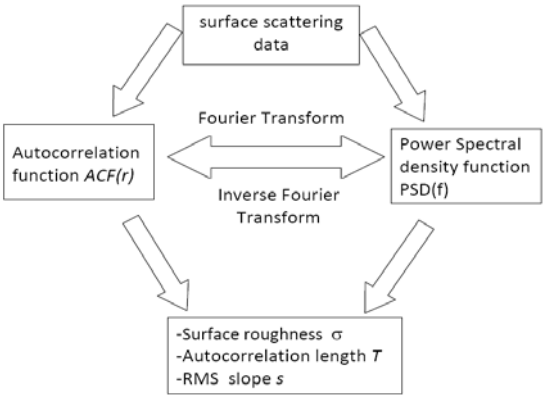
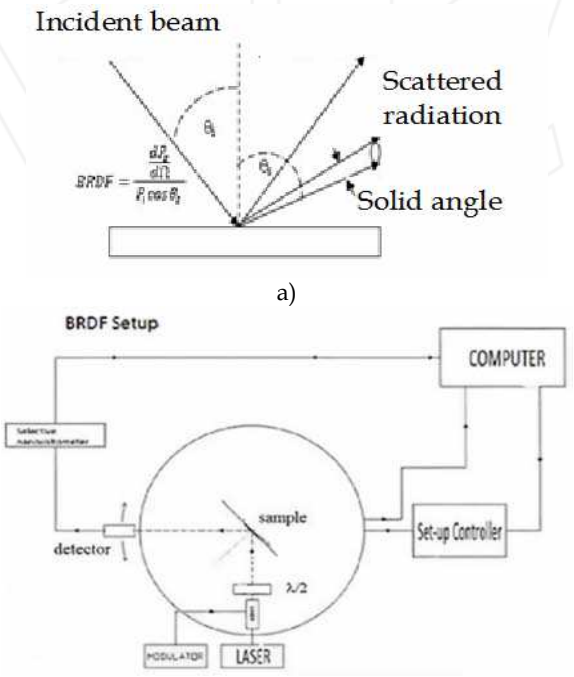


Fig. 1. Scheme of calculating surface parameters procedure from optical measurements

3. The BRDF technique

Optical measurements allow to find macroscopic parameters of surfaces and films from angular measurements of scattered light. The very good tool for single surface and multilayered films characterization is bidirectional reflection function method (BRDF) [Nicodemus 1965].

In BRDF the differential power of scattered beam dP per solid angle of receiver aperture $d\Omega$ in the θs direction and per incident power P_i coming from the θ_i direction is measured. In Fig. 2a the geometry of incident and measured reflected radian beams have been presented.



a)

b)

Fig. 2. Geometry of BRDF measurement (a) and scheme of BRDF measurement setup (b)

Practically $dP/d\Omega$ is equal to the measured scatter power P_s per acceptance angle Ω of a detector namely:

$$BRDF = \frac{dP / d\Omega}{P_i \cos\theta_s} = \frac{P_S}{P_i \Omega \cos\theta_S} \left[sr^{-1} \right] \tag{9}$$

If studied surface is relatively flat, the excellent approach give vector The Raleigh-Rice vector perturbation theory [Elson and Benett 1979] is best applied. It shows simple dependence between scattered radiation expressed by BRDF and PSD of investigated surface.

$$BRDF = \frac{16\pi^2}{\lambda^2} \cos\theta_i \cos\theta_s Q \, PSD(f) \tag{10}$$

where Q is a factor dependent on polarization state of the light source, the optical constants and λ where λ is the light wavelength. PSD is represented in spatial frequency f that is related with angles θ_i and θ_s by following formula:

$$f = \frac{\sin(\theta_s) - \sin(\theta_i)}{\lambda} \tag{11}$$

If the surface roughness is spatially isotropic, as is usually assumed for randomly polished surfaces, the PSD depends only on the magnitude of the surface spatial frequency and is independent of its direction on the surface plane. All information about the surface spectrum can be obtained from scattering measurements made in the plane of incidence. This allows to extract the topographic structure of the single surface from scattering phenomena, which means that BRDF and PSD (except for the factor $Q\cos\theta_s$) are directly proportional. The so called "golden rule" is a principle of surface investigations by the use angular scatterometry [Church et al. 1977, Stover 1995b].

BRDF measurements presented in this chapter were performed with an automatic scatterometer (Fig. 2b). It consists of a 650 nm laser diode as the light source (with the beam diameter of 2 mm) mounted on a goniometric table with 0.01 deg resolution. The light scattered at the sample surface is measured with a Si photodiode detector. The rotations are obtained by a computer controlled stepper motors. For a fixed angle of incidence, the scattered intensity in the plane of incidence are measured by varying the detector orientation. All measurements are carried out with the *s*-polarized incident beam. In any case, the sample surface size was much larger than the beam diameter. Moreover, the minimal illuminated area (4 mm^2) is large enough to yield meaningful statistical description of the surface [Stover 1984].

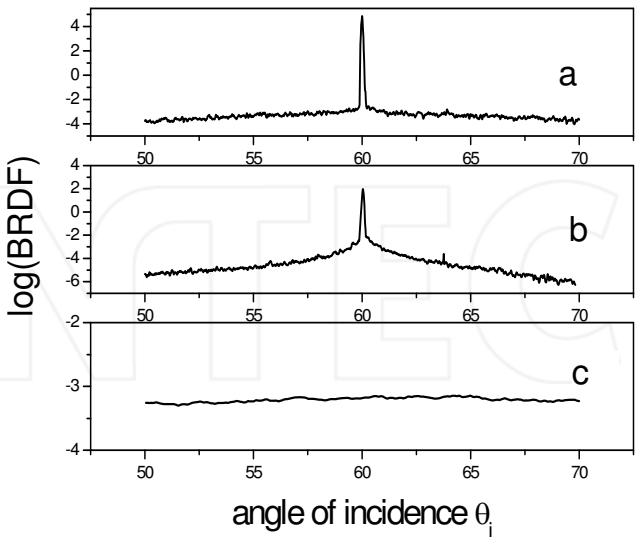


Fig. 3. BRDF vs. angle of incidence for 3 different samples: flat silicon–a, polished steel – b, SWM7 and Spectralon – c surfaces

Fig. 3 shows the results of BRDF measurements as a function of scattering angle for 3 samples. The image in Fig. 3 consists of polished silicon with roughness $\sigma = 1$nm (Fig. 3a), mechanically polished steel $\sigma = 69$ nm SWM7 with (Fig. 3b) and lambertian white diffuser Spectralon (Fig. 3c). Because scattered intensity strongly depends on scattering angle and intensity ratio for near specular to far specular scattering may reach several orders so generally the BRDF is shown in logarithmic scale. The incidence angle was the same $\theta_i = 60^0$ for all samples. As it is easy to notice the angular BRDF's are different. The BRDF of polished Si as other good mirrors has high narrow specular peak. Outside specular reflection mirror reflectors scatter diffusely light at very low level. Polished steel with moderately high roughness has BRDF with lower and wider specular peak and intensity of scattered light monotonically decrease in larger scattering angle. In Spectralon [Workmann 1998] BRDF specular reflection disappears and light is scattered diffusively and homogeneously in all direction.

In the case of polished surfaces the power spectrum of the surface errors is generally a smooth and broad function of spatial frequency, and the smoothed value of the BRDF is independent of the system pupil function.

In a situation when optical constants are unknown, the polarization factor Q_s for "s" polarization could be found from angle dependence of specular reflection coefficient. It is given by the geometric mean of the sample specular reflectances $R(\theta_i)$ and $R(\theta_s)$ at angles θ_i and θ_s. For scattered radiation measured in plane of incident Q_s is equal [Church et al. 1989, Stover 1995b]:

$$Q_s = \{R_s(\theta_i) R_s(\theta_s)\}^{\frac{1}{2}} \tag{12}$$

This equation allows to determine Q_s without knowing the optical constants of the sample. Because Q_s is a smooth function, a good curve fit may be found by measuring sample reflectance for several angles of incidence. Of course, if optical indices are known, the Q_s factor can be calculated directly from the definition.

Fig. 4 Shows *PSD(f)* calculated from BRDF. The polarisation factor Q_s was determined from specular reflectance measurements performed for 7 incidence angles θ_i from range 60^0 to 81^0. The ABC model was fitted to experimental data.

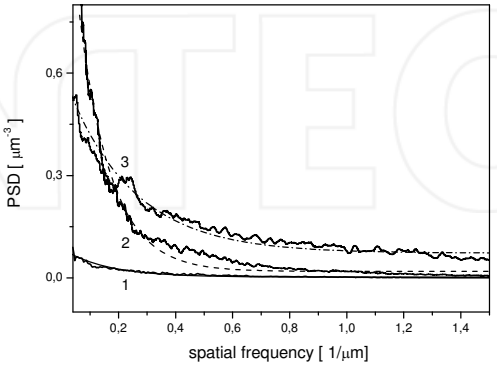


Fig. 4. PSD function extracted from BRDF measurements for TiN films on different substrates on polished: 1 —Si, 2 —SWM7 steel and 3 — on WC plate respectively

Roughness σ of TiN films from equation (4) and autocorrelation length *T*, and power factor *C* from *ABC* model (5) have been found. The values of determined parameters are shown in Table 1. The result strongly suggest the influence of substrate type on topography of TiN films obtained under the same technological conditions.

Log-Log analysis of *PSD(f)* is very convenient and easy to interpretation of *PSD(f)* dependence. Many optical surfaces are described by the means ABC model. On Fig. 5 the log-log *PSD(f)* for Duralcan alloy before and after CO_2 laser annealing [Kaczmar et al. 2000] has been presented.

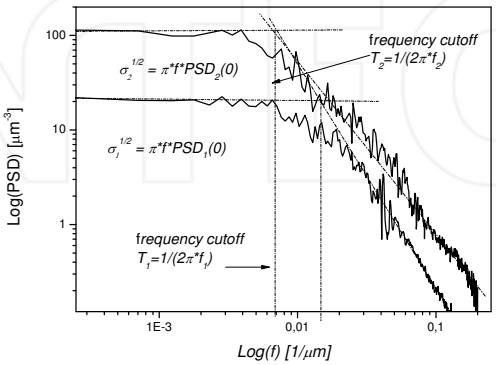


Fig. 5. The values of topographic parameters have been obtained as following:

For low spatial frequencies Bf << 1 the *PSD(f)* has constant value. For high frequencies Bf >> 1 and the formula (3) may be approached by the use ABC model as follows:

$$Log(PSD) = Log(A) - CLog(Bf) \tag{13}$$

Substituting *for y = Log(PSD)* and for *x=Log(Bf)*, we obtain linear dependence: *y = -C x + b*, where *C* qualify type of random distribution of irregularietes. For the special case *C=2* or *C=4*, the distribution of surface highs is Lorentzian or Gaussian respectively.

If we try to extrapolate *PSD(f)* curves in low and high spatial frequency bandwidth by straight lines one then we will obtain break point on cut of them. It allows to determine from this point so called cut-off frequency. By knowing its value one may to calculated autocorrelation length according formula: $T=1/2\pi*f_c$. According to (2) the surface roughness is equal field under the PSD curve. For low frequency bandwidth i.e from zero to cut-off frequency f_c one may estimate the σ follows to approached formula: $\sigma^{1/2} = \pi*f_c*PSD(0)$, where *PSD(0)* denote power spectral density at low frequency.

For some surfaces surfaces in their PSD or BRDF versus frequency the break point do not occur (that is Bf>>1). In that case the PSD and BRD may be approximated linear with slope *C*. For random one dimensional surface from ABC model we obtain from (4):

$$PSD(f) = ABf^{-C} \tag{14}$$

The formula (14) expressed in logaritmic terms is the same as (12). Surfaces without break point are called fractal surface [Church 1988]. The autocorrelation length *T* of fractal profile

is equal to the length of investigated surface. On Fig. 6 the BRDF versus spatial frequency for 3 fractal surfaces: 1-glass BK7, 2-polished stainless steel, 3-$BaSO_4$ white standard is shown.

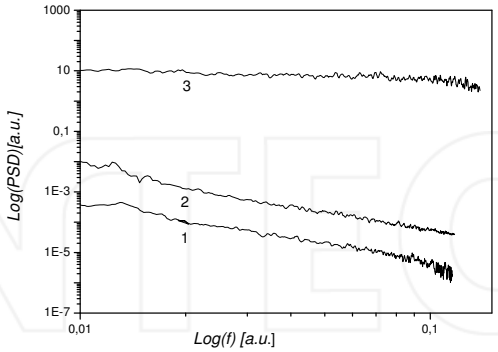


Fig. 6. The power spectral density for fractal profile, 1-BK7 glass, 2-stainless steel, 3-BaSO4

The curves 1 and 2 represents fractal glass and steel surfaces. The *C* coefficients of power spectrum slopes are nearly 4, so their highs of fractal surface have Gaussian distribution. Curve 3 presents power spectrum for barium sulphate which is applied as Lambertian diffuser [Workmann 1998]. As may easy conclude white standard diffuser are fractal surfaces with zero slope of PSD spectra.

The many practical surfaces in optics or semiconductor application exhibit the fractal-like power spectral density over bandwidths of interest.

As it is known the roughness value depends on spatial bandwidth which is specified by measurements setup and other conditions such as sample size etc. Because PSD spectra of measured surfaces may differ considerably results of roughness may be influenced by position and width of chosen bandwidth. Fig. 7 presents the calculated roughness versus spatial frequency area for SWM7 steel, stainless steel and silicon surfaces respectively.

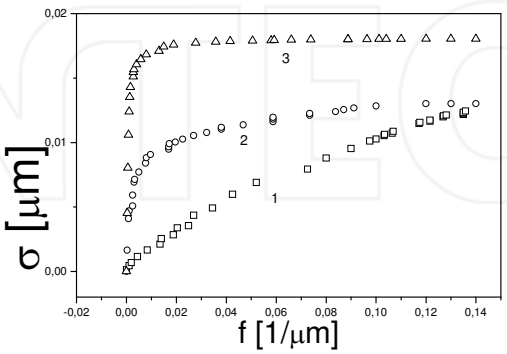


Fig. 7. Surface roughness as a function of spatial frequency bandwidth obtained from BRDF for, SWM7 steel, stainless steel and silicon surface - curve 1, 2, 3 respectively

As it may be concluded from the above figure, for some surfaces roughness values reach constant value at low frequencies and topographic features observed at high frequencies essentially do not contribute to total roughness.
If the surface roughness contains periodic components, such as tool marks in precision-machine surfaces, the BRDF will contain sharp diffraction lines. The positions and intensities of those lines are related to the feed rates and amplitudes of the tool marks, while their widths are determined by the system pupil function. The case of periodic surface is shown on Fig. 8. The BRDF for ST3SY steel concurrent (curve 2) and reverse concurrent (curve 1) in milling process is shown.

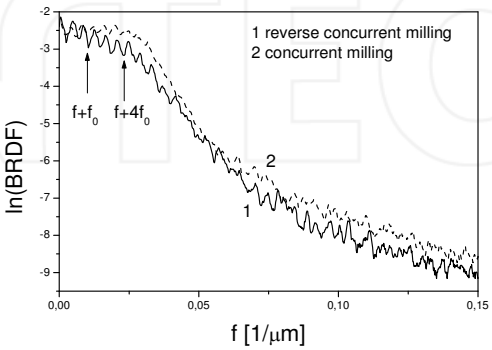


Fig. 8. BRDF for periodic steel surface obtained in milling process

Spatial wavelength of fundamental mill calculated from BRDF is equal Λ_0=154μm. All remaining peaks from low frequency bandwidth occur on BRDF spectrum in the same interval frequency $f_0 = 1/\Lambda_0$.
Some random optical surface may not be described by the use ABC model. It is for mono-crystalline silicon wafer with (100) orientation etched in KOH solution. In this case reduction of surface reflectivity is a key factor for improving the crystalline silicon solar cell efficiency. In Fig. 9 the relative BRDF frequency spectra for that wafer have been shown.

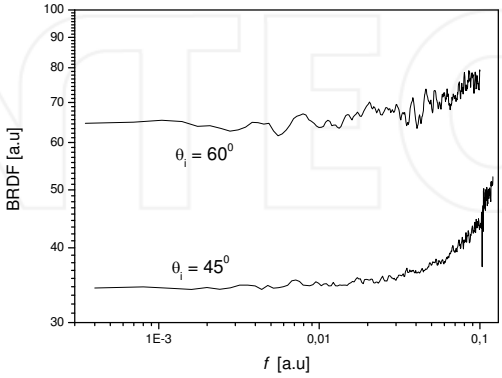


Fig. 9. BRDF of texturized monocrystalline Si in KOH for two angles of incidence

For high spatial frequency slopes have positive values unlike to most optical surfaces. This peculiar behaviour in *BRDF(f)* dependce is due to surface texture of etched silicon wafer. It consists of pyramid-like irregularietes differing in sizes and shapes [Lipiński and Cichoszewski 2008]. So created texture scatters light into hight angles (frequencies). It allows to substantially reduce reflection for near specular angles.
Additionally one influences of non-topographic features on power spectral density function should be discussed. Such cases can be found in composite (alloys) structures. If components are different the large variations in optical constants may occur. Apart from large reflectance differences, the phase change of reflection for absorbing materials appear. This phase change resulted from reflectance and may equal several tens of degrees for conductors and can be interpret as pseudo-height differences [Bennet and Mattsson 1999]. In other words a hypothetical, perfectly smooth surface made up of elements having different optical constants may show shadowing effect ordinarily interpreted as height differences in optical measurements. For flat surfaces dielectric function variation plays key role in light scattering. This situation can be found in optical study of Al-Si-SiC composites [Kaczmar el al. 2000]. Additionally in Al-Si-SiC structures such a metal-semiconductor-dielectric composites particular monocrystals may occurr on the surface and may be considerably larger than light wavelength. They have own surface direction and reflect light according to geometrical law of reflection.
The phase changes of reflected light which are components of the investigated composites (alloys) result from appreciable differences in their optical constants. In experimental data analysis, the differences in n and k indices particular components could be treated as pseudo-height caused by phase differences. For example for the aboveAl-Si-SiC composites for light with wavelength λ =550 nm, the pseudo-heights are for Al -24 nm, 2.5 nm for Si and 0 nm for SiC for which is dielectric material and was treated as reference. Therefore the pseudo-heights must be taken into consideration when predicting values of power spectral density and autocorreation functions are obtained by the use of optical methods.

4. The Power spectral density of thin films from BRDF study

The BRDF results for film depend on topographies of its upper and lower side and in the case of diffusive layers depend also on light scattering in the bulk.

1. If the topography of both upper and bottom interfaces are identical the PSD is given by [Elson 1992]:

$$BRDF = \left|F_1 + F_2\right|^2 PSD \tag{15}$$

   where F_1 and F_2 are optical function of interfaces film- air and film – substrate respectively, and PSD represents power spectral density function of film system.
2. In the case completely uncorrelated surfaces the BRDF function should be expressed by the formula:

$$BRDF = \left|F_1\right|^2 PSD_1 + \left|F_2\right|^2 PSD_2 \tag{16}$$

   where PSD_1 and PSD_2 denote power spectral density for top and bottom film surface respectively. In this case, the scattered power from the different interfaces is added. Anyway one can extract part of scattering coming only from the upper surface using full internal reflection effect in the film.

Fig 10. shows reflection from transparent film on rough substrate. Part of the light reflect from the top surface, part after refraction on the boundary of the media traverse the film twice due to reflection from substrate and goes out from the sample (see Fig. 10). This situation occurs for specular reflection for any angle of incidence. If however bottom surface is rough, the part radiation is scatter into angles θs that are larger than Brewster angle θ_B. For this case the all radiation scattered at angles $\theta s > \theta_B$ is internally reflected.

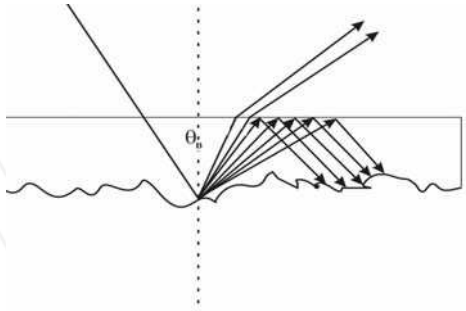


Fig. 10. Propagation of scattered light in layer with rough bottom

It means that the reflected light at suitably great angles comes only from scattering from upper interface of the film and allows to determine PSD from BRDF measurements for upper surface of the film.

If a bulk scattering does not occur the optical losses originate from surface and interface only. In that case determining power spectral density can be done from BRDF measurements. Fig. 11 shows determined PSD function for bare polished silicon 1 – curve 1, thermally obtained SiO_2 on the same Si substrate – curve 2, porous silica and titania-silica blend films on BK7 glass obtained by sol-gel technique [Karasiński 2005] –curve 2 and 3 respectively.

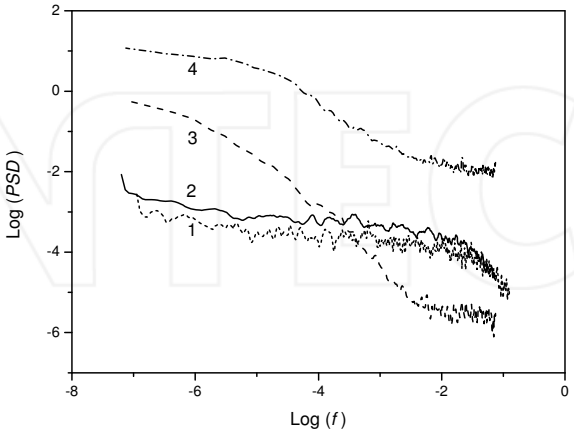


Fig. 11. PSD function of thin films. Curve 1 – bare silicon, 2 –Thermal SiO_2 on Si, 3 – porous silica and 4 porous SiO_2-TiO_2 on BK7 glass

As it may be noticed on Fig. 11 for films 2 and 3 the irregularities with low spatial frequency (longer spatial wavelength) contribute larger fraction to total roughness than the higher ones. Therefore autocorrelation lengths *T* for layers 1 and 2 are bigger than for thermally obtained silica. The PSD obtained for silicon and SiO_2 on the same Si before after annealing are very similar. It results from the fact that upper and lower interfaces of SiO_2 film are identical. The values of σ determined from AFM and BRDF measurements are similar however roughness calculated from BRDF study are larger. As a matter of fact in BRDF we measure scattered radiation from much bigger surface area than in AFM. The total roughness enlarge due to fraction of irregularities coming from longer spatial wavelengths measured by the use BRDF technique.

The values of macroscopic parameters for surface and films presented in this work have been shown in Table 1. The results given in Table 1 concern optical surfaces. Such surfaces can meet I optics and optoelectronics applications. For those surfaces roughness is commonly less than 5 nm and their autocorrelation length are order hundred manometers to micrometers. So well average slopes are less than 10^{-2}. In optics surfaces with slopes less than 10^{-2} are regarded as flat [Ohlidal 1988].

sample	Thickness [nm]	Roughness σ [nm]	T [nm]	C
Polished Si	-	1.7	2730	3.91
Thermal SiO2 on Si	352	2.9	2890	3.88
Porous silica on Si	672	2.1	770	1.97
titania-silica on Si	211	1.2	455	2.11
TiN on Si	330	11.7	6780	3.61
TiN on steel	393	34	13890	3.81
Polished aluminium		7.3	3100	3.64

Table 1. The values of topographic parameters of random layers and surfaces

5. Determination of the PSD and ACF function from optical profilometry

Measurements of optical reflectance by means of classical reflectometry inform us about optical properties on a large area, i.e. of the order of 1-5 cm^2. The results obtained on a much smaller will be similar if coatings and surfaces are homogenous over the investigated area and inside the layers. For inhomogeneous surfaces, when topographic or materials non-uniformities differ from tens μm to several mm, the measurements taken from the integrating sphere measurement and standard reflectometry give rather an averaged reflectance over a larger scale reflected samples.

The scattered radiation measured by optical profilometry (OP) is a function of heights of irregularities and slopes of microfacets, but sensitivity of this method derives mainly from detection of the slope change [Brown and Breitmeier 1988, Whitehead 2003]. The presence of

long lateral irregularities is often caused by manual or mechanical treatments and may have a periodical nature. The short spatial waves result rather from random process of the surface formation and their contribution to the total profile is easy to determine from atomic force microscopy (AFM) technique.

Optical profilometry measures intensity of reflected radiation from the surface point. Resolution of OP depends on beam diameter where diameters changes from 1 μm to 1 mm. In this chapter we presented result of profilometric studies. The long spatial wavelength irregularities detected in OP investigations may contribute substantially to the total roughness. Optical profilometry measurements complete the topography description in long spatial wavelengths.

The optical profilometer described in this chapter is multifunctional experimental set up for surface topography investigations. It work in two modes. The first mode – specular mode employs with laser He-Ne as the light source with collimating system allowing to achieve light beam with 12 μm diameter. OP measurements have been normalized with the calibration sphere method [Mainsah et al. 1996]. A lead screw stepper motor actuated device may scan 30 mm × 30 mm surface with step of 0.02 mm. It allows one to obtain the optical profiles of surface with 20 μm lateral resolution. The scheme of optical profilometer (OP) is shown in Fig. 12.

The second mode of OP – BRDF mode which is more preferred for rougher films surfaces is based on reflection probe R-200-7. This mode is used to get reflectance over smaller areas at different angles of incidence. The probe consists a bundle of 7 optical fibres, i.e. 6 fibres around one "illumination fibre" (which illuminates sample), each with a diameter of 200 μm. Of course the size of illumination defines lateral resolution of measurements. The laser diode (λ = 635 nm) is used as the light source. The reflected light is collected by six fibres around the illumination fibre. The probe is coupled with a collecting fibres to a A/D converter connected with computer. The axis of the probe may be angled at 15°, 30° and normal to the sample plane.

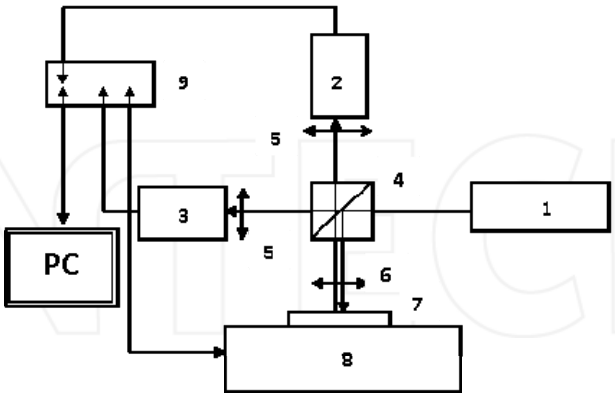
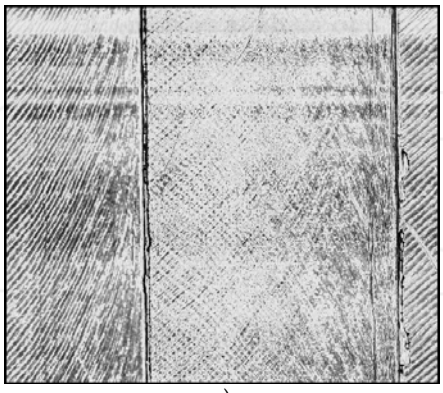


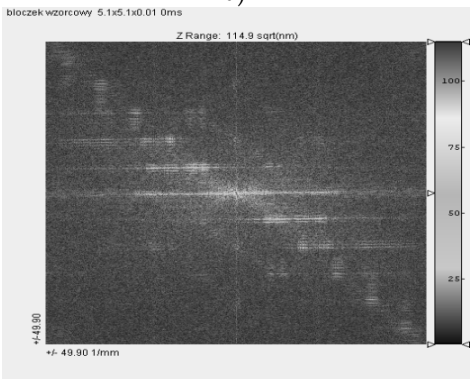
Fig. 12. Scheme of optical profilometer: 1 – laser diode (λ = 635 nm), 2,3 – detectors, 4 - beam splitter, 5 - colleting lens, 6 - objective, 7- sample, 8 -XY stage, 9 – controlling/collecting unit

The analysis of two-dimensional Fourier transforms of surface profiles expressed as PSD function of f_x and f_y coordinates is an effective tool for any texture characterization. Fourier

transformations of surface profiles are determinate by the use of fast Fourier transform (FFT) procedure. images allow to find distribution of surface heights in inverse space, distinct their periodicity and evaluate anisotropy of surface films in.
In Fig. 13 a and b, the mechanically polished steel SWM7 block profile, and its two-dimensional FFT is presented.



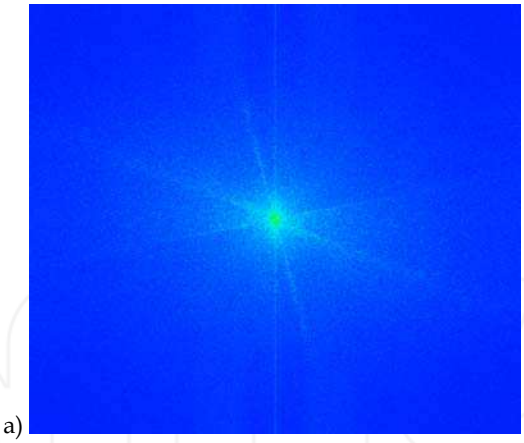
a)



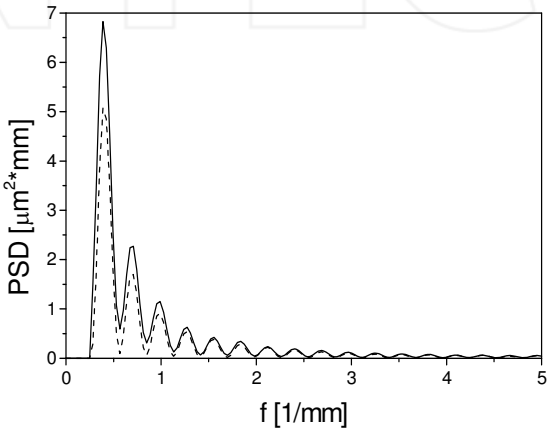
b)

Fig. 13. Surface profile –a and its FT –b of mechanically polished steel SWM7

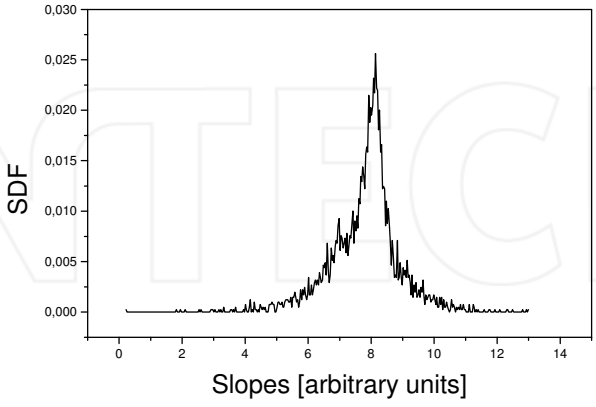
Grooves and scratches visible in Fig. 13a are the result of the polishing process. These features cover the whole surface in a more or less uniform way and overlap. FFT allows toassess the repeatability of the processes forming the periodic inequalities. As may be noticed, in its two-dimensional FFT (Fig 13b) periodic dependence of the PSD along the direction forming an angle 45^0 with the X and Y axes may be distinguished. It is result of interference of spatial waves appearing in X and Y directions. The occurrence of spatial waves interferences indicate the same processes along mutually perpendicular directions.
Fig. 14a shows FFT of optical profile of polyvinylcarbazole (PVK) film deposited on Corning glass 7059, obtained by spin -coating method [Patel et al. 2007], The scanning area was 10x10 mm which enables detecting of surface features in longer space wavelengths.



a)



b)



c)

Fig. 14. FFTprofile of PVK film on Si substrate obtained by means OP BRDF mode –a and one dimensional PSD along X and Y axis –b, relative slopes distribution function for PVK film –c respectively

Because refractive index of PVK for 633 nm light wavelength is nearly the same as Corning 7059 glass [10], almost the whole scattered light originates from upper surface of the film. FFT analysis of optical profiles shown in spatial frequencies, allows to distinguish periodic features and specify interference of spatial waves and to evaluate ratio of surface anisotropy. Because PVK was formed by centrifugal force in spin–coating process the PSD functions obtained for X and Y axes are similar, what is presented in Fig 14b. The determined PSD demonstrates as well periodic structure of surface. Also the relative slope density function (SDF) was determined for this sample (Fig. 14c).
OP measures the local slope of microfacets (local surfaces), the autocorrelation function is given as:

$$ACF(s) = \sigma^2 p(s) \tag{17}$$

This relation may be applied to ACF determination directly from OP measurements if $p(s)$ is known.
Fig. 15a shows FT from optical profile of satinated glass texture used in glass houses. In Fig. 15b the calculated PSD function for x and y spatial frequency is presented.

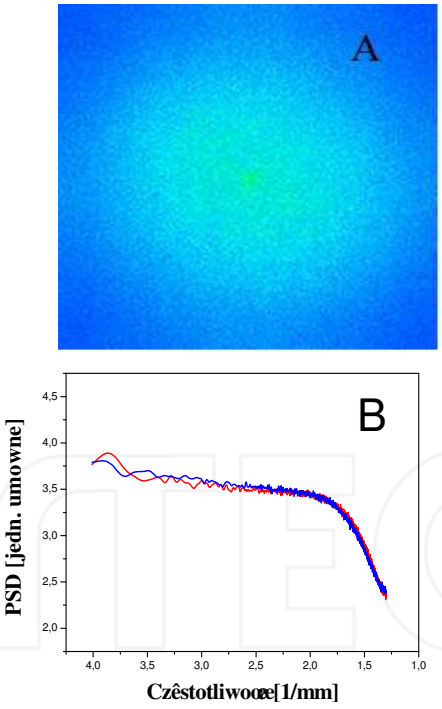


Fig. 15. FT of satinated glass surface and arbitrary one-dimensional PSD for x and y direction

As results from Fig 15.b the PSD for x and y directions are nearly the same. Thus one may be conclude that satinated glass surface exhibits very good isotropy. It is important factor for glass use in green houses.

On the other hand the OP measurements allow to detect anisotropy of surface creation process. Optical profilometry could be applied to waviness detection occurring on surface or to inspection of periodic variation of film thickness. As an example in Fig. 16a the profilometric map of TiO2-SiO2 film on BK7 glass substrate obtained by dip coating sol-gel method is shown.

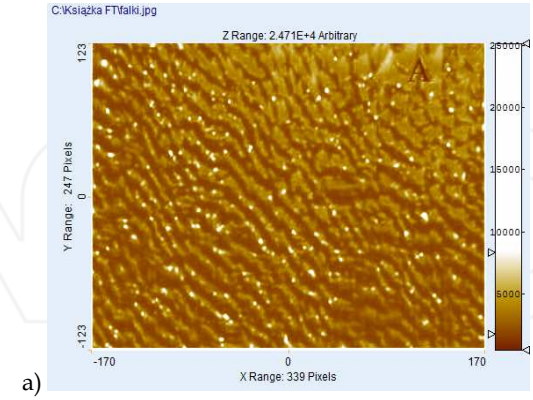


Fig. 16. PO profile of TiO2-SiO2 film on –a and its FFT – b and ACF –c

It presents periodic changes of reflection coefficient for TiO2-SiO2 film. In Fig. 16b the Fourier Transform of OP profile from Fig. 16a is shown. This is due to two harmonic vibrations caused by mechanical elements of technological equipment Elimination of vibration allows to obtain very flat TiO2-SiO2 films [Karasiński 2005].

6. Conclusion

As it is shown in this chapter the optical methods may be used as effective tools for surfaces and interfaces thin films and layers. Surface topographic parameters may be obtained from biredirectional reflectance distribution function and optical profilometry measurements. The BRDF and OP techniques allow to get information about surface topography. From these optical investigations the surface power spectral density (PSD) and autocorrelation (ACF) functions for surfaces and films may be found. Both functions are calculated from the Fourier transformation of surface profiles. The root mean square (rms) surface roughness, slopes and correlation length for studied samples from PSD function have may be determined.
The resolution of BRDF method allows to determine surface parameters in spatial wavelength range from 0.1 to hundreds of micrometers. The OP may be applied as complementary method for measurements of longer spatial wavelengths.
The most effective analysis of the surface statistic is calculating power spectral density or autocorrelation functions from surface profiles in spatial frequency or wavelength domains respectively.
The scattered radiation is a function of highs of irregularities and slopes of microfacets but the sensitivity of PO studies is resulted mainly from sensitivity of detection of slopes change. In PO BRDF mode the technique is sensitive for upper surface of a film when substrate is optically flat. The presented techniques may be also take advantage in control manual painting or varnishing. The presence of long lateral irregularities is often caused by manual or mechanical treatments (i.e paintings, varnishing, manual polishing) and may have periodical nature. The short spatial waves result rather from random process of surface formation and their contribution in total profile is easy to determine from light scattering by the using BRDF method.
If films are transparent and their interfaces are partly correlated, the description of film topography is difficult.
The PO specular mode may be used for transparent film thickness determination if upper and lower interfaces are flat. This technique allows measures thickness of layer If interfaces correlated are in transparent films. If surface are partially correlated the BRDF -PO mode is more preferred for their topography measurements

7. Acknowledgment

This work was supported by Grant N N503 137235 of Polish Science Committee.

8. References

Print Books

Stover J.C., Optical Scattering. (1995). *Measurements and Analysis* (2d ed.), SPIE, ISBN 0-8194-1934-6, Belligham. (b)

Bennett J. M.,. Mattsson L. (1999). *Introduction to surface roughness and Scattering* (2 ed.) SA ISBN 1-55752-609-5, Washington, D.C.

Workmann J. (st ed.) (1998). *Applied spectroscopy*, Academic Press, ISBN 0-112-764070-3.
Whitehouse D. J. (2003). *Handbook of surface and nanometrology*, (1st-ed.) IOP Publishing, ISBN 0-7503-0583-5, Bristol and Philadelphia.
Beckmann P, Spizinochino B. (1963) *The scattering of electromagnetics waves from rough surfaces*, Pergamon Press Oxford, London, New York, Paris:

Papers in Journals

Elson J.M., Bennett J.M., (1979). Vector scattering theory, *Opt. Eng.*, 18, 116.
Church E.L., Jenkinson H.A.,. Zavada J.M, (1977). Measurement of the finish of diamond-turned metal surfaces by differential light scattering, *Opt. Eng.*, 16 360.
Chandley P.J. (1976). Determination of the autocorrelation function of height on rough surface from coherent light scattering, *Opt. Quant. Elect.* 8 329.
Church E. L. (1988). Fractal surface finish, Appl. Opt. 27 1518.
Ohlidal I., (1988). General formulas for the optical characterization of single layers with spectroscopic reflectometry, *J. Mod. Opt.* 35, 1373.
Stover J.C., Serati S.A., Gillespie C.H., (1984). Calculation of surface statistic from light scattering, Opt. Engng, 23 ,406.
Karasiński P. (2005) Influence of technological parameters on the properties of sol-gel silica films, *Opt. Appl.* Vol. 35, No. 1 117.
ANSI/ASME (1985). Standard B46.1-1985, *Surface Texture (Surface Roughness, Waviness, and Lay),* Am. Soc. Mech. Eng., New York.
Elson J. M. (1984). Theory of light scattering form rough surface with inhomogeneous dielectric permittivity. *Phys. Rev. B30* 5460.
Elson J. M. Benett J. M. (1995) Calculation of the power spectral density from surface profile data, *Appl. Opt*., 34, 20.
Stover J.C., Serati S.A., Gillespie C.H., (1984).Calculation of surface statistic from light scattering, *Opt. Engng*, 23 406.
Bendat, J.S., Piersol, A.G. (1971) *Random Data: Analysis and Measurement, Proceedures*, Wiley-Interscience, New York
Bendat, J.S., Piersol, A.G. (1986) Engineering Applications of Correlation and Spectral Analysis*,* (2d ed.), John Willey &Sons, New York.
Nicodemus F. E. (1965). Directional Reflectance and Emissivity of an Opaque Surface *Applied Optics*, 4 767.
Mainsah E., Dong W.P., Stout K.J., (1996) .*Measurement* 17 173.

Papers in Conference Proceedings

Kaczmar J.W., Pietrzak K., Wlosinski W. (2000) The production and application of metal matrix composite materials. *J Mater Process Technol*., 106 58.
Karasiński P. (2005). Influence of technological parameters on the properties of sol-gel silica films, *Opt. Appl.* 35, 117
M.J. Elson, (1992) Theory and Software for Light Scattering from Multilayer Optical Components with Interfacial Roughness, Naval Air Warfare Center, Weapons Division, China Lake, CA, Technical Publication 8084.
Stover J.C. (ed.), (1995). Optical Scattering in the Optics, Semiconductor and Computer Disk Industries, Proc. SPIE 2541. (a)
Patel S., Ramrakhiani M., Bisen D., P. (2007) *J. Appl. Polym. Sci*, 104, 722.

Church E. L., Leonard T.A., Takacs P. Z. (1989). The prediction of BRDF from surface profile measurements. Proc. SPIE 1165 196.

Brown A.J.C., Breitmeier U., (1988) Industrial application of an optical profilometer. Proceedings SPIE, 954.

Lipiński M., Cichoszewski J. (2008) in Proc. 23nd EU PVSEC, (eds. G. Willeke, H. Ossenbrink, and P. Helm, WIPRenewable Energies, Munich, Germany, 1911.

18

Application of FTIR Spectroscopy to Agricultural Soils Analysis

Linker Raphael
Faculty of Civil and Environmental Engineering
Technion – Israel Institute of Technology
Israel

1. Introduction

Soil is a very complex medium that contains minerals, organic matter, micro-organisms, air and water. Soil is one of the most important factors for agriculture and some soils are deemed more fertile than others. Soil fertility is directly related to factors such as nutrients concentrations or availability, organic matter content, acidity, moisture, etc., as well as to agricultural practice such as till vs. no-till (Desbiez et al. 2004) (Sá et al. 2009). Ensuring soil fertility is a basic requirement for any form of sustainable agriculture, yet in practice this seemingly trivial goal is very difficult to achieve due not only to the complexity of the soil medium itself, but also due to the complexity of the soil-crop-air interactions and to the fact that some processes require years before having any visible impact. Various recent studies have shown that soil fertility is declining in many farmlands due mainly either to inadequate farming practices (Gobeille et al. 2006), insufficient fertilization, in which case the soil reserves are depleted, or over-fertilization that results in pollution to the groundwater or toxic accumulation of chemicals in the soil. Avoiding such under- or over-fertilization is the chief goal of the so-called *precision fertilization* concept, which aims at delivering exactly the amount of nutrients required to sustain optimal growth of the crop. One of the main obstacles to the application of the precision fertilization concept, or the more general *precision farming* concept, is soil heterogeneity. Hence, although it is technically possible to perform a wide range of analyses and derive a soil fertility or health index such as the one proposed by Idowu et al. (2008), most of the required analyses are time-consuming, which in practice makes it impossible to map the soil properties of a field with the required spatial and/or temporal resolution. The need for fast and cheap methods that would enable the analysis of a large number of samples has been stressed in numerous studies (Viscarra Rossel&McBratney 1998b) and infrared spectroscopy has long been recognized as one of the most promising techniques (e.g. McCarty&Reeves 2006; Mouazen et al. 2007; Cécillon et al. 2009). As in other applications, the initial works were conducted mainly in the near-infrared (NIR) range for which instruments were readily available. Technological achievements in the mid-infrared (mid-IR) range during the last decade have made this spectral range much more attractive and an increasing number of soil studies are conducted using this spectral range. Although there is no doubt that near-infrared spectroscopy can be very useful for some analyses, several comparative studies have shown

the superiority of the mid-infrared techniques in most cases (McCarty&Reeves 2006; McCarty et al. 2002; Viscarra Rossel et al. 2006; Canasveras et al. 2010) and therefore the present chapter will focus solely on mid-IR techniques.

2. Mid-Infrared techniques commonly used for soil analysis

Infrared (IR) spectroscopy is based on the interaction of molecules with electromagnetic energy in the infrared spectral region, which is in the wavelength range of 0.8-1000 µm. This IR range is commonly divided into four regions, labeled near-, mid-, thermal- and far-infrared, respectively (Figure 1). The particularity of the mid-infrared range is that it includes the so-called fundamental vibrations of the molecules. When a molecule absorbs IR radiation at frequencies matching that of its own molecular vibrations, it results in an increase of the amplitude of the vibrations at these frequencies. Since each frequency corresponds to a given amount of energy and a specific molecular motion (e.g. stretching, bending or contracting of chemical bonds), the mid-IR spectrum can reveal the kind of molecular motions and bonds (functional groups) that are present in the molecule and hence can serve as a unique fingerprint of a specific compound. Furthermore, most functional groups have characteristic IR absorption bands that do not change much from one compound to another. By comparison, the near-infrared range is dominated by overtones and combinations of these fundamental vibrations, which makes the interpretation of the NIR spectra much more difficult.

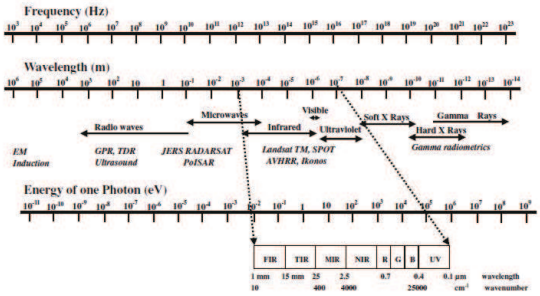


Fig. 1. The electromagnetic spectrum with emphasis on the ultra-violet – infrared range Reproduced from Viscarra Rossel et al. (2006).

Nowadays, Fourier Transform Infrared (FTIR) is the preferred method for mid-IR spectroscopy as it provides quantitative information in a rapid and accurate fashion. A typical FTIR spectrometer obtains an infrared spectrum by collecting the interferogram of a sample signal, which contains all the infrared frequencies, applies the Fourier transform to the digitized signal, and outputs the spectrum. Such a FTIR spectrometer relies on an interferometer, which splits to radiation beam into two beams that are recombined after a path difference has been introduced (Griffiths&de Haseth 1986). The most common interferometer is the so-called Michelson interferometer, which consists of a beam splitter located between two perpendicular mirrors, one of which can move along an axis perpendicular to its plane (Figure 2). At the beam splitter, the radiation beam from the infrared source is partially reflected to the fixed mirror and partially transmitted to the moveable mirror. The moveable mirror is moved in a highly controlled fashion to create the

path difference between the two beams. After the beams return to the beam splitter, they interfere and are again partially transmitted and partially reflected to the detector. Due to the effect of the interference, the intensity of each beam passing to the detector and returning to the source depends on the path-length difference between the two beams. The two beams can undergo constructive interference, destructive interference or a combination of both, depending on the path-length difference. Constructive interference, which yields a maximum detector signal, occurs when the optical path difference is an integer multiple of the wavelength. The variation in the energy that reaches the detector as a function of the path difference yields the interferogram, which is the integral of all interference patterns produced by each wavelength. The detected interferogram can not be interpreted directly, but has to be "decoded" using the well-known Fourier Transformation (Griffiths&de Haseth 1986). Fourier transform is typically thought of as decomposing a signal into its component frequencies and their amplitudes. The Fourier transform is an integral transform that re-expresses a function in terms of sinusoidal basis functions, i.e. as a sum or integral of sinusoidal functions multiplied by some coefficients ("amplitudes"). The general idea is that a multiplication of the input waveform of unknown amplitude and frequency (interferogram signals) by a known reference frequency of unity amplitude (the analyzing wave) can give us the unknown amplitude and original frequency. Thus, by using a frequency-adjustable analyzing wave, each digitized point of the interferogram can be transformed from the time (or optical retardation) domain to the frequency domain, which results in the IR spectrum. When a single interferogram is thus Fourier-transformed, a so-called single-beam spectrum is generated, which is the raw detector response versus the wavelength. In order to produce the absorption spectrum of a sample, the sample single-beam spectrum must be normalized against a background spectrum taken with no sample in the beam path. The absorption spectrum can be presented equivalently as a transmittance ($T= I/I_0$) or absorbance ($A= \log_{10} I_0/I$) spectrum, where I is the intensity measured with the sample in the beam and I_0 is the intensity measured from the background spectrum.

Direct transmittance is the oldest and most straightforward spectroscopic technique, which is based on the absorption of the IR radiation as it passes through the sample. Clearly, this technique is applicable only to samples that do not absorb all the incoming IR energy and are sufficiently transparent in this spectral range. For highly absorbent samples, such as soils, it is necessary to prepare a pellet that embeds the soil sample in a transparent matrix, most usually KBr. The pellet preparation involves grounding 2-3 mg of soil with ~1 g of KBr using a mortal and pestle and using a hydraulic press and die to create a thin, IR-transparent disk. The main advantage of this technique is that it yields very clear and information-rich signals. Its main disadvantages are the lengthy preparation required to prepare the pellets and the fact that it is difficult to obtain quantitative results. Also, since only a few mg of soils are used for the preparation of each pellet, the resulting spectrum may not be representative the bulk soil. For these reasons, transmittance measurements are rather rarely used in soil analysis and the reflectance and photoacoustic methods described below are prefered.

2.1 Diffuse reflectance

Nguyen et al. (1991) pointed out to two main drawbacks of the transmittance technique in addition to the time-consuming sample preparation: (1) possible reaction of the sample with the halide matrix, and (2) scattering and/or total absorption for samples with high concentrations and large particles relative to the infrared wavelengths. To circumvent these limitations, Nguyen et al. (1991) introduced the use of diffuse reflectance (DRIFT) for soil

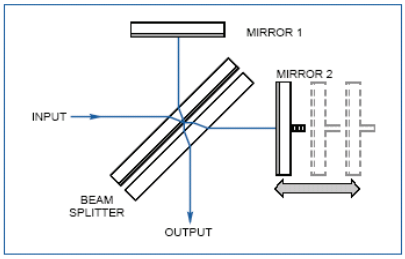


Fig. 2. Schematic description of a Michelsen interferogram.

analysis. While the study of Nguyen et al. (1991) was purely qualitative and restricted to band assignments, the subsequent articles of Janik et al. (1995) and Janik&Skjemstad (1995) showed that mid-IR DRIFT could be used to quantify various soil components. Since then, the use of mid-IR DRIFT has been investigated in numerous soil studies (see excellent review paper of Reeves III 2010), and it is arguably the mid-IR spectroscopic method most commonly used for soil analysis. DRIFT spectroscopy is mainly concerned with radiation emerging from a non-mirror or matt surface after it has undergone absorption, refraction, reflection and scattering in the bulk material. DRIFT spectra are subject to nonlinear scaling of the intensity which reduces the intensity of strongly absorbing bands in comparison to weaker bands. Thus, low intensity bands appear to be enhanced in DRIFT spectra. The specular component becomes increasingly significant at high sample concentrations and for large particles, and the occurrence of strong specular reflection explains for instance why the presence of inorganic carbon interferes with the development of calibrations for soil organic carbon (Reeves et al. 2005). In order to avoid such non-linearities and spectral distortions, the use of samples diluted with KBr has long been advocated and such KBr dilution is still used in many studies. However, using KBr-diluted samples increases the sample preparation time and may cause interferences due to ion exchange between the sample and the KBr matrix or adsorption of water onto the KBr (Bertrand et al. 2002). Janik&Skjemstad (1995), Janik et al. (1998) and Reeves et al. (2001) have shown that quantitative analysis can also be performed on mid-IR spectra of neat (undiluted) soil samples. These findings were further supported by the comparative study of Bertrand et al. (2002) who compared results obtained using neat and KBr-diluted samples and concluded that KBr dilution does not improve the accuracy of the measurements.

2.2 Attenuated total reflectance

In the Attenuated Total Reflectance (ATR) method, the IR radiation propagates through a crystal with a high refractive index that is in contact with the sample (Figure 3). Mirrors are used to direct the IR beam toward the crystal at an angle that exceeds the critical angle for internal reflection, so that the radiation undergoes multiple reflections within the crystal.

This critical angle θ_C depends on the refractive indices of the sample and ATR crystal according to

$$\theta_C = \mathbf{sin}^{-1}\left(n_2 / n_1\right) \tag{1}$$

where n_1 and n_2 are the refractive indices of the crystal and the sample, respectively. Due to the quantum mechanic properties of the light, the electromagnetic field extends beyond the

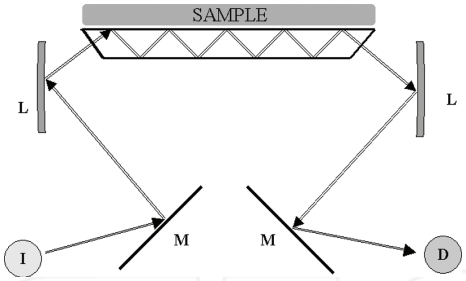


Fig. 3. Attenuated total reflectance (ATR) configuration. I denotes the incoming light (from the interferometer), D denotes the detector, L and M are lenses and mirrors. Reproduced from (Etzion et al. 2004)

crystal surface for a short distance known as the *evanescent field*. If a sample is applied directly onto the surface of the ATR crystal, some of the IR radiation (i.e., the evanescent wave) is absorbed by this sample, so that the sample absorbance spectrum can be obtained. The evanescent wave decays exponentially with the distance from the surface of the crystal. The depth of penetration of the evanescent wave d_p is defined as the distance from the crystal-sample interface at which the intensity of the evanescent wave decays to 37% (1/e) of its original value, and is given by

$$d_p = \frac{\lambda}{2\pi n_1 \sqrt{\left[\sin^2\theta - \left(n_2/n_1\right)^2\right]}} \tag{2}$$

where λ is the radiation wavelength. The penetration depth and the total number of reflections along the crystal can be controlled to some extend either by varying the angle of incidence of the radiation (θ) or through the selection of the crystal material. The penetration depth is typically less than 10 μm, so that very good contact between the sample and the crystal is critical in order to obtain reliable and reproducible results.

The first use of ATR for soil analysis was reported by Ehsani et al. (2001), who attempted to determine nitrate concentration using ATR spectra of dry soil samples. This study demonstrated the difficulty of obtaining adequate contact between the soil and the ATR crystal, but also showed how the results could be greatly improved by using a soil paste or slurry. Such an technique was further developed by Shaviv et al. (2003), Linker et al. (2004), Linker et al. (2005) and Borenstein et al. (2006) who used samples consisting of soil pastes close to water saturation. Since nitrate is highly soluble in water and is not fixed in the soil matrix, all the nitrate is present in the liquid phase of the soil paste, which has several advantages. First, since the paste moisture content is less than 1 g [H_2O]/g [soil], the nitrate concentration in the liquid phase is higher than in the dry soil. Second, much better contact is obtained between the ATR crystal and the liquid phase than could be obtained between the ATR crystal and the dry soil. Finally, the original moisture of the soil sample has no influence on the measurement. However, water exhibits very strong absorbance bands in the mid-IR range, which may distort or hide bands of interest. This is illustrated in Figure 4 which shows the spectra of de-ionized water, de-ionized water with 1000 mg[N]/kg[water] nitrate, and a paste of sandy soil with 870 mg[N]/kg[water] nitrate. Distortion and shift of

the spectra can be observed, especially in the regions indicated by circles. The magnitudes of these changes are comparable with the size of the nitrate signal and would cause significant inaccuracies if straightforward estimation of nitrate was performed. Therefore an accurate water subtraction procedure such as the one developed by Linker et al. (2004, 2005) should be applied to the spectra prior to the quantitative analysis.

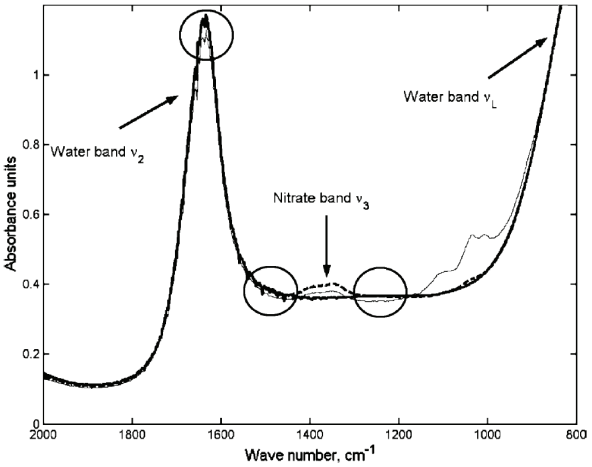


Fig. 4. Spectra of de-ionized water, de-ionized water with 1000 mg[N]/kg[water] nitrate, and a paste of sandy soil with 870 mg[N]/kg[water] nitrate. The circles indicates the regions in which shifts and distorsions of the spectra are most clearly visible. Reproduced from Linker et al. (2005).

2.3 Photoacoustic spectroscopy

Fourier transform infrared photoacoustic spectroscopy (FTIR-PAS) differs from the previous methods in the sense that it relies on complete absorption of the incoming radiation by the sample. The sample is placed in a sealed enclosure that is purged with He to avoid atmospheric interferences and to which is connected a highly sensitive microphone which records the pressure waves that result from the local heating induced by the absorbed radiation (McClelland et al. 2002) (Figure 5). The recorded spectrum depends on the absorption properties of the sample, its thermal diffusivity, and the thermal penetration depth. Therefore, photoacoustic spectra are more difficult to interpret than reflectance ones and there is no one-to-one correspondence between the spectra obtained with both techniques. Also, photoacoustic spectra are much more prone to measurement noise. Nonetheless, since this method can be used with highly absorbing samples without any pretreatment, it is well suited for soil samples analyses and its use has been reported in several recent studies (Du et al. 2008a, 2008b, 2009b).

3. Data processing

Due to the large amount of data generated by FTIR spectrometers and the complexity of the spectra, it is imperative to use chemometrics procedures to analyze the data. A detailed

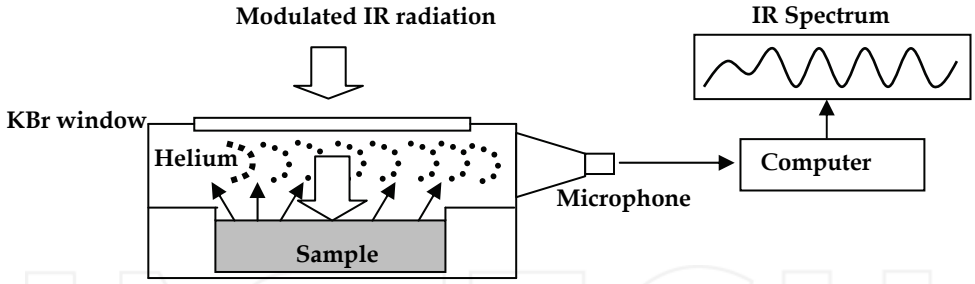


Fig. 5. Schematic description of FTIR photoacoustic spectroscopy. Reproduced from Du et al. (2008a).

description of such procedures can be found in various textbooks such as Brereton (2003) or Chau et al. (2004). Partial Least Squares (Wold et al. 2001) is the procedure most commonly used for quantitative determination of one or several soil components (see for instance the excellent compilation table in Viscarra Rossel et al. (2006)). PLS is an easy-to-use straightforward procedure that is available in standard statistical and chemometrics softwares, which probably explains its popularity. The main limitation of PLS is that it assumes linearity of the data, but the good results reported in most soil studies show that this does not appear to be too restrictive an assumption when analyzing soil spectra. Exceptions to this generally linear behavior were reported for instance by Janik&Skjemstad (1995) and Janik et al. (2009), especially at very high concentrations. However, Janik et al. (2009) reported that in those cases where PLS was not capable of providing a good prediction model, adding a non-linear element in the form of a neural network did not improve the results significantly.

The *a priori* assumption of linearity can be avoided altogether by using procedures such as wavelet decomposition and neural networks. An example of such a procedure is depicted in Figure 6. In this procedure, wavelet decomposition is used to compress the data and obtain a low-dimensional representation of the spectra. Although wavelet decomposition by itself does not produce a compressed representation of the original data, data reduction can be achieved by eliminating the wavelet coefficients that do not contain valuable information. Various approaches have been reported in the literature for selecting the most relevant coefficients, such as eliminating all "small" coefficients using for instance either thresholding (Kai-man Leung et al. 1998; Ehrentreich 2002), entropy (Kai-man Leung et al. 1998), mutual information (Alsberg et al. 1998), maximum likelihood (Leger&Wentzell 2004), or genetic algorithms (Depczynski et al. 1999), or retaining only the coefficients with the highest variance (Trygg&Wold 1998) as depicted in Figure 6. Once data compression has been achieved, the remaining coefficients can be used as input variables for a neural network that creates a non-linear mapping between these inputs and the property (or properties) of interest. The neural network can also be replaced by a simpler linear regression as in Viscarra Rossel&Lark (2009), in which case a hybrid model that contains both linear and non-linear stages is obtained.

Since the wavelet transform decomposes the signal into components at different scales (or loosely speaking, components of different widths), it is a very powerful tool for resolving overlapping bands and separate the bands of interest from the background and interferences. For instance, Jahn et al. (2006) used such an approach to distinguish between the strongly-overlapping absorbance bands of nitrate and calcium carbonate in the ATR spectra of calcareous soils (Figure 7).

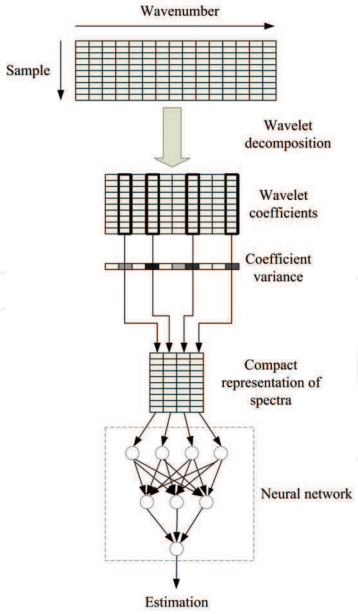


Fig. 6. Schematic representation of a prediction model combining wavelet decomposition of the spectra and a neural network

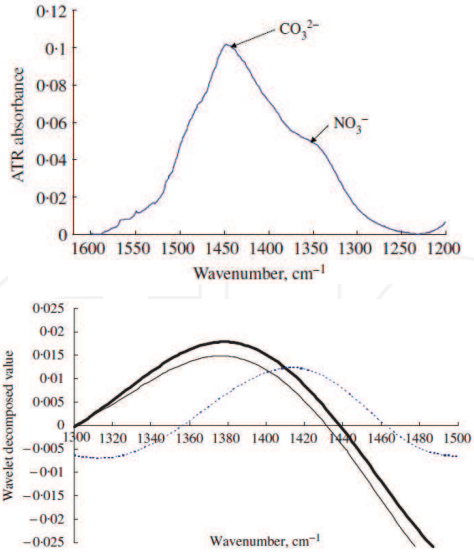


Fig. 7. Overlapping bands of nitrate and carbonate in ATR spectra of calcareous soils (Top frame) and the wavelet decomposed values at scale 2 (dashed, attributed to carbonate) and at scale 3 (attributed to nitrate,14ppm and 120ppm N-NO_3) (Bottom frame). Reproduced from Jahn et al. (2006).

4. Most common applications

4.1 Estimation of soil nutrients and organic matter

Soil nutrients, such as C, N, P, K, S, Ca and microelements, play a primordial role in the development of agricultural crops and hence determination of their concentrations is crucial to the application of the precision agriculture concept. Viscarra Rossel et al. (2006) compiled an excellent review of the studies dedicated to the estimation of these parameters using diffuse reflectance. Table 1 present a non-exhaustive list of studies that focused on the determination of carbon, total nitrogen, nitrate, potassium, phosphorus and organic matter using diffuse reflectance, ATR and photoacoustic spectroscopy. Table 1 clearly shows that carbon, and in particular organic C, is the property most commonly determined, using DRIFT. Most studies reported very good results with correlation coefficients (R^2) between the actual and estimated values higher than 0.90. A noticeably much poorer result was reported by Reeves&Smith (2009) (R^2=0.60). This latter study was not restricted to agricultural soils as the previous studies (some of which included a number of samples similar to that of Reeves&Smith (2009)), but included soil samples from national forests, rangelands, woodlots, etc. and Reeves&Smith (2009) hypothesized that this wide range of soil uses was responsible for the poor performance of the regression models. Acceptable results were also reported for estimation of total nitrogen (R^2>0.80), using either DRIFT or photoacoustic spectroscopy. For nitrate, Viscarra Rossel et al. (2006) did not find any correlation between DRIFT spectra and nitrate concentration. This finding contrasts sharply with the good results that can obtained using the ATR technique, for which the typical determination errors range from 5 to 25 mg[N]/kg[dry soil] for realistic agricultural field concentrations. The largest determination errors correspond to calcareous soils for which nitrate determination is hindered by the absorbance band of carbonate that overlaps the nitrate band (Linker et al. 2004, 2005, 2006; Borenstein et al. 2006; Jahn et al. 2006). For potassium, phosphorous and organic matter conflicting findings have been reported. Bertrand et al. (2002) reported that DRIFT could be used to estimate potassium concentration (R^2=0.85). Acceptable results were also reported by Du et al. (2009b) who used photoacoustic spectroscopy (R^2=0.79). McCarty&Reeves (2006) and Reeves&Smith (2009) obtained relatively poor but still potentially useful results using DRIFT (R^2 of 0.60 and 0.76, respectively), but Viscarra Rossel et al. (2006) concluded that DRIFT spectroscopy was not suitable for estimating potassium concentration (R^2=0.40). For phosphorous, Janik et al. (2009) and Du et al. (2009b) reported R^2 values of 0.87 and 0.81 using DRIFT and photoacostic spectroscopy, respectively. However, Bertrand et al. (2002), McCarty&Reeves (2006) and Viscarra Rossel et al. (2006) did not find a relationship between DRIFT spectra and phosphorous (R^2<0.40). For organic matter, very good results were obtained using photoacoustic spectroscopy (Du et al. 2007, 2009b) and DRIFT (Masserschmidt et al. 1999), with R^2 values above 0.90 in all three studies. However, very poor correlation between DRIFT spectra and organic matter content was reported by Canasveras et al. (2010).

The somewhat conflicting results reported in some of the DRIFT studies emphasize the need for standardization of the analyses (both analytical reference analyses and DRIFT spectroscopy) and underscore the problem of transferability of the results (See Section 5.1). Photoacoustic spectroscopy appears to be a very promising technique which should certainly be investigated in further studies.

	Diffuse reflectance	Attenuated total reflectance	Photoacoustic
Organic C	Janik&Skjemstad (1995); Janik et al. (1998); McCarty et al. (2002); Reeves et al. (2002); Reeves (2009, 2010); Reeves&Smith (2009) ;McCarty&Reeves (2006) ;Viscarra Rossel et al. (2006, 2008); Viscarra Rossel&Lark (2009) ;Janik et al. (2009) ;Zimmermann et al. (2007) ;McBratney et al. (2006)		
Total C	McCarty et al. (2002) ;Reeves et al. (2001, 2002); Reeves (2009, 2010); Viscarra Rossel et al. (2008); Minasny et al. (2009); Reeves&Smith (2009)		
Total N	Janik&Skjemstad (1995); Reeves et al. (2001); McCarty&Reeves (2006); Viscarra Rossel et al. (2008)		Du et al. (2009b)
Nitrate	Viscarra Rossel et al. (2006) ;Verma&Deb (2007)	Ehsani et al. (2001); Shaviv et al. (2003); Linker (2004); Linker et al. (2004, 2005, 2006); Borenstein et al. (2006); Jahn et al. (2006);	
Potassium	Bertrand et al. (2002); McCarty&Reeves (2006); Viscarra Rossel et al. (2006); Minasny et al. (2009); Reeves&Smith (2009)		Du et al. (2009b)
Phosphorus	McCarty&Reeves (2006); Viscarra Rossel et al. (2006) ; Janik et al. (2009); Minasny et al. (2009)		Du et al. (2009b)
Organic matter	Masserschmidt (1999); Canasveras et al. (2010)		Du et al. (2007, 2009b)

Table 1. List of studies that used reflectance and photoacoustic spectroscopy for the estimation of soil carbon, total nitrogen, nitrate, potassium phosphorus and organic matter

4.2 Determination of soil moisture

Water in itself is essential for plant growth and its presence in the soil influences numerous properties such as the availability of oxygen and nutrients, microbial activity, swelling and shrinking of clays or soil stability. Since the methods traditionally used to estimate water-related soil properties are time consuming, much research has been devoted to the development of alternative methods. In particular, very good results have been reported for some of these properties using near-infrared spectroscopy under laboratory conditions (Mouazen et al. 2005, 2006; Viscarra Rossel&McBratney 1998a) and encouraging results have been reported for in-situ field measurements (Mouazen et al. 2005). A summary of the main results obtained with mid-IR DRIFT spectroscopy is presented in Table 2. With the exception of Dataset #1 in Tranter et al. (2008), mid-IR spectroscopy was capable of estimating to some extent water retention at -10kPa and -1500kPa, with R^2 values ranging from 0.64 to 0.81 and from 0.66 to 0.89, respectively. However, the study of Bertrand et al. (2002) clearly shows that KBr dilution of the samples or high $CaCO_3$ concentration deteriorate the results. Although this has not been specifically verified, the very poor results obtained by Tranter et al. (2008) on Dataset #1 were most probably due to the fact that the spectra were recorded from intact soil cores rather than from grounded samples as in all the other studies. Regarding moisture of air-dried and oven-dried soil, the few results available are very good, unless the soil has a high $CaCO_3$ concentration.

4.3 Soil characterization

The soil physical structure affects various physical, chemical and biological processes such as for instance water infiltration and retention or root penetration and proliferation. With this respect, clay is the component most often studied by mid-IR spectroscopy, although some studies also attempted to estimate sand or silt content (McBratney et al. 2006; Viscarra Rossel et al. 2008). For clay, R^2 values of 0.85 or more were reported by Janik&Skjemstad (1995), Viscarra Rossel et al. (2008) and Viscarra Rossel&Lark (2009). Reasonably good results were reported by Janik et al. (2009), Canasveras et al. (2010) and Janik et al. (2009), with R^2 value of 0.84 and 0.82, respectively, but lower R^2 values (0.74-0.79) were reported by Bertrand et al. (2002), McCarty&Reeves (2006), Minasny et al. (2009) and McBratney et al. (2006). No explanation has been offered to explain such a variability of the prediction performances. In addition to the previous DRIFT studies, Du et al. (2007) reported very good results (R^2>0.90) using photoacoustic spectroscopy and Du et al. (2008b) further showed that this method can be used to differentiate between different types of clays. Another important soil component that can be estimated by mid-infrared spectroscopy is carbonate. In this case the results are much more consistent, with R^2 values of 0.95 or more reported in the studies of Janik&Skjemstad (1995), Bertrand et al (2002), Canasveras et al. (2010) and Du et al. (2008b). Linker et al. (2006) and Du et al. (2008a) showed that the spectral information related to carbonate and other soil constituents can be used to discriminate between soils, and Linker et al. (2006) further showed that such a automatic discrimination can improve the estimation of nutrients such as nitrate.

5. Current challenges and emerging applications

5.1 Spectral libraries and model transferability

The results presented in the previous Section clearly demonstrated that mid-IR spectroscopy can be used to determine accurately a series of soil characteristics and properties under

Study	Water retention at -10kPa or 120cm	Water retention at -1500kPa	Moisture of air-dried soil	Moisture of oven-dried soil
Bertrand et al. (2002)	0.92‡		0.92	
	0.84‡ (KBr diluted samples)		0.86 (KBr diluted samples)	
	0.74‡ (CaCO3>2%)		0.68 (CaCO3>2%)	
McBratney et al. (2006)	0.64†	0.66†		
Janik et al. (2007)	0.79	0.84		
Tranter et al. (2008) (Dataset #1 – intact cores)	0.02	0.50		
	0.48†	0.71†		
Tranter et al. (2008) (Dataset #2 – crushed samples)	0.85	0.89		
Janik et al. (2009)	0.81	0.89		0.89

Table 2. R^2 values between the actual soil moisture or water retention and the one estimated from mid-IR spectra. The results were obtained using neat ground samples unless otherwise indicated. † indicates that results obtained indirectly using properties estimated by mid-IR spectroscopy and pedotransfer functions. ‡ indicates that the result correspond to water retention at 120cm water tension

laboratory conditions. However, two questions still remain mostly unanswered: (1) could a regression model developed by one group of researchers be readily applied by other researchers and (2) would it be possible to establish spectral libraries that combine spectra from different sources? Such issues are of course not restricted to mid-IR spectroscopy but are also relevant to NIR spectroscopy (Cécillon et al. 2009). These two issues are related since in the absence of such libraries each research team tends to work with spectra recorded "in-house" using experimental procedures that are, at least to some extend, specific to that specific laboratory. Furthermore, due to the extensive work required to collect soil samples and to perform detailed chemical analyzes of these samples, most studies are still based on soil samples belonging to a specific geographic area. The issue of mid-IR model transferability was recently addressed by Minasny et al. (2009) who concluded that "Due to the differences in the origins of the soil and in the laboratory measurement procedures, calibration functions from one region do not perform well on another one". They further emphasized that in their opinion the cause for this lack of transferability lies mainly in the difference in laboratory techniques, different analytical procedures, and inter-laboratory

bias. The creation of a library of mid-IR DRIFT spectral of soils was considered by Viscarra Rossel et al. (2008) who emphasized three main prerequisite for the development of such a library: (1) it should describe adequately the soil variability in the region in which it is to be used, (2) experimental procedures, and in particular sample handling, preparation, storing and scanning must be standardized and (3) the analytical data used as reference should be acquired using reliable and accredited analytical procedures. This issue was also addressed by Reeves III in his recent review on the use of NIR and mid-IR DRIFT for soil analysis (Reeves III 2010) in which he emphasized that "Spectral libraries for quantitative analysis will be useless, or worse deceptive, if users base results on combining spectra from differing sets of samples that have not been standardized." However, as the study of Reeves&Smith (2009) clearly demonstrated, standardization in itself does not guarantee accurate results, even for properties that seem trivial to determine such as organic carbon. In this case, Reeves&Smith (2009) explained their poor results by the fact that their database included soils of various usages rather than to be restricted to agricultural soils. Finally, it must be noted that since specular reflection occurs for both organic and non-organic fractions of soil, libraries of spectra collected by other means than diffuse reflectance would be largely useless for comparing mineral spectra to soil spectra (Reeves et al. 2005). All these factors underline the complexity of the task at hand if one was to contemplate the creation of a truly universal soil spectral library.

5.2 Indirect estimation of additional soil properties

In addition to those properties that can be measured directly by mid-infrared spectroscopy, there is a large number of important soils properties that are not associated with distinct absorbance bands. In such cases McBratney et al. (2006) propose to use the estimates obtained by mid-IR spectroscopy in an inference system that uses pedotransfer functions. This approach is illustrated in Figure 8: Once regression models based on mid-IR spectra have been calibrated for properties such as clay, silt, sand and organic carbon, these models are used to estimate the same properties of a new sample, together with their respective uncertainties. These values are then used to predict the bulk density, the water content at field capacity and wilting point, and finally the available water capacity.

An important feature of this approach is that the determination errors are propagated through the various estimation stages so that a confidence interval can be associated with each prediction (Figure 9). Tranter et al. (2008) showed how such an approach improved considerably the estimation of soil water retention, which was very poorly estimated directly from the mid-IR spectra (See Table 2). Minasny&McBratney (2008) also reported more parsimonious models and greater accuracy when using an inference system then when using straightforward PLS regression on the mid-IR spectra. Such an approach may help solving the "model transferability" problem discussed in the previous section and further work should be devoted to validating and improving this interesting approach.

5.3 On-the-go sensing and combination with GIS

In order to achieve "precision farming", the information obtained from soil analyzes must be combined with a Geographic Information System (GIS) platform in order to create overlaying maps of soil characteristics, water content, nutrient availability, yield, etc. In addition to measurement accuracy, a key issue is the number of sampling points and the spatial resolution of the measurements. The development of "on-the-go" sensors is of

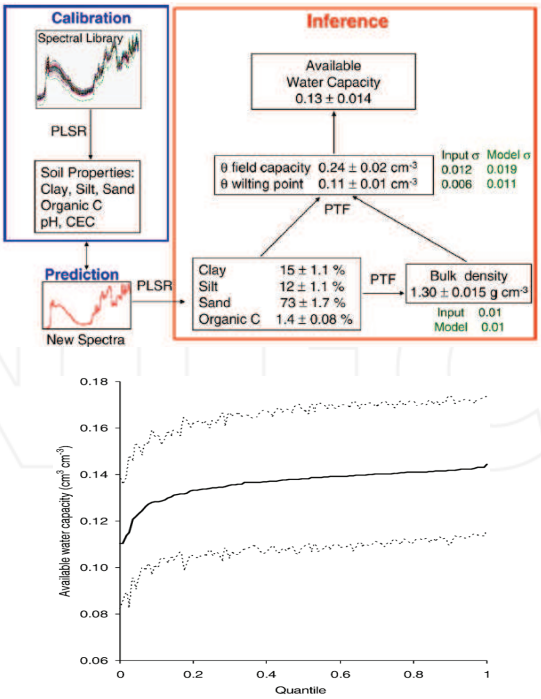


Fig. 8. Top frame: Schematic description of the combination of mid-IR spectroscopy and inference models as suggested by McBratney et al. (2006). Bottom frame: Quantiles of the available water capacity estimated according to the inference scheme. Reproduced from McBratney et al. (2006).

primordial importance (Viscarra Rossel&McBratney 1998b) and considerable research efforts have been devoted to the development of fast, fully automated and geo-referenced measurement systems (Sibley et al. 2009; Mouazen et al. 2007; Christy 2008; Adamchuk et al. 2004, 2005; Sinfield et al. 2010). Among the spectroscopy-based methods, the use of NIR has been investigated more actively, mostly due to the lower price and higher robustness of the equipment. Indeed, in their review paper Sinfield et al. (2010) concluded that the "the FT-IR ATR technique, while very accurate and fast, makes use of expensive and delicate equipment (e.g., moving mirror in interferometer) which is not readily amenable to an on-the-go setting" but that development of sensors based on a limited set of wavebands, as suggested by Linker (2004) and Jahn et al. (2006), could "enhance the potential to apply reflectance based approaches in the field by limiting the sophistication of requisite equipment". At any rate, considerable improvements are still required before large-scale implementation could be considered.

5.4 Monitoring of N isotopes

Nitrogen, either naturally present in the soil or added as fertilizer, may undergo a series of complex transformations. These transformations are interdependent and depend on a large number of variables, so that isolating specific processes or pathways is very challenging. One of the methods used to study these processes consists of enriching the soil with stable

^{15}N isotopes and monitoring the concentration of ^{15}N in the by-products of the various reactions (Stevens et al. 1993; Stevens&Laughlin 1994; Baggs 2008). Such studies require the quantification of ^{14}N and ^{15}N, which is done by isotope ratio mass spectrometry (IRMS). However, expensive and laborious preparation of the samples is required to convert the N species into forms suitable for the IRMS measurements and the method can not realistically be used to analyze a large number of samples during an experiment. Du et al. (2009a) showed how mid-infrared spectroscopy could provide a much faster and cheaper alternative for N-isotope tracing, albeit with much lower accuracy than IRMS. The method is based on the observation that since the mid-infrared range corresponds to the fundamental vibrations of a molecule, which depend on the atoms' mass, the absorbance bands of ^{14}N-based and ^{15}N-based compounds are slightly shifted relative to each other, so that the concentration of each species can be estimated. This spectral shift between ^{14}N-NO_3 and ^{15}N-NO_3 is shown in Figure 9.

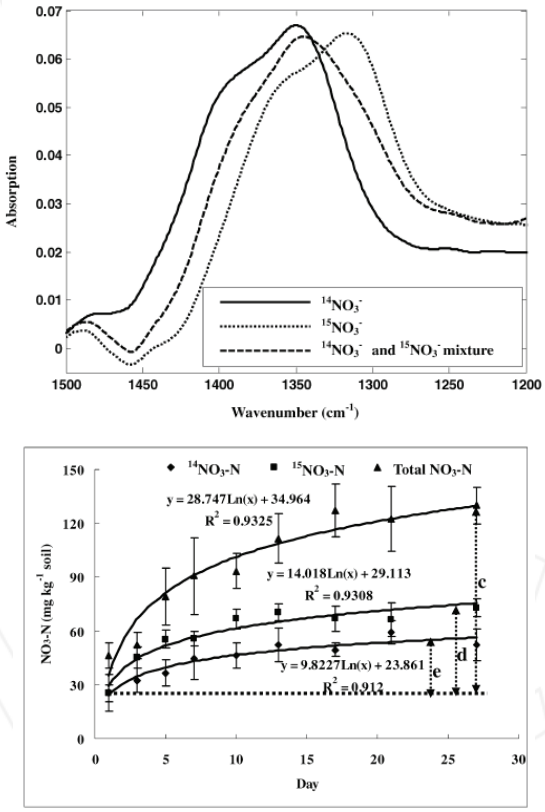


Fig. 9. Top frame: FTIR-ATR spectra of ^{14}N-NO_3 and ^{15}N-NO_3 in water. Bottom frame: Nitrate formation during an incubation experiment as estimated from FTIR-ATR measurements. Nine sub-samples were analyzed at each sampling point. ''c'' is the total amount of nitrate formed. ''d'' and ''e'' are the formed ^{15}N-NO_3 and ^{14}N-NO_3, respectively. Reproduced from Du et al. (2009a).

Using PLS regression models calibrated using saturated soil pastes spiked with $^{14}N\text{-}NO_3$ and $^{15}N\text{-}NO_3$, Du et al. (2009a) were able to estimate the $^{14}N\text{-}NO_3$ and $^{15}N\text{-}NO_3$ concentrations in a non-calcareous soil with determination errors of less than 6 mg[N]/kg[dry soil], which enabled them to monitor separately the formation of $^{14}N\text{-}NO_3$ and $^{15}N\text{-}NO_3$ during an incubation experiment (Figure 9). Although the measurement uncertainties reported in Figure 9 may seem rather large, the low cost and short time required for the FTIR-ATR measurements would make it possible to perform and average more measurements than as done in the study of Du et al. (2009a), and thus reduce the measurement uncertainties. After further validation of the approach with other soils, including with calcareous soils in which the carbonate absorbance band overlaps the nitrate absorbance band and interferes with nitrate determination (Linker et al. 2004, 2005, 2006), and extension of the method to monitoring soil NH_4 and/or NO_2 species, this method would provide a very powerful and cheap tool for studying soil nitrogen transformations.

6. Conclusion

The abundant studies conducted in the mid-IR range have demonstrated the potential that this technique holds for rapid and inexpensive soil analysis. A major advantage of spectroscopy techniques in general and mid-IR in particular is that several properties can be determined from a single spectrum, which greatly reduces the costs of analysis compared to conventional laboratory techniques. In addition, the measurement is very rapid so that a large number of samples can be easily screened or measurements could be conducted "on-the-go", at least in principle. However, several problems still need to be solved before this technology could be upgraded from the research laboratory into routine analyses. The most pressing issue appears to be the standardization of the FTIR and conventional procedures used to analyze the samples, which would make it possible to consider the establishment of spectral libraries similar to those existing in other FTIR fields. Such libraries would provide the basis necessary for developing robust chemometrics models, based either solely on mid-IR spectra or combining mid-IR spectra and pedotransfer functions, valid not only at the local but also at the regional scale.

7. References

Adamchuk, V. I.; Hummel, J. W.; Morgan, M. T. & Upadhyaya, S. K. (2004). On-the-go soil sensors for precision agriculture. *Computers and Electronics in Agriculture* 441: 71-91.

Adamchuk, V. I.; Lund, E. D.; Sethuramasamyraja, B.; Morgan, M. T.; Dobermann, A. & Marx, D. B. (2005). Direct measurement of soil chemical properties on-the-go using ion-selective electrodes. *Computers and Electronics in Agriculture* 483: 272-294.

Alsberg, B. K.; Woodward, A. M.; Winson, M. K.; Rowland, J. J. & Kell, D. B. (1998). Variable selection in wavelet regression models. *Analytica Chimica Acta* 368: 29-44.

Baggs, E. M. (2008). A review of stable isotope techniques for N2O source partitioning in soils: recent progress, remaining challenges and future considerations. *Rapid Communications in Mass Spectrometry* 2211: 1664-1672.

Bertrand, I.; Janik, L. J.; Holloway, R. E.; Armstrong, R. D. & McLaughlin, M. J. (2002). The rapid assessment of concentrations and solid phase associations of macro- and micronutrients in alkaline soils by mid-infrared diffuse reflectance spectroscopy. *Australian Journal of Soil Research* 408: 1339-1356.

Borenstein, A.; Linker, R.; Shmulevich, I. & Shaviv, A. (2006). Determination of soil nitrate and water content using attenuated total reflectance spectroscopy. *Applied Spectroscopy* 60: 1267-1272.

Brereton, R. G. (2003). *Chemometrics. Data analysis for the laboratory and chemical plant*, Wiley & Sons. Chichester.

Canasveras, J. C.; Barron, V.; del Campillo, M. C.; Torrent, J. & Gomez, J. A. (2010). Estimation of aggregate stability indices in Mediterranean soils by diffuse reflectance spectroscopy. *Geoderma* 158: 78-84.

Cécillon, L.; Barthès, B. G.; Gomez, C.; Ertlen, D.; Genot, V.; Hedde, M.; Stevens, A. & Brun, J. J. (2009). Assessment and monitoring of soil quality using near-infrared reflectance spectroscopy (NIRS). *European Journal of Soil Science* 605: 770-784.

Chau, F. T.; Liang, Y.; Gao, J. B. & Shao, X. (2004). Chemometrics. From basics to wavelet transform, In: *Chemical analysis: A series of monographs on analytical chemistry and its applications*, J. D. Winefordner (Ed.).

Christy, C. D. (2008). Real-time measurement of soil attributes using on-the-go near infrared reflectance spectroscopy. *Computers and Electronics in Agriculture* 611: 10-19.

Depczynski, U.; Jetter, K.; Molt, K. & Niemoller, A. (1999). Quantitative analysis of near infrared spectra by wavelet coefficient regression using a genetic algorithm. *Chemometrics and Intelligent Laboratory Systems* 472: 179-187.

Desbiez, A.; Matthews, R.; Tripathi, B. & Ellis-Jones, J. (2004). Perceptions and assessment of soil fertility by farmers in the mid-hills of Nepal. *Agriculture Ecosystems & Environment* 1031: 191-206.

Du, C.; Linker, R. & Shaviv, A. (2008a). Identification of agricultural Mediterranean soils using mid-infrared photoacoustic spectroscopy. *Geoderma* 143: 85-90.

Du, C. W.; Linker, R. & Shaviv, A. (2007). Characterization of soils using photoacoustic mid-infrared spectroscopy. *Applied Spectroscopy* 61: 1063-1067.

Du, C. W.; Linker, R.; Shaviv, A. & Zhou, J. M. (2009a). In Situ Evaluation of Net Nitrification Rate in Terra Rossa Soil Using a Fourier Transform Infrared Attenuated Total Reflection N-15 Tracing Technique. *Applied Spectroscopy* 63: 1168-1173.

Du, C. W. & Zhou, J. M. (2009). Evaluation of soil fertility using infrared spectroscopy: a review. *Environmental Chemistry Letters* 72: 97-113.

Du, C. W.; Zhou, J. M.; Wang, H. Y.; Chen, X. Q.; Zhu, A. N. & Zhang, J. B. (2009b). Determination of soil properties using Fourier transform mid-infrared photoacoustic spectroscopy. *Vibrational Spectroscopy* 491: 32-37.

Du, C. W.; Zhou, J. M.; Wang, H. Y.; Mang, J. B. & Zhu, A. N. (2008b). Study on the soil mid-infrared photoacoustic spectroscopy. *Spectroscopy and Spectral Analysis* 286: 1242-1245.

Ehrentreich, F. (2002). Wavelet transform applications in analytical chemistry. *Analytical and Bioanalytical Chemistry* 3721: 115-121.

Ehsani, M. R.; Upadhyaya, S. K.; Fawcett, W. R.; Protsailo, L. V. & Slaughter, D. (2001). Feasibility of detecting soil nitrate content using a mid-infrared technique. *Transactions of the ASAE* 446: 1931-1940.

Etzion, Y.; Linker, R.; Cogan, U. & Shmulevich, I. (2004). Determination of protein concentration in raw milk by mid-infrared Fourier transform infrared/attenuated total reflectance spectroscopy. *Journal of Dairy Science* 879: 2779-2788.

Gobeille, A.; Yavitt, J.; Stalcup, P. & Valenzuela, A. (2006). Effects of soil management practices on soil fertility measurements on Agave tequilana plantations in Western Central Mexico. *Soil & Tillage Research* 871: 80-88.

Griffiths, P. R. & de Haseth, J. A. (1986). *Fourier transform infrared spectrometry*, Wiley-Interscience. New York.

Idowu, O. J.; van Es, H. M.; Abawi, G. S.; Wolfe, D. W.; Ball, J. I.; Gugino, B. K.; Moebius, B. N.; Schindelbeck, R. R. & Bilgili, A. V. (2008). Farmer-oriented assessment of soil quality using field, laboratory, and VNIR spectroscopy methods. *Plant and Soil* 3071-2: 243-253.

Jahn, B. R.; Linker, R.; Upadhyaya, S. K.; Shaviv, A.; Slaughter, D. C. & Shmulevich, I. (2006). Mid-infrared spectroscopic determination of soil nitrate content. *Biosystems Engineering* 944: 505-515.

Janik, L. J.; Forrester, S. T. & Rawson, A. (2009). The prediction of soil chemical and physical properties from mid-infrared spectroscopy and combined partial least-squares regression and neural networks (PLS-NN) analysis. *Chemometrics and Intelligent Laboratory Systems* 972: 179-188.

Janik, L. J.; Merry, R. H.; Forrester, S. T.; Lanyon, D. M. & Rawson, A. (2007). Rapid prediction of soil water retention using mid infrared spectroscopy. *Soil Science Society of America Journal* 712: 507-514.

Janik, L. J.; Merry, R. H. & Skjemstad, J. O. (1998). Can mid infrared diffuse reflectance analysis replace soil extractions? *Australian Journal of Experimental Agriculture* 387: 681-696.

Janik, L. J. & Skjemstad, J. O. (1995). Characterization and analysis of soils using mid infrared partial least-squares. 2. Correlations with some laboratory data. *Australian Journal of Soil Research* 334: 637-650.

Janik, L. J.; Skjemstand, J. O. & Raven, M. D. (1995). Characterization and analysis of oils using mid infrared partial least-squares. 1. Correlations with XRF-determined major-element composition. *Australian Journal of Soil Research* 334: 621-636.

Kai-man Leung, A.; Chau, F. T.; Gao, J. B. & Shih, T. M. (1998). Application of wavelet transform in infrared spectrometry: spectral compression and library search. *Chemometrics and Intelligent Laboratory Systems* 431-2: 69-88.

Leger, M. N. & Wentzell, P. D. (2004). Maximum likelihood principal components regression on wavelet-compressed data. *Applied Spectroscopy* 587: 855-862.

Linker, R. (2004). Waveband selection for determination of nitrate in soil using mid-infrared attenuated total reflectance spectroscopy. *Applied Spectroscopy* 58: 1277-1281.

Linker, R.; Kenny, A.; Shaviv, A.; Singher, L. & Shmulevich, I. (2004). Fourier Transform Infrared-attenuated total reflection nitrate determination of soil pastes using principal component regression, partial least squares, and cross-correlation. *Applied Spectroscopy* 58: 516-520.

Linker, R.; Shmulevich, I.; Kenny, A. & Shaviv, A. (2005). Soil identification and chemometrics for direct determination of nitrate in soils using FTIR-ATR mid-infrared spectroscopy. *Chemosphere* 61: 652-658.

Linker, R.; Weiner, M.; Shmulevich, I. & Shaviv, A. (2006). Nitrate determination in soil pastes using attenuated total reflectance mid-infrared spectroscopy: Improved accuracy via soil identification. *Biosystems Engineering* 94: 111-118.

Masserschmidt, I.; Cuelbas, C. J.; Poppi, R. J.; De Andrade, J. C.; De Abreu, C. A. & Davanzo, C. U. (1999). Determination of organic matter in soils by FTIR/diffuse reflectance and multivariate calibration. *Journal of Chemometrics* 13: 265-273.

McBratney, A. B.; Minasny, B. & Rossel, R. V. (2006). Spectral soil analysis and inference systems: A powerful combination for solving the soil data crisis. *Geoderma* 136: 272-278.

McCarty, G. W. & Reeves, J. B. (2006). Comparison of NFAR infrared and mid infrared diffuse reflectance spectroscopy for field-scale measurement of soil fertility parameters. *Soil Science* 1712: 94-102.

McCarty, G. W.; Reeves, J. B.; Reeves, V. B.; Follett, R. F. & Kimble, J. M. (2002). Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Science Society of America Journal* 662: 640-646.

McClelland, J. F.; Jones, R. W. & Bajic, S. J. (2002). *Photoacoustic spectroscopy*, In: *Handbook of Vibrational Spectroscopy*, J. M. Chalmers and P. R. Griffiths (Ed.), Wiley & SonsNew York

Minasny, B. & McBratney, A. B. (2008). Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. *Chemometrics and Intelligent Laboratory Systems* 94: 72-79.

Minasny, B.; Tranter, G.; McBratney, A. B.; Brough, D. M. & Murphy, B. W. (2009). Regional transferability of mid-infrared diffuse reflectance spectroscopic prediction for soil chemical properties. *Geoderma* 153: 155-162.

Mouazen, A. M.; De Baerdemaeker, J. & Ramon, H. (2005). Towards development of on-line soil moisture content sensor using a fibre-type NIR spectrophotometer. *Soil & Tillage Research* 80: 171-183.

Mouazen, A. M.; Karoui, R.; De Baerdemaeker, J. & Ramon, H. (2006). Characterization of soil water content using measured visible and near infrared spectra. *Soil Science Society of America Journal* 70: 1295-1302.

Mouazen, A. M.; Maleki, M. R.; De Baerdemaeker, J. & Ramon, H. (2007). On-line measurement of some selected soil properties using a VIS-NIR sensor. *Soil & Tillage Research* 93: 13-27.

Nguyen, T. T.; Janik, L. J. & Raupach, M. (1991). Diffuse reflectance infrared Fourier-transform (DRIFT) spectroscopy in soil studies. *Australian Journal of Soil Research* 29: 49-67.

Reeves III, J. B. (2010). Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? *Geoderma* 158: 3-14.

Reeves, J.; McCarty, G. & Mimmo, T. (2002). The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. *Environmental Pollution* 116: S277-S284.

Reeves, J. B. (2009). Does the Spectral Format Matter in Diffuse Reflection Spectroscopy? *Applied Spectroscopy* 63: 669-677.

Reeves, J. B.; Francis, B. A. & Hamilton, S. K. (2005). Specular reflection and diffuse reflectance spectroscopy of soils. *Applied Spectroscopy* 59: 39-46.

Reeves, J. B.; McCarty, G. W. & Reeves, V. B. (2001). Mid-infrared diffuse reflectance spectroscopy for the quantitative analysis of agricultural soils. *Journal of Agricultural and Food Chemistry* 49: 766-772.

Reeves, J. B. & Smith, D. B. (2009). The potential of mid- and near-infrared diffuse reflectance spectroscopy for determining major- and trace-element concentrations in soils from a geochemical survey of North America. *Applied Geochemistry* 24: 1472-1481.

Sá, J. C. d. M.; Cerri, C. C.; Lal, R.; Dick, W. A.; de Cassia Piccolo, M. & Feigl, B. E. (2009). Soil organic carbon and fertility interactions affected by a tillage chronosequence in a Brazilian Oxisol. *Soil and Tillage Research* 1041: 56-64.

Shaviv, A.; Kenny, A.; Shmulevitch, I.; Singher, L.; Raichlin, Y. & Katzir, A. (2003). Direct monitoring of soil and water nitrate by FTIR based FEWS or membrane systems. *Environmental Science & Technology* 3712: 2807-2812.

Sibley, K. J.; Astatkie, T.; Brewster, G.; Struik, P. C.; Adsett, J. F. & Pruski, K. (2009). Field-scale validation of an automated soil nitrate extraction and measurement system. *Precision Agriculture* 102: 162-174.

Sinfield, J. V.; Fagerman, D. & Colic, O. (2010). Evaluation of sensing technologies for on-the-go detection of macro-nutrients in cultivated soils. *Computers and Electronics in Agriculture* 701: 1-18.

Stevens, R. J. & Laughlin, R. J. (1994). Determining N-15 in nitrite or nitrate by producing nitrous-oxide. *Soil Science Society of America Journal* 584: 1108-1116.

Stevens, R. J.; Laughlin, R. J.; Atkins, G. J. & Prosser, S. J. (1993). Automated-determination of nitrogen-15-labeled dinitrogen and nitrous-oxide by mass-spectroscopy. *Soil Science Society of America Journal* 574: 981-988.

Tranter, G.; Minasny, B.; McBratney, A. B.; Rossel, R. A. V. & Murphy, B. W. (2008). Comparing spectral soil inference systems and mid-infrared spectroscopic predictions of soil moisture retention. *Soil Science Society of America Journal* 72: 1394-1400.

Trygg, J. & Wold, S. (1998). PLS regression on wavelet compressed NIR spectra. *Chemometrics and Intelligent Laboratory Systems* 42: 209-220.

Verma, S. K. & Deb, M. K. (2007). Nondestructive and rapid determination of nitrate in soil, dry deposits and aerosol samples using KBr-matrix with diffuse reflectance Fourier transform infrared spectroscopy (DRIFTS). *Analytica Chimica Acta* 582: 382-389.

Viscarra Rossel, R. A.; Jeon, Y. S.; Odeh, I. O. A. & McBratney, A. B. (2008). Using a legacy soil sample to develop a mid-IR spectral library. *Australian Journal of Soil Research* 46: 1-16.

Viscarra Rossel, R. A. & Lark, R. M. (2009). Improved analysis and modelling of soil diffuse reflectance spectra using wavelets. *European Journal of Soil Science* 60: 453-464.

Viscarra Rossel, R. A. & McBratney, A. B. (1998a). Laboratory evaluation of a proximal sensing technique for simultaneous measurement of soil clay and water content. *Geoderma* 851 19-39.

Viscarra Rossel, R. A. & McBratney, A. B. (1998b). Soil chemical analytical accuracy and costs: implications from precision agriculture. *Australian Journal of Experimental Agriculture* 38: 765-775.

Viscarra Rossel, R. A.; Walvoort, D. J. J.; McBratney, A. B.; Janik, L. J. & Skjemstad, J. O. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131: 59-75.

Wold, S.; Sjostrom, M. & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58: 109-130.

Zimmermann, M.; Leifeld, J. & Fuhrer, J. (2007). Quantifying soil organic carbon fractions by infrared-spectroscopy. *Soil Biology & Biochemistry* 39: 224-231.

19

The Application of FT-IR Spectroscopy in Waste Management

Ena Smidt, Katharina Böhm and Manfred Schwanninger
BOKU - University of Natural Resources and Life Sciences, Vienna
Austria

1. Introduction

This synopsis of FT-IR spectroscopic applications in waste management covers relevant issues regarding monitoring, process and quality control. Quality in this context means low reactivity, low gas forming potential and compliance with limit values of materials to be landfilled, appropriate compost ingredients and improvement of the stable carbon pool by humification of compost organic matter. The division into sections was carried out according to specific materials and processes and the related questions to be answered.

Waste materials have not posed a problem in the past, as long as they were returned to the natural cycle. The higher degree of utilisation, the long lasting life cycle of goods and their re-use caused minor waste amounts and contributed to the balance of input and output streams. As natural materials were the basis for good production, waste was integrated in the natural cycle. Organic waste materials only escaped degradation if they were preserved under particular conditions. Inorganic residues from ore mining have led to local contamination by heavy metals. In contrast, fragments of ancient pottery or ruins have gained in importance as historical witnesses. The economical and social increase in prosperity and urban development has been paralleled by strongly rising amounts of organic and inorganic waste and the acceleration of turnover rates. These have led primarily to sanitation problems, especially in expanding urban areas. Disposal of these wastes by spreading them over the surrounding countryside or filling dumps were only temporary "solutions". Natural waste such as foliage falling in autumn can be considered as an intermediate product in a closed circle. Except for natural disasters, turnover rates and the related contents of substances are subjected to regulations. Anthropogenic waste is in most cases the end of a one-way street that causes system imbalances. Therefore new approaches and waste management strategies aim at copying natural cycles.

The amount, changes in the chemical composition of our goods, the use of hazardous substances and the careless landfilling of waste in dumps have caused serious environmental problems and demonstrated the need for action. The awareness of this issue emerged when soils and groundwater were contaminated by leachate emissions. Due to the missing separation of hazardous waste, landfill remediation has often implied a complete excavation of the landfilled material. Discussions on climate change, the role of waste management in the global carbon cycle and its contribution to the carbon budget have drawn attention to relevant gaseous emissions from landfills with a global warming potential, especially of methane.

Analytical approaches in the past differed in that the main analytical interest focused on the effect of gaseous emissions on the environment and on soil and water pollution. Environmental analytical methods were applied to find out the effects in these spheres caused by waste materials. Waste materials themselves were not investigated. The first attempts to encapsulate waste materials were aimed at prevention of contact with the environment. Analytical investigations in the past concentrated on pollution control and on the chemical and physical quality of landfill liners with regard to their density, tightness and stability.
Later on the interest in the waste material itself came to the fore as it represents the source of future emissions. The environmental impact and the economic implication caused the government to react in terms of regulations regarding the handling of waste materials. During the last decades waste management has become an important industrial sector in countries with a high environmental awareness and adequate standards. The approach of waste encapsulation was abandoned due to the missing knowledge about the durability of landfill liners and technical facilities. Moreover, it was evident that the encapsulation would have promoted the preservation of the material and postponed the problem of reactivity.
The concept of a "multi-barrier system" that includes pre-treatment of municipal solid waste in terms of stabilisation besides the technical equipment and an adequate geological basement was implemented. The reduction of waste reactivity within a manageable space of time should avoid the long lasting after care period of several decades that does not comply with the objective of the Austrian Waste Act (BMLFUW, 1990) not to burden the next generations with environmental problems from the past.
Among the wide range of current research topics in waste management the development of appropriate analytical tools is an overall concern at the national and the European level. Waste materials represent an analytical challenge due to their widely varying composition and structure contrary to all principles of natural order. The implementation of limit values by national rules and the need for environmental compliance have stimulated the discussion on adequate methods, their reliability and usefulness. Their availability is a prerequisite to achieving the protection of the environment according to the principles of social welfare and sustainability as required by the aims of the Austrian Waste Act (BMLFUW, 1990). For waste management practice analytical methods should be fast, cheap, easy to handle, and marginally error-prone. Many requirements in view of the complex matrix!
There are two main problem areas in waste management to be dealt with: toxic effects and reactivity of the material. Heavy metals, organic pollutants such as hydrocarbons, polycyclic hydrocarbons and chlorinated substances are quantified individually. The complex matrix of waste materials makes the determination and quantification of single substances difficult. Interferences of different compounds affect their extraction and separation for quantification. The complex mixture of different materials regarding chemical components, texture and behaviour implies the application of new analytical tools. With respect to pollutants the determination of single compounds remains indispensable in many cases. Analytical advances concentrate on improved extraction and disintegration methods and on modern instruments for determination. For an overall waste characterisation including chemical properties and behaviour more holistic approaches have gained in importance.

1.1 Fourier transform infrared (FT-IR) spectroscopy and data evaluation

FT-IR spectroscopy has proved to be a powerful tool to comply with the purpose of comprehensive characterisation. The unique characteristic of the material presented by the

spectrum sheds light on material properties, its behaviour as well as on specific components represented by their functional groups. The substantial progress regarding the infrared spectroscopic measurement is achieved by the recording of the interferogram, the fast detection systems and the Fourier transformation. This combination provides many advantages and has caused a high interest in this method for process and quality control in many areas (Mantsch & Chapman, 1996; Moron & Cozzolino, 2004; Pollanen et al., 2005; Zhang et al., 2005). Infrared spectroscopy is based on interactions of infrared radiation with matter. Infrared light causes functional groups to vibrate. The uptake of energy is indicated by absorption bands in the spectrum. Measured band intensities depend on the content of the substance to be determined and on the individual interaction of the functional group with infrared radiation at a distinct energy level (Hesse et al., 1995; Smith, 1999; Socrates, 2001). The near-infrared (NIR) region from 14,000 cm^{-1} – 4000 cm^{-1} and the mid-infrared (MIR) region from 4000 cm^{-1} – 400 cm^{-1} are commonly applied for process and product control in many industrial fields. The spectral pattern of substances reveals inherent features and allows the proof of their identity. These characteristics support the solution of frequent problems in industry. Infrared spectroscopic investigations of waste materials presented in this chapter have focused on the KBr (potassium bromide) pellet and the attenuated total reflection (ATR) technique in the MIR area. Most experiences and insight regarding spectra interpretation, assignment and behaviour of bands during waste degradation or stabilisation have been gained from the KBr technique (Smidt et al., 2002; Smidt & Schwanninger, 2005; Smidt & Meissl, 2007). Subsequently, investigations were extended to the NIR area and the reflection mode using a fibre probe and the integrating sphere, as well as to the ATR technique in the MIR area (Meissl et al., 2008a). During the last decade the applicability for waste characterisation and assessment has been proved and gives impetus to new ways in waste management practice (Chen, 2003; Michel et al., 2006; van Praagh et al., 2009). In association with multivariate statistical methods these techniques provide an additional benefit for application in practice in terms of time saving and handling. Multivariate statistical methods extract a maximum of latent information from a huge data pool (Brereton, 2002; Esbensen, 2002) and transform the complex spectral pattern into a generally understandable result. The presented results for classification and prediction of the applied multivariate statistical methods focus on Principal Component Analysis (PCA), Soft Independent Modelling of Class Analogy (SIMCA), Partial Least Squares - Discriminant Analysis (PLS-DA), and Partial Least Squares Regression (PLS-R). PCA is used to analyse large data sets. This procedure extracts the information of the original data matrix which leads to a smaller number of dimensions, called principal components (PCs). The required number of PCs depends on the complexity or dissimilarity of the data-set. The variance explained by each PC decreases with increasing number of PCs. PCA reveals the inherent data structure and underlying features which supports the identification of similarities and differences between materials. The loadings spectra, corresponding to the PCs, provide information on the contribution of the spectral regions to the differentiation between samples visible in the scores plot. Based on the PCA, classification and prediction models can be developed. SIMCA is a well known pattern recognition method which describes each class separately in a principal components space. New objects are considered to belong to the class if their distance (e.g. Euclidean) to the constructed PC space is not significantly larger than the distance of the class objects to their PC space. In a further classification method PLS-DA, samples are distinguished by means of the partial least squares regression according to their membership that is defined by the dummy variable (+1) and (-1).

PLS-R relates the variation in one variable (questioned parameter = Y-variable) to the variation of spectral data (X-variables). Based on this correlation models were calculated and further used for prediction. It is a prerequisite that the questioned parameter is reflected by the spectral pattern (Michel et al., 2006; Meissl et al., 2007). The quality of classification and prediction models is expressed by several characteristics (model parameters). R^2 is the coefficient of determination. The root mean square error of cross validation (RMSECV) reflects the error of the model. Cross validation is a validation method where samples out of the sample set are alternately excluded from calibration and used for prediction. This procedure is repeated until all samples have been kept out once. The bias indicates the systematic difference between the measured and the predicted values. The ratio of standard deviation to standard error of performance (RPD) provides information on the precision of analyses for a specific purpose (Williams & Norris, 2004).
This study shows in many cases the development or changes of spectral characteristics, therefore the spectra were shifted in parallel for a clear presentation. Most spectra were recorded from KBr pellets except the ATR spectra shown in section 2.3. For multivariate data analysis spectra were vector normalised.

1.2 Interpretation of infrared spectral data originating from waste materials

Waste is a complex material and heterogeneous regarding the chemical composition and texture. Due to the low amount of sample needed for MIR spectroscopic measurement adequate sample preparation is a prerequisite to obtain reliable and reproducible results. A convenient procedure of sample preparation was described by Meissl et al. (2008b).
Pure substances are characterised by sharp distinct absorption bands that can be assigned to functional groups. By contrast, waste materials display broad overlapping bands due to the complex mixture and manifold interactions between degrading organic molecules. During the last decade MIR and NIR spectroscopy has been applied with increasing success to complex samples. New approaches have gained in importance due to the association with multivariate statistical methods. The proof of identity and the classification of unknown waste materials are based on similarities or differences of the spectral pattern. Identification of functional groups and assignment to substances are an essential target to follow and to understand the chemical changes during degradation and stabilisation processes. Due to the complexity of the material and overlapping bands the identification is limited to a few but significant indicator bands.
Spectra interpretation of waste materials is primarily based on theoretical locations of functional groups cited in literature for pure substances as well as for waste materials or components of waste materials (Hesse et al., 1995; Smith, 1999; Socrates, 2001; Smidt et al., 2002). Due to interactions of degrading organic molecules band shifts in waste spectra occur frequently. Several bands that are assigned to organic functional groups, such as aliphatic methylene bands (Table 1) have a stable band position in waste spectra. The location of many other bands may vary due to the stronger influence of the whole molecule and the waste matrix. Bands of inorganic compounds are mostly sharp and support identification by their characteristic shape and relatively stable position. For many functional groups wavenumber regions are indicated. Band assignment is supported by recording the spectral signature of waste ingredients or by addition of a specific component to waste materials of which the presence should be verified. The most relevant indicator bands that provide information on the stage of degradation are compiled in table 1. The list is limited to bands that are clearly visible in the spectrum of the complex waste matrix even though many substances are

represented by several bands. Intensities of bands that are assigned to organic functional groups decrease with progressing degradation until a nearly constant level is reached that indicates the slow-down of metabolic activities. Bands of some metabolites appear temporarily and disappear. Degradation of organic matter leads to a relative increase of mineral compounds that is reflected by the increase of the corresponding absorption bands (Hesse et al., 1995; Naumann et al., 1996; Haberhauer et al., 1998; Grube et al., 1999; Smith, 1999; Ouatmane et al., 2000; Socrates, 2001; Reig et al., 2002; Smidt et al., 2002; Zaccheo et al., 2002; Chen, 2003; Madejova, 2003; Tan, 2003; Smidt & Schwanninger, 2005; Smidt & Meissl, 2007).

Location wavenumber (cm^{-1})	Vibration	Functional group or component
Organic compounds		
2920	C-H stretching (as)	aliphatic methylene group
2850	C-H stretching (s)	aliphatic methylene group
2590-2520	S-H stretching	thiols
1740-1700	C=O stretching	aldehyde, ketone, carboxylic acids, esters
1685-1630	C=O, COO^- stretching C=C stretching	amide I, carboxylates aromatic ring modes, alkenes
1600-1590	C=C	aromatic skeleton
1570-1540	N-H in plane bending	amide II and secondary amines
1515-1505	aromatic skeletal	lignin from lignocellulosic materials
1430-1420	COO^- stretching	carboxylic acids
1350-1250	C-N stretching	primary and secondary aromatic amines
1265-1240	C-O-C stretching C-N stretching	esters amide III
1250-900	C-O-C, C-O C-O-P	polysaccharides phosphodiesters
Inorganic compounds		
3700-3200	SiO-H stretching	silica
3700-3400	O-H stretching	bonded and non-bonded hydroxyl groups and water
1635	O-H bending	water
1450-1410	C-O stretching	carbonate
1384 (1400-1340)	N-O stretching	nitrate (leachate)
1140-1080	S-O stretching	sulphate
1080	Si-O stretching	quartz
1030	Si-O stretching Si-O-Si stretching	clay minerals silica
875	C-O out of plane bending	carbonate
713	C-O in plane bending	carbonate
680-610	S-O bending	sulphate

Table 1. Location of indicator bands, corresponding vibration and assignment to functional groups and components (as = antisymmetric, s = symmetric)

The nitrate band at 1384 cm^{-1} in solid waste materials has been proved by addition of KNO_3 (potassium nitrate) (Fig. 1a). The appearance of the nitrate band in the compost spectrum at

an advanced stage of composting corresponds to the evolution of nitrate in the biological process (Fig. 1b). The sharp band shape is an additional indicator of an inorganic compound that does not interact with other molecules as degrading organic molecules do. Both facts support identification of nitrate characterised by a stable band position in solid waste samples (Smidt et al., 2002). It is evident that all background information available contributes to band assignment and spectra interpretation. It should be mentioned that the position of the nitrate band in liquid samples vary in a wider range (cp. freeze-dried leachate in Fig. 5).

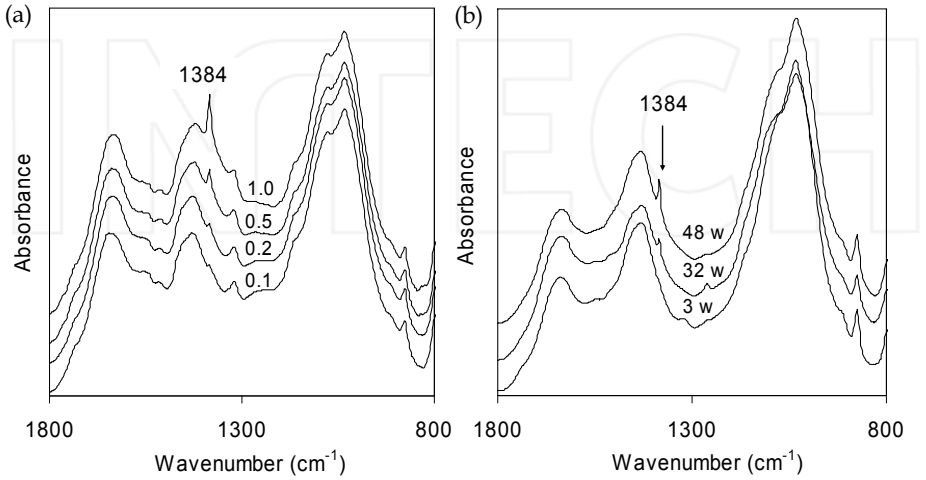


Fig. 1. (a) Increasing nitrate band at 1384 cm^{-1} due to addition of KNO_3, corresponding to mass fractions of 0.1, 0.2, 0.5, and 1% nitrate; (b) emerging nitrate band during a composting process (3, 32, 48 weeks)

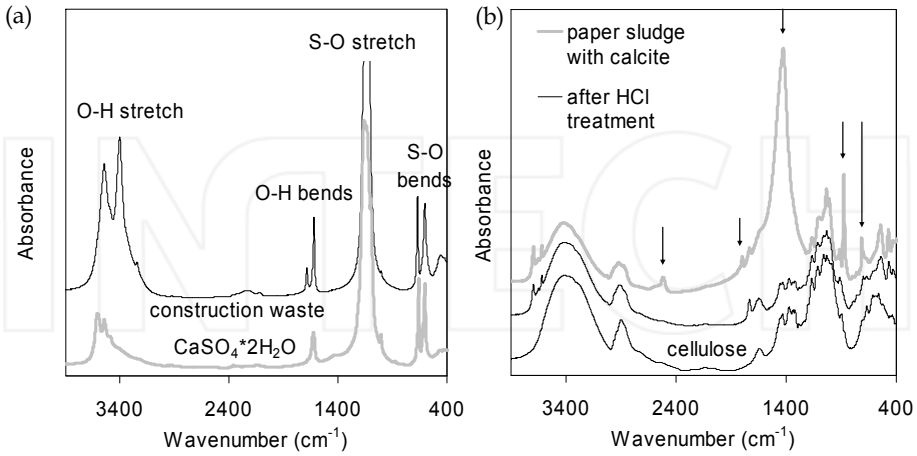


Fig. 2. (a) Infrared spectra of mineral lumps from a construction waste landfill and $CaSO_4*2H_2O$; (b) spectral characteristics of the original industrial landfill sample (carbonate bands indicated by arrows), the HCl treated sample and cellulose as reference

Figure 2 shows approaches of band assignment by comparison of waste ingredients with a pure substance and by sample treatment. Figure 2a displays spectra of mineral lumps from a construction waste landfill and $CaSO_4*2H_2O$ (calcium sulphate) to reveal the specific bands assigned to sulphate that plays a relevant role in leachate. Apart from the sulphate (S-O) stretching and bending vibrations at 1136 cm^{-1} and 620 cm^{-1}, respectively, the mineral lumps also show strong O-H stretching vibrations. Figure 2b illustrates an example of waste material originating from an abandoned landfill with unknown composition. The sample was treated with HCl (c = 0.1 mol/L) to convert carbonates into carbon dioxide and water, and to reveal bands of other compounds that are overlapped and covered by strong bands of carbonate. Cellulose was identified by comparison with a spectrum of pure cellulose. The material was found to be industrial paper sludge mixed with carbonate for stabilisation before landfilling.

2. Abandoned sites and old landfills

The environmental damages resulting from abandoned landfills and dumps have revealed the problems of careless disposal. At the beginning of a well-regulated and organised waste management system pollutants and toxic substances were in the foreground due to their immediate impact on water and soil. They covered the more subtle contribution of unsuspicious goods of daily use to relevant emissions and their effect on the imbalance in the global carbon budget. The registration of abandoned landfills is part of exploration programmes in terms of risk assessment and adequate remediation measures. The majority of abandoned landfills under investigation contain municipal solid waste and construction waste originating from the sixties and seventies of the last century. The measurement of landfill gas emissions is still state of the art and part of monitoring programmes. Long-path spectroscopic instruments developed for air pollution control are also applied for landfill monitoring. Long-path FT-IR spectroscopic investigations provide several advantages such as fast scanning of large surfaces at long distances without contact and online measurement of emissions from areas that are difficult to access. Due to infrared absorption of many air pollutants they can be detected and quantified by FT-IR long-path instruments (Bacsik et al., 2005). Quantification of gaseous compounds is carried out according to Lambert-Beer's law (Weber et al., 1996; Galle et al., 2001; Hegde et al., 2003). Methane and nitrous oxide are the main components to be measured in the context of landfill monitoring. The measurement of gaseous emissions only provides information on the current microbial activity that depends on environmental conditions such as water supply. The gas forming potential according to the present chemical compounds that could be degraded under appropriate conditions cannot be determined in this way. Preferential gas flows make the quantification of emissions difficult. Due to the missing base seal, the leachate that provides additional information is not available in most cases. Therefore investigations of the solid waste are indispensable.

2.1 Classification of waste materials

The classification of landfilled materials should be a first step. The knowledge about the landfilled material supports the interpretation of data as the composition influences the behaviour, especially biological tests of reactivity. Municipal solid waste (MSW) comprises all components of consumer goods and features a typical spectral pattern. Different pre-treatment and landfilling conditions of MSW lead to divergent spectral characteristics of the

waste material. Old landfills (old-LF) are characterised by the high portion of construction waste. The 5-year-old recent reactor landfill (reactor-LF) differs from the landfill with mechanically-biologically treated waste (MBT-LF) due to the missing pre-treatment. The waste bank represents a special kind of deposit where scarcely biologically treated MSW was piled up to form a constructional wall of about 6 m. Despite the differences in composition and degradation conditions similarities are distinctive. Industrial waste (ind-LF), by contrast, features a more one-sided composition. The following PCAs are based on the mean-centred MIR spectra that cause zero-lines in the scores and loadings plots, indicating the averages of all samples. The PCA supports the discovery of differences and similarities that are visualised by long or short distances between the samples in the scores plot (Figs. 3a, c). They illustrate the differences between landfills containing MSW and an industrial landfill containing carbonate stabilised paper sludge. The loadings plots elucidate how much the spectral regions contribute to the first (PC1, explained variance 69%) and the second (PC2, 15%) principal component (Figs. 3b, d). The PCA calculated with the whole spectrum (Figs. 3a, b) reveals the strong influence of inorganic components, especially of carbonate and clay minerals that dominate the loadings spectrum of PC1. The position of the ind-LF samples found at the positive side of PC1 in the scores plot corresponds to the positive loadings (e.g. at 1420 cm^{-1}). The position of most of the MSW landfill samples found at the negative side of PC1 in the scores plot corresponds to the negative loadings (e.g. at 1030 cm^{-1}). To sum it up: positive loadings plus positive scores = ind-LF with more carbonate (1420 cm^{-1}), negative loadings plus negative scores = MSW landfills with more clay minerals (1030 cm^{-1}). This means that ind-LF and MSW landfills differ basically in composition especially in carbonate and clay minerals. Different types of MSW landfills are distinguished along PC2 that reflects the development from reactor-LF to old-LFs. Apart from inorganic compounds (clay minerals) spectral regions that can be assigned to organic functional groups (2920 and 1640 cm^{-1}) become more relevant in the loadings spectrum (Fig. 3b). The characteristics are in accordance with higher organic matter contents and reactivity in the reactor-LF compared to old-LFs that display higher mineral contents due to mineralisation and a portion of construction waste. However, the transition from one landfill type to another one is not distinct as old-LFs can still feature more reactive sections. Vice versa, the reactor-LF also contains sections of low organic matter content and moderate reactivity. Most of the material from the 15-year-old waste bank is located in the scores plot between the reactor-LF and the old-LFs. The variance within the bank samples is caused by different environmental conditions and development during the 15-years of disposal. More details were reported by Smidt et al. (2007).

Figs. 3c and 3d present the scores plots and the loadings spectra of a PCA that was based on the organic indicator bands represented by C-H and C=O vibrations. Separation of MSW landfill types due to organic compounds is performed along PC1. MSW landfill types and the ind-LF are differentiated along PC2. In terms of these characteristics several old-LF samples move closer to ind-LF samples in the scores plot. This indicates a weaker discrimination power of the selected spectral regions.

The PCA is an appropriate tool to obtain a general idea on the similarities and differences of waste materials and the contributing spectral regions. For practical application the development of classification models is indispensable in order to immediately attribute waste materials to defined classes depending on the problem to be solved. A classification model according to SIMCA was reported by Smidt et al. (2008b) to distinguish different waste materials.

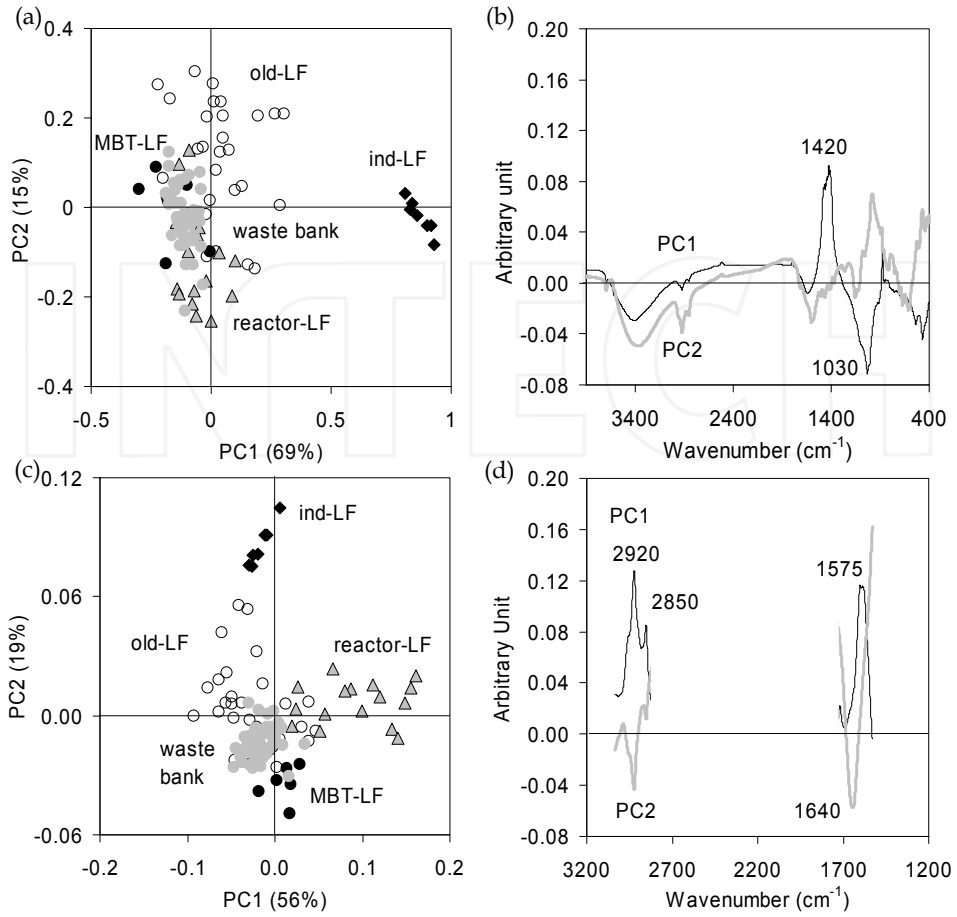


Fig. 3. Scores plots and loading spectra of samples from landfilled wastes based on the whole spectra (a, b) and on selected wavenumber regions from 3030 to 2800 cm^{-1} and from 1795 to 1530 cm^{-1} (c, d)

2.2 Risk assessment of abandoned landfills and dumps regarding the reactivity

The careless landfilling of waste in the past has caused extensive investigations to assess the remaining risk and the necessity of remediation measures. In this context the stability of waste organic matter is a fundamental criterion (Binner & Zach, 1999; Tesar et al., 2007; van Praagh et al., 2009). The proof of stability can be carried out by means of time-consuming biological tests that provide information on the current microbial activity under aerobic conditions and the gas forming potential under anaerobic conditions. These tests were primarily established for MBT waste to verify its stability prior to landfilling (Binner & Zach, 1999; Adani et al., 2004). The dissolved organic carbon (DOC) and the total organic carbon (TOC) are relevant parameters for the assessment of abandoned landfills according to the Austrian Landfill Ordinance (BMLFUW, 2008). The DOC provides information on

soluble organic compounds. Their portion decreases with progressing stabilisation. The stipulated limit values of TOC restrict the organic matter content in all landfill types. Only for MBT waste the TOC limitation is not valid. Figure 4 shows the correlation between predicted and measured (references) parameters using selected regions of the MIR spectrum and PLS-R. The model parameters are summarised in table 2.

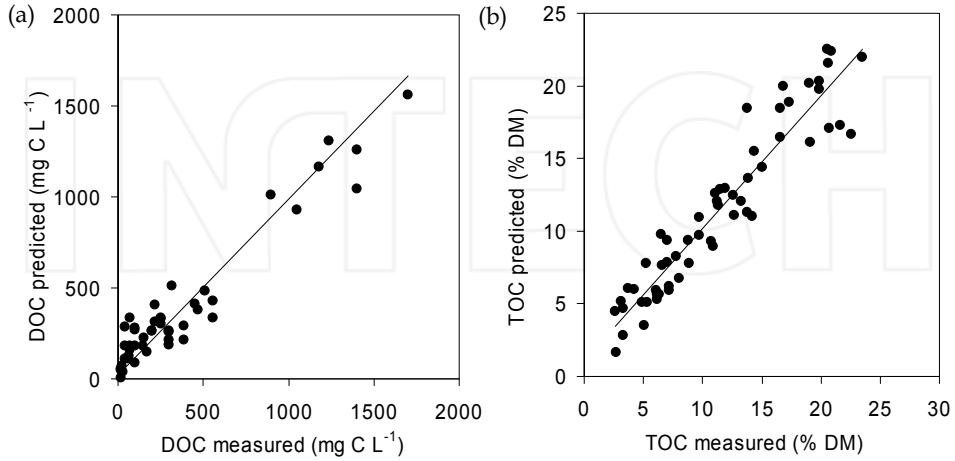


Fig. 4. Correlation between (a) the predicted and the measured DOC and (b) the predicted and the measured TOC, both based on the spectral ranges from 3003 - 2816 and 1770 - 400 cm^{-1} and PLS-R

Model parameters	DOC	TOC
Wavenumber ranges (cm^{-1})	3003 - 2816 and 1770 - 400 cm^{-1}	
Calibrated range	15 – 1700 mg C L^{-1}	2.6 – 23.5 % DM
R^2	89%	90%
RMSECV	158 mg C L^{-1}	1.9 % DM
RPD	3.0	3.1
No. of PLS components	6	7
Bias	-5.94	-0.0719

Table 2. Parameters for the PLS-R models dissolved organic carbon (DOC) and total organic carbon (TOC) content

2.3 Success control of remediation measures

In situ aeration of old reactive landfills is one measure to avoid methane emissions and to accelerate mineralisation in that anaerobic conditions are exchanged for aerobic ones. The spectral pattern of the solid waste matrix reflects the chemical changes mainly by decreasing band heights of organic functional groups (Tesar et al., 2007). Aeration of the solid waste is paralleled by oxidation of soluble components in the leachate. The most conspicuous changes in the spectral pattern are caused by transformation of N-H and S-H groups containing compounds to the mineralisation products nitrate and sulphate that are

presented in the spectrum by strong bands of the N-O and the S-O vibrations (Smidt & Schwanninger, 2005; Smidt & Meissl, 2007). ATR spectra (4000 – 600 cm^{-1}) of freeze-dried leachate from aerated landfill material and leachate from landfill material under anaerobic conditions were classified by means of PLS-DA (Fig. 5). The corresponding loadings spectrum of the 1st PLS component (Fig. 5b) shows the relevant regions, especially the S-O, N-O, C-H, S-H and N-H vibrations (Table 1) that contribute to the discrimination of leachate obtained from materials under different aeration conditions. Three PLS components explain 97% of the total variance, whereas the 1st one already explains 94%, indicating a distinct differentiation and the efficiency of aeration.

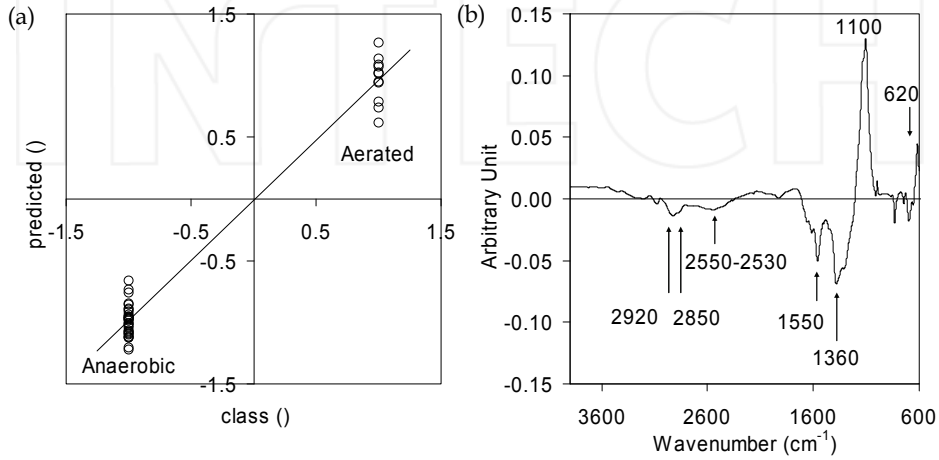


Fig. 5. PLS-DA results based on spectral data of freeze-dried leachate from landfill material under anaerobic and aerated conditions: (a) correlation between predicted and class (dummy variable) and (b) loadings spectrum of the 1st PLS component

2.4 Long-term behaviour of deposits

The long-term behaviour of landfills and dumps is a relevant issue with respect to remediation activities and the re-use of these areas. Although organic matter degradation leads to similar metabolic and mineralised products, the individual composition causes a specific spectral pattern as shown in figure 3. MSW that had been piled up to a waste bank (Fig. 3) showed different characteristics after 15 years corresponding to depths and air supply. Aerobic conditions in the upper layer of the profiles accelerated degradation and corresponding changes in the spectral pattern (Smidt et al., 2007). The classification of samples according to specific stages of degradation by means of spectral characteristics is a fast method for the risk assessment of abandoned landfills, besides the precise determination of relevant parameters as shown in section 2.2. The long-term behaviour is strongly influenced by both the chemical composition and environmental conditions.

The scores plot and loadings spectra in figure 6 illustrate the distance of samples and the responsible spectral regions. The samples originated from an old landfill. They were taken as a composite sample along the depth profile. Two samples (black dots) that were collected between tight clayey layers, where degradation was inhibited, differ considerably from the other ones. The loadings spectra of PC1 and PC2 reveal the strong influence of mineral

compounds (1430 – 400 cm^{-1}). The loadings spectrum of PC1 additionally shows the contribution of organic components (1630 cm^{-1}). Besides the poor mineralisation the enclosed material was protected against degradation and subsequent mixing with construction waste.

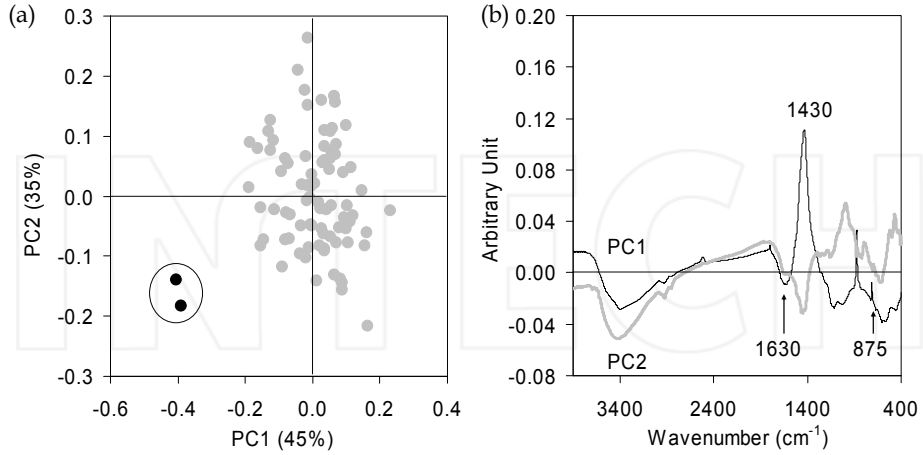


Fig. 6. PCA based on spectra of samples from an old landfill; (a) scores plot with 2 samples (black dots) between tight clayey layers, (b) loadings spectra of PC1 and PC2

3. Separation of recyclables and mechanical-biological or thermal pre-treatment of municipal solid waste prior to landfilling

The maximum possible separation of recyclables is a crucial issue. In addition to the individual separation by the waste producer, the recovery of recyclables is improved by mechanical sorting systems in the plant. The pre-treatment of municipal solid waste prior to landfilling comprises two main strategies: the mechanical-biological and the thermal treatment. Both processes aim at reducing reactivity, either by mineralisation and stabilisation or by incineration of organic matter.

3.1 Identification and separation of recyclables by NIR spectroscopy

The development of mechanical separation and sorting systems has become an important technical approach in the improvement of the recycling proportion and the required quote respectively. Identification of different materials is a prerequisite in order to achieve an efficient and correct separation. NIR spectroscopy is an appropriate tool for a comprehensive characterisation of wastes (Bonifazi et al., 2009) and differentiation of polymers. In this field of application the measurement in the near infrared area ranks first (Feldhoff et al., 1997; Van Den Broek et al., 1998; Kulcke et al., 2003; Leitner et al., 2003; Li et al., 2005; Dou et al., 2006; Luiken & Bos, 2010).

3.2 Mechanically-biologically treated (MBT) waste - process control and compliance with limit values according to the Austrian Landfill Ordinance

National and international regulations and limit values for MBT waste require verification of compliance before final disposal. For MBT waste reactivity and stability are relevant

properties. Process kinetics, the release of organic components and their long-term behaviour are fields of intensive research in waste management (Barlaz et al., 1990; Barlaz, 1998; Tintner et al., 2010). Despite the scientific interest, the performance of industrial processes strongly depends on this basic knowledge. Waste organic matter primarily consists of non-hazardous organic molecules responsible for the reactivity of the material and emerging gaseous and liquid emissions. Chemical and physical sum parameters are preferably applied for waste characterisation in practice. Waste organic matter is usually quantified by the determination of loss of ignition, TOC and nitrogen content. The significance of these parameters is limited as they do not differentiate the particular behaviour of the molecules subsumed therein. Mineralisation products of organic matter such as ammonium and nitrate provide information on progressing biological degradation. High ammonium contents are assigned to early stages of degradation, whereas nitrate indicates the late phase when ammonium has been oxidised. Nevertheless, a wide range of contents is observed during the biological treatment. The coexistence of ammonium and nitrate and the repeated increase and decrease of ammonium contents make unambiguous assessment difficult. Last but not least these parameters depend on the uniformity of metabolic processes in the rotting unit. The equal progress is strongly related to process operation in terms of mechanical mixing and evenly distributed aeration and moisture.

Biological tests provide more comprehensive information on reactivity in that chemical properties of all compounds contribute to the behaviour of waste materials under aerobic and anaerobic conditions. Biodegradability of organic compounds is revealed by the oxygen uptake or CO_2 release due to microbial activity, indicating mineralisation of organic molecules. The gas sum over a period of 21 days, generated under anaerobic conditions, is measured by means of the incubation test (Austrian Standards Institute, 2004). For MBT materials to be landfilled biological tests replace the determination of the total organic carbon content (TOC) because the limit value of 5% dry matter for this landfill type cannot be achieved by biological degradation only. However, biological parameters are time-consuming and require considerable expertise. In addition to biological tests limitation of the calorific value by 6600 kJ kg^{-1} dry matter was stipulated for MBT waste. The calorific value indicates the efficiency of plastic separation and progressing degradation of organic matter by its decline (BMLFUW, 2008). Chemical changes during the biological treatment of MSW are reflected by spectral characteristics that allow a fast process control, evaluation and optimisation. The progress of organic matter degradation is influenced by the technical system and process operation. The development of spectral characteristics is illustrated in figure 7. Fresh input materials feature strong aliphatic methylene bands at 2920 and 2850 cm^{-1}. Absorption bands of organic components (cp. Table 1, cellulose) decrease or disappear as far as they indicate temporarily present metabolic products (1740, 1320 cm^{-1}). This process showed extensive mineralisation after 9 weeks. It can be assumed that the chemical composition of MSW is related to the biological behaviour. PLS-R models based on MIR spectra were developed for the prediction of the biological parameters "respiration activity" and "gas generation sum" to get information on reactivity of MBT waste faster (Böhm et al., 2010a). Bands that feature considerable changes during the biological treatment such as the aliphatic methylene bands and the fingerprint region (1788 - 1532 cm^{-1}, 1350 - 1028 cm^{-1}), were selected for the prediction models.

The loadings spectrum of the first PLS component out of 8 (Fig. 8b) indicates that organic (methylene bands) and inorganic (carbonate bands at 1430 and 875 cm^{-1}) compounds contribute to the thermal behaviour of the waste sample. Whereas waste organic matter

causes the calorific value to increase, water and carbonates affect its decrease due to the endothermic reactions of water evaporation and carbonate decay.

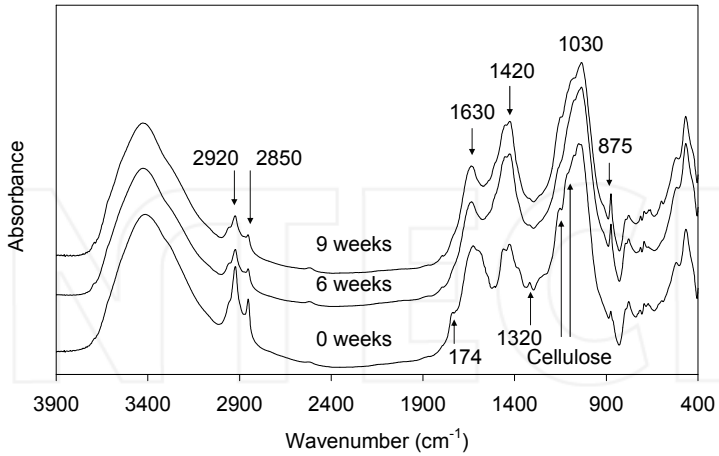


Fig. 7. Changing spectral characteristics of MSW during the biological treatment

Figure 8 illustrates the correlation between the predicted and the measured calorific value of the PLS-R model (Model parameters: R^2 = 76%, RMSECV = 711 kJ kg^{-1} dry matter (DM), RPD = 2.1; bias = 2.1).

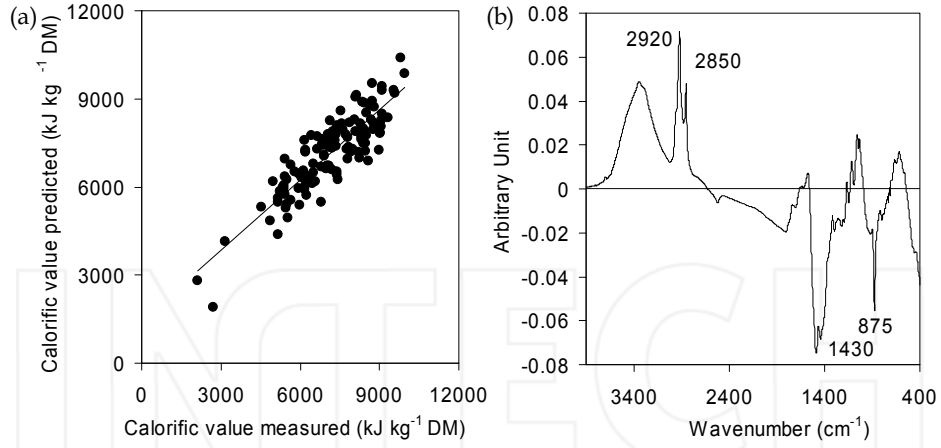


Fig. 8. (a) Correlation between the predicted and the measured calorific values based on the whole MIR spectrum, (b) the corresponding loadings spectrum of the 1st PLS component

3.3 Thermal treatment of MSW - natural and accelerated ageing (carbonation) of MSW incinerator bottom ash

Inorganic residues from incineration have gained in significance during the last decades due to the increase of the thermal treatment of waste for energy recovery. The long-term behaviour of landfilled bottom ash, the process of carbonation and the fate and release of

heavy metals all dependent on the ageing process, have become relevant topics. In figure 9a the main differences between spectra of MSW (I) and MSW incinerator bottom ash (II) are elucidated. Spectra of incineration residues differ from MSW by the loss of bands that are assigned to organic functional groups. Figure 9b illustrates two spectra of MSW incinerator bottom ash before (III) and after CO_2 exposure (IV), two spectra of a bottom ash deposit from a depth of 50 cm (V), from the surface (VI), and a spectrum of the calcite reference. It is evident that the progress of carbonation in abandoned deposits depends on the available surface and the access of air. Carbonates that are part of MSW decay during incineration above 650 °C. The reaction of CaO with water leads to $Ca(OH)_2$ which causes the pH-value of the inorganic residues to increase.

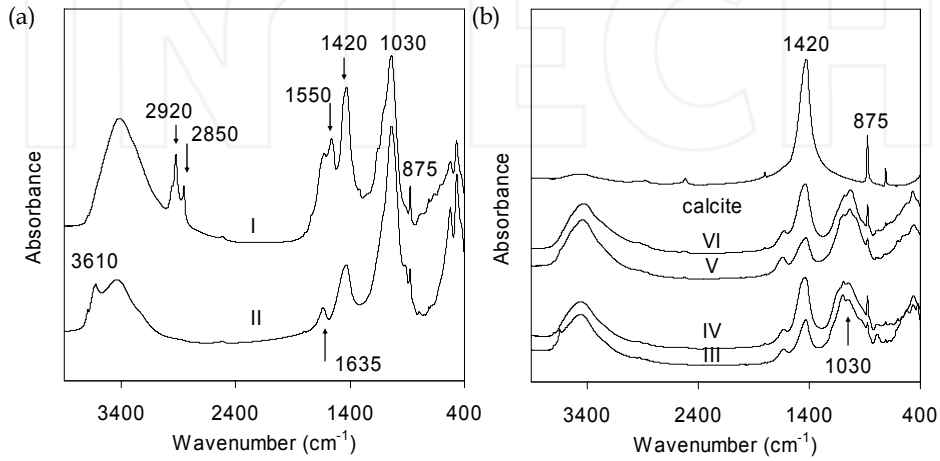


Fig. 9. Spectra of (a) MSW (I) and MSW incinerator bottom ash (II) and (b) MSW incinerator bottom ash (III) before and (IV) after CO_2 exposure, two samples from different depths (V = 50 cm and VI = surface layer) of a bottom ash deposit and the spectrum of the reference (calcite)

The uptake of CO_2 from air effects carbonation. This process in the opposite direction leads to stabilisation, immobilisation of heavy metals and the decrease of the pH-value to a neutral level (Chimenos et al., 2000; Polettini & Pomi, 2004; Rendek et al., 2006). This process can be accelerated by forced CO_2 supply (Mostbauer et al., 2008).

The ageing and stabilisation of incinerator bottom ash are solely chemical processes compared to the stabilisation of waste organic matter that is promoted by microbial activity. Carbonation is reflected by the increase of typical "carbonate bands" (Table 1) in the spectrum. Quantification of carbonates in MSW incinerator bottom ash by means of FT-IR spectra leads to reliable results (Smidt et al., 2009). The standard method is based on the measurement of the CO_2 release from carbonates by acid treatment which can be affected by the complex matrix.

4. Recovery of biogenic waste materials, secondary products

In the context of resource recovery biogenic waste materials have become important ingredients for valuable composts that are applied as soil conditioners. Composting has been practiced for many centuries all over the world at different technical levels (Ahmad et

al., 2007). The variety of biogenic waste materials available from the public separate collection and from industrial processes requires quality criteria and suitable input materials for specific applications. This purpose has been stipulated by the Austrian Compost Ordinance (BMLFUW, 1992). The limitation of heavy metals and pollutants is in the focus of legislation. Nutrients (Courtney & Mullen, 2008), humic substances (Meissl et al., 2007; Smidt et al., 2008d) and phytosanitary properties (Bruns et al., 1996; Erhart et al., 1999) are additional quality criteria that emphasise the positive effects on plants and soils and improve the reputation of composts. Biogenic waste and sewage sludge are usually processed in composting plants whereas manure compost is produced locally by farmers.

4.1 Differentiation of various input materials and assignment of composts to specific biogenic waste materials

Figure 10 displays five composting processes in different plants, grouped by a PCA.

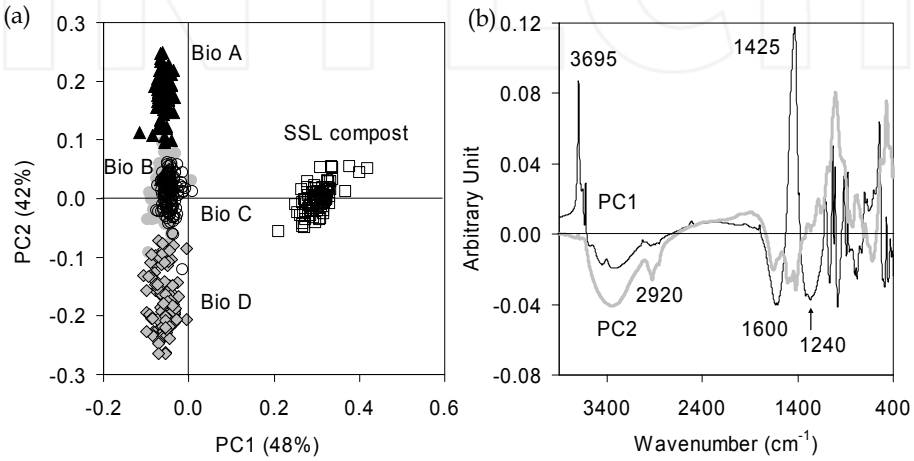


Fig. 10. (a) Scores plot and (b) loadings spectra (PC1 and PC2) of a PCA based on spectral characteristics of five different composting processes (Bio = biowaste, SSL = sewage sludge)

Input materials are distinguished by their characteristic spectral pattern. Along PC1 the particular composition of biowaste and sewage sludge compost contributes to the separation. Biowaste composting processes (Bio A-D) are separated along PC2 according to specific features such as anaerobic pre-treatment (Bio D) or addition of mineral compounds (Bio A). Bio B and Bio C represent the typical mixture of yard and kitchen waste that is processed in open windrow systems.

The loadings spectra of PC1 and PC2 indicate the relevant bands that contribute to the differentiation of composts: mineral components, O-H vibrations of kaolin at 3695 and 910 cm^{-1} (additive to sewage sludge compost) and organic components at 1600 cm^{-1} (aromatic compounds such as humic substances) and at 1240 cm^{-1} (C-O vibrations).

4.2 Process control during the aerobic treatment (composting) of biowaste (yard waste, market waste) and sewage sludge

The degradation of organic matter during composting goes through several phases. The most intensive rotting phase is paralleled by high microbial activity indicated by the high

oxygen demand and the release of metabolic products such as ammonium and volatile fatty acids responsible for odour emissions. During the maturation phase the respiration activity reaches a nearly constant low level. The increasing pH-value affects losses of ammonia. Oxidation of ammonium causes nitrate contents to increase. Compost maturity is characterised by a C/N ratio of 10 – 12 (Ouatmane et al., 2000), low microbial activity and plant compatibility that has to be verified by plant tests according to the Austrian Compost Ordinance (BMLFUW, 1992). Figure 11 and 12 show different stages of three composting processes by spectral characteristics. The indicator bands (cp. Table 1) that are identifiable by visual inspection are the same as indicated for MSW. Nevertheless, the complete pattern of composts differ sufficiently to separate MSW from biogenic waste as shown by the SIMCA classification model (Smidt et al., 2008b). Figure 11a illustrates a typical biowaste composting process over a period of 121 days. Apart from organic matter mineral compounds such as carbonates and clay minerals are usually occurring constituents in biowaste from the separate collection. Their relative increase due to mineralisation of organic substances is obvious. Figure 11b demonstrates the degradation of straw mixed with liquid manure in a fast composter device during a period of nine days. The spectra are dominated by straw characteristics. It is evident that straw is only partially degraded during the short period of composting. Besides the disappearing band at 1720 cm^{-1} and the decreasing band at 1260 – 1230 cm^{-1} that can be assigned to esters in e.g. hemicelluloses (Stewart et al., 1995) and the decreasing bands at 1165 and 1060 cm^{-1} (cellulose marked by arrows in the spectrum) intact cell structures are indicated by sharp and distinct bands in the wavenumber region 1520 - 1200 cm^{-1}. Figure 12 demonstrates the development of primary sludge and anaerobically stabilised sludge from a wastewater treatment plant to sewage sludge compost. The latter originates from a composting plant. The FT-IR spectrum of the primary sludge is dominated by bands that can be assigned to cellulose, hemicelluloses (1740 and 1240 cm^{-1}) and to the microbial biomass that is identified by aliphatic methylene and amide bands. The mature sewage sludge compost features low methylene bands and a considerable nitrate band at 1384 cm^{-1}. Due to the specific mixture in the composting plant the portion of clay minerals is higher, the carbonate content lower.

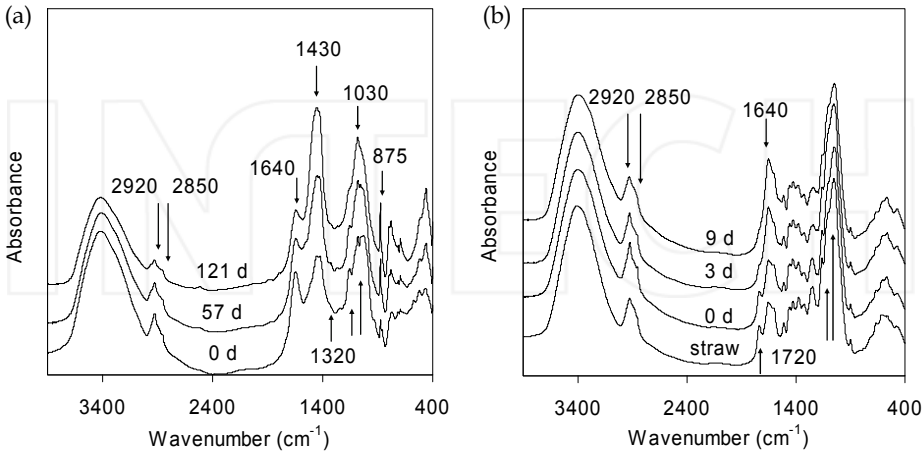


Fig. 11. Development of spectral characteristics during composting processes of (a) biowaste and (b) straw/liquid manure

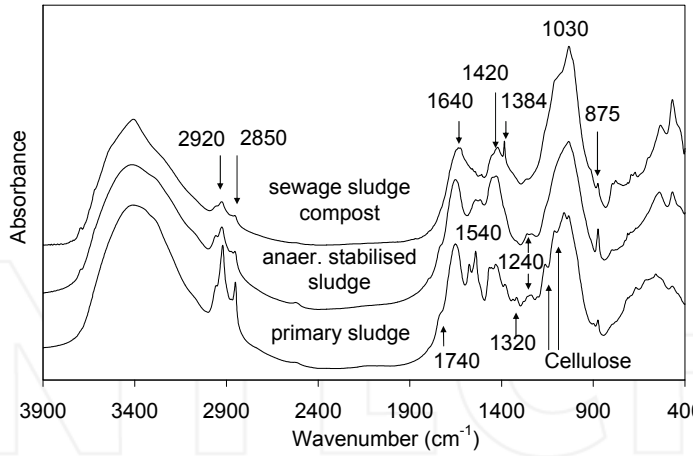


Fig. 12. Spectra of different process stages of sewage sludge treatment and composting

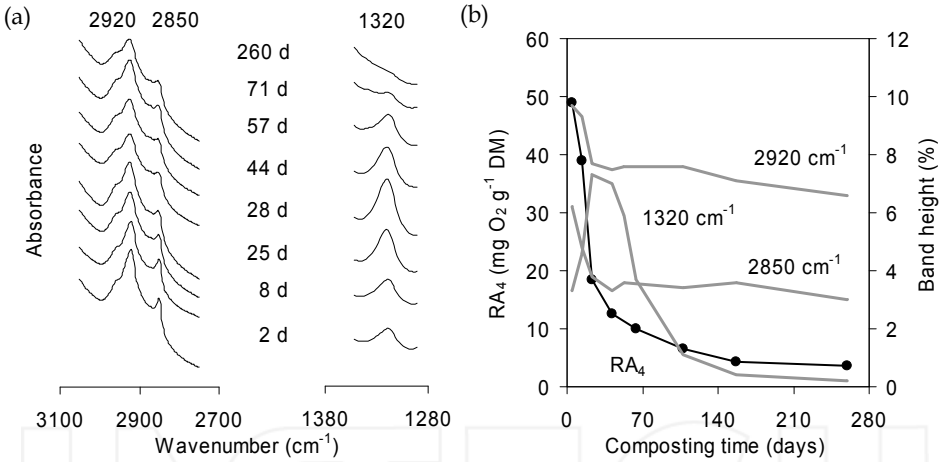


Fig. 13. (a) Decreasing aliphatic methylene bands at 2920 and 2850 cm^{-1}, emerging and disappearing band at 1320 cm^{-1}; (b) development of corresponding relative band heights and respiration activity (RA_4) during a biowaste composting process (2-260 days)

Data evaluation can be performed in different ways. Changing heights or areas of bands indicate the progressing mineralisation. Haberhauer et al. (1998) suggested the ratio of band heights (2920 cm^{-1}/1640 cm^{-1}) as an indicator of stability. Figure 13a shows the development of selected bands (2920, 2850, and 1320 cm^{-1}), figure 13b the corresponding band heights during a biowaste composting process over a period of 260 days. The individual band height is expressed in per cent of ten measured band heights in the spectrum (Smidt et al., 2002). Aliphatic methylene groups are part of many biomolecules with different degradability. Degradation of easily degradable substances is indicated by the decreasing band height at 2920 and 2850 cm^{-1} that reached a low constant level (Fig. 13). Recalcitrant

molecules such as long chain fatty acids and waxes are responsible for the remaining weak bands. They are still found in soil organic matter (Jandl et al., 2002). The emerging and disappearing band at 1320 cm^{-1} that can be assigned to amines shows the typical behaviour of metabolic products that are temporarily visible in the spectrum. The corresponding microbial activity of the composting process is indicated by the respiration activity over a period of 4 days (RA_4). The interaction of microbial activity and changing chemical composition is revealed by the conformity of the curves, each of them reflecting process kinetics.

4.3 Quality assessment of the final product - prediction of parameters using PLS-R

For products such as compost, quality criteria have been defined and require analytical control. Various parameters are involved depending on the waste material to be investigated. Waste organic matter is affected by different mechanisms of degradation and transformation. Composts that comply with quality standards according to the Austrian Compost Ordinance (BMLFUW, 1992) leave the waste management regime and become products for soil amelioration. Mineralisation and humification contribute substantially to stabilisation. Whereas mineralisation causes enrichment of scarcely degradable substances, humification leads to synthesis of new biomacromolecules. Therefore humification counteracts carbon losses by CO_2 release in that it is fixed in humic substances. Extractable humic acids are a suitable parameter for assessing the increase of stable humic substances in composts. Prediction of humic acid contents and respiration activity (RA_4) by FT-IR spectroscopy and PLS-R is an adequate approach for avoiding the time consuming procedure of humic acid (HA) and RA_4 determination. Prediction models being valid for biowaste composts were developed by Meissl et al. (2007; 2008a). Additional PLS-R models for TOC and total nitrogen (TN) were established by Böhm et al. (2010b). Table 3 compiles the parameters for the TOC, TN, RA_4, and HA prediction models. The quality of the models depends on the precision of the reference analyses. Considering this fact all developed MIR based PLS-R models are suitable for practical application, especially for time-consuming parameters such as the determination of humic acid contents and respiration activity.

Model parameters	TOC	TN	RA_4	HA
Wavenumber ranges (cm^{-1})	3000-2800, 1790-1492, 1373-1030			1745-1685, 1610-1567
Calibrated range	9.4 - 27.3 % DM	0.9 - 2.6 % DM	1.0 - 60.6 mg O_2 g^{-1} DM	4.5 - 45.6 % ODM
R^2	89%	83%	88%	88%
RMSECV	1.3 % DM	0.14 % DM	2.9 mg O_2 g^{-1} DM	2.4 % ODM
RPD	3	2.4	2.9	2.9
No. of PLS components	6	9	7	7
Bias	0.005	0.001	-0.01	0.004

Table 3. Parameters of the models for total organic carbon (TOC), total nitrogen (TN), humic acids content (HA), and respiration activity (RA_4); ODM = organic dry matter

Prediction of commonly used parameters in waste management has been reported by several authors, especially for composts (Reeves & Van Kessel, 2000a; Reeves & Van Kessel, 2000b;

Michel et al., 2006; Huang et al., 2007). Apart from organic matter, TOC, TN and nutrient contents, both physical and biological parameters were predicted by FT-NIR spectroscopy.

4.4 Compost application on soils; development of soil organic matter by compost application – characterisation of humic acids by FT-IR spectroscopy

The loss of organic matter in soils by agricultural activities has become a serious problem in many countries (Montanarella, 2002). The application of composts is an appropriate measure for improving the stable organic matter content long-term (Franko et al., 1997; Jensen et al., 1997; Li et al., 1997) as confirmed by field experiments. Figure 14a visualises the increase of humic acid contents (% ODM) in three different plots of agricultural soil where well humified biowaste compost has been applied since 1994, 1999 and 2004 (C). These plots were compared to the reference soils (R) without any amendment. The spectral region of aromatic compounds where humic acids are expected (1610 - 1567 cm^{-1}) were selected (Meissl et al., 2007; Meissl et al., 2008a). The increase is visualised by the highlighted area in figure 14b (1610 - 1567 cm^{-1}) between the spectra of the reference soil (R) and of the soil with compost application (C). It corresponds to the duration of compost application (Smidt et al., 2008c). Humic acids belong to the stable organic matter fraction in soils with low turnover rates. Regarding the age of soil humic acids and the diversity of these molecules the question arises, if compost humic acids are similar regarding the composition and stability. Figure 15a demonstrates the changing spectral pattern of compost humic acids after 4, 25 and 260 days of composting. In the scarcely composted material humic acids feature strong aliphatic methylene bands and bands that can be assigned to polysaccharides (1150 –950 cm^{-1}). In the mature compost, spectral characteristics become similar to soil humic acids. The intensity of the band at 1030 cm^{-1} is caused by impurities of clay minerals that were not completely separated by the extraction procedure of soil humic acids (Smidt et al., 2008d).

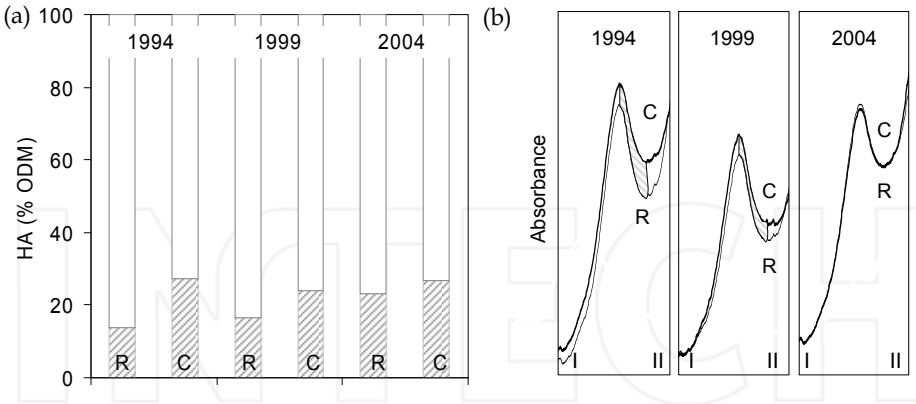


Fig. 14. Increasing contents of stable organic matter in agricultural soil due to compost application since 1994, 1999 and 2004, verified by (a) increasing humic acid (HA) contents (shaded bar) and (b) corresponding spectral characteristics (C = compost amended soil, R = reference soil, presented wavenumber region 1795 (I) – 1505 (II) cm^{-1})

Lignin is known to serve as a precursor in humic substance formation (Tan, 2003). A lab scale composting process with biogenic waste was carried out to reveal the effect of lignin on humic acid synthesis. Addition of 5% lignin from an industrial process could improve

the humic acid content in the mature compost compared to the reference without lignin application. Figure 15b shows the spectra of freeze-dried humic acids extracted from the fresh biogenic waste with lignin addition (1 d), from the 42-day-old compost with lignin (42 d) and from the reference without lignin (R-42 d). Whereas spectral characteristics of lignin are clearly visible at the beginning (1 d), they disappear during the progressing process (42 d). Humic acids become similar to the reference. Based on FT-IR spectra it can be assumed that lignin building blocks were integrated in the humic acid molecule (Smidt et al., 2008a).

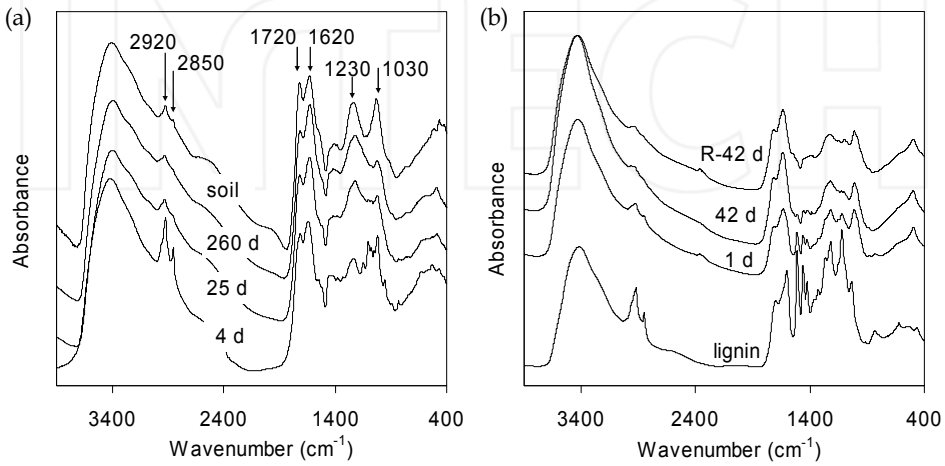


Fig. 15. (a) Spectral pattern of humic acids originating from a biowaste composting process (4, 25, 260 days) and from soil; (b) integration of lignin in compost humic acids

5. Conclusion

The application of FT-IR spectroscopy in waste management provides fast, cheap and reliable process and product control. Moreover, environmental monitoring can be based on spectral characterisation. Besides conventional methods of spectral data evaluation multivariate statistical methods enable the extraction of the inherent information. The development of classification and parameter prediction models is a prerequisite for using this analytical tool in waste management practice. There are many perspectives for the adoption of the method in this area. Depending on the questions to be answered tailor made models can be developed in order to replace time-consuming analyses. The strength of FT-IR spectroscopic investigations is the comprehensive characterisation of complex materials by their unique signature. Due to the small sample amount used for infrared spectroscopic measurements careful sample preparation is a crucial issue in view of the heterogeneous mixture of waste materials.

6. References

Adani, F.; Tambone, F. & Gotti, A. (2004). Biostabilization of municipal solid waste. *Waste Management*, 24, 775-783, 0956053X

Ahmad, R.; Jilani, G.; Arshad, M.; Zahir, Z.A. & Khalid, A. (2007). Bio-conversion of organic wastes for their recycling in agriculture: An overview of perspectives and prospects. *Annals of Microbiology*, 57, 471-479, 15904261

Austrian Standards Institute (2004). *Stability parameters describing the biological reactivity of mechanically-biologically pretreated residual wastes - Part 2: Gas generation by incubation test (GS_{21})*, OENORM S 2027-2

Bacsik, Z.; Mink, J. & Keresztury, G. (2005). FTIR spectroscopy of the atmosphere Part 2. Applications. *Applied Spectroscopy Reviews*, 40, 327-390, 05704928

Barlaz, M.A. (1998). Carbon storage during biodegradation of municipal solid waste components in laboratory-scale landfills. *Global Biogeochemical Cycles*, 12, 373-380, 08866236

Barlaz, M.A.; Ham, R.K. & Schaefer, D.M. (1990). Methane production from municipal refuse: A review of enhancement techniques and microbial dynamics. *Critical Reviews in Environmental Control*, 19, 557-584, 1040838X

Binner, E. & Zach, A. (1999). Biological reactivity of residual wastes and dependence on the duration of pretreatment. *Waste Management and Research*, 17, 543-554, 0734242X

BMLFUW (1990). *Abfallwirtschaftsgesetz,* BGBl. Nr. 325/1990

BMLFUW (1992). *Verordnung zur getrenten Erfassung von Bioabfällen - Austrian ordinance for the seperate collection of biogenic waste from households,* BGBl. 68/1992

BMLFUW (2008). *Verordnung des Bundesministers für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft über Deponien (Deponieverordnung 2008),* BGBl. Nr. II 39/2008

Böhm, K.; Smidt, E.; Binner, E.; Schwanninger, M.; Tintner, J. & Lechner, P. (2010a). Determination of MBT-waste reactivity - An infrared spectroscopic and multivariate statistical approach to identify and avoid failures of biological tests. *Waste Management*, 30, 583-590, 0956053X

Böhm, K.; Smidt, E.; Binner, E.; Tintner, J. & Lechner, P. (2010b). Prediction of various quality parameters of biowaste composts by means of FourierTransform infrared (FTIR) spectroscopy. *Proceedings of 7th international conference ORBIT 2010*, 9789606865251, Heraklion, Greece, Grafima publications, Thessaloniki, Greece

Bonifazi, G.; Serranti, S.; Bonoli, A. & Dall'Ara, A. (2009). Innovative recognition-sorting procedures applied to solid waste: The hyperspectral approach. *Proceedings of WIT Transactions on Ecology and the Environment*, pp. 885-894, 9781845641818,

Brereton, R.G. (2002). *Chemometrics: Data analysis for the laboratory and chemical plant*. John Wiley & Sons Ltd., 0471489778, Chichester, England

Bruns, C.; Ahlers, S.; Gattinger, A.; Schuler, C.; Vogtmann, H. & Wolf, G. (1996). The suppressive effects of composted seperately collected organic waste and yard waste compost on two important soilborne plant pathogens, In: *Science of composting*, M. de Bertoldi, P. Sequi, B. Lemmes and T. Papi (Ed.), 1094-1095, Chapman & Hall, 0751403830, England

Chen, Y. (2003). Nuclear Magnetic Resonance, Infra-Red and Pyrolysis: Application of Spectroscopic Methodologies to Maturity Determination of Composts. *Compost Science & Utilization*, 11, 152-168, 1065657X

Chimenos, J.M.; Fernández, A.I.; Nadal, R. & Espiell, F. (2000). Short-term natural weathering of MSWI bottom ash. *Journal of Hazardous Materials*, 79, 287-299, 03043894

Courtney, R.G. & Mullen, G.J. (2008). Soil quality and barley growth as influenced by the land application of two compost types. *Bioresource Technology*, 99, 2913-2918, 09608524

Dou, Y.; Zhang, W.; Zhang, Y.; Zhou, H.; Wang, D. & Xu, J. (2006). Identification of plastics by fourier transform raman and infrared spectroscopy. *Fenxi Huaxue/ Chinese Journal of Analytical Chemistry*, 34, 1615-1618, 02533820

Erhart, E.; Burian, K.; Hartl, W. & Stich, K. (1999). Suppression of Pythium ultimum by biowaste composts in relation to compost microbial biomass, activity and content of phenolic compounds. *Journal of Phytopathology*, 147, 299-305, 09311785

Esbensen, K. (2002). *Multivariate Data Analysis - in practice*. Alborg University, 8299333032, Esbjerg

Feldhoff, R.; Wienke, D.; Cammann, K. & Fuchs, H. (1997). On-line post consumer package identification by NIR spectroscopy combined with a FuzzyARTMAP classifier in an industrial environment. *Applied Spectroscopy*, 51, 362-368, 00037028

Franko, U.; Crocker, G.J.; Grace, P.R.; Klír, J.; Körschens, M.; Poulton, P.R. & Richter, D.D. (1997). Simulating trends in soil organic carbon in long-term experiments using the CANDY model. *Geoderma*, 81, 109-120, 00167061

Galle, B.O.; Samuelsson, J.; Svensson, B.H. & Borjesson, G. (2001). Measurements of methane emissions from landfills using a time correlation tracer method based on FTIR absorption spectroscopy. *Environmental Science and Technology*, 35, 21-25, 0013936X

Grube, M.; Zagreba, E.; Gromozova, E. & Fomina, M. (1999). Comparative investigation of the macromolecular composition of mycelia forms Thielavia terrestris by infrared spectroscopy. *Vibrational Spectroscopy*, 19, 301-306, 09242031

Haberhauer, G.; Rafferty, B.; Strebl, F. & Gerzabek, M.H. (1998). Comparison of the composition of forest soil litter derived from three different sites at various decompositional stages using FTIR spectroscopy. *Geoderma*, 83, 331-342, 00167061

Hegde, U.; Chang, T.C. & Yang, S.S. (2003). Methane and carbon dioxide emissions from Shan-Chu-Ku landfill site in northern Taiwan. *Chemosphere*, 52, 1275-1285, 00456535

Hesse, M.; Meier, H. & Zeeh, B. (1995). *Spektroskopische Methoden in der organischen Chemie*. Georg Thieme Verlag, 3135761053, Stuttgart

Huang, G.; Han, L. & Liu, X. (2007). Rapid estimation of the composition of animal manure compost by near infrared reflectance spectroscopy. *Journal of Near Infrared Spectroscopy*, 15, 387-394, 09670335

Jandl, G.; Schulten, H.R. & Leinweber, P. (2002). Quantification of long-chain fatty acids in dissolved organic matter and soils. *Journal of Plant Nutrition and Soil Science*, 165, 133-139, 14368730

Jensen, L.S.; Mueller, T.; Nielsen, N.E.; Hansen, S.; Crocker, G.J.; Grace, P.R.; Klír, J.; Körschens, M. & Poulton, P.R. (1997). Simulating trends in soil organic carbon in long-term experiments using the soil-plant-atmosphere model DAISY. *Geoderma*, 81, 5-28, 00167061

Kulcke, A.; Gurschler, C.; Spöck, G.; Leitner, R. & Kraft, M. (2003). On-line classification of synthetic polymers using near infrared spectral imaging. *Journal of Near Infrared Spectroscopy*, 11, 71-81, 09670335

Leitner, R.; Mairer, H. & Kercek, A. (2003). Real-time classification of polymers with NIR spectral imaging and blob analysis. *Real-Time Imaging*, 9, 245-251, 10772014

Li, C.; Frolking, S.; Crocker, G.J.; Grace, P.R.; Klír, J.; Körchens, M. & Poulton, P.R. (1997). Simulating trends in soil organic carbon in long-term experiments using the DNDC model. *Geoderma*, 81, 45-60, 00167061

Li, H.; Kumagai, M.; Takahashi, Y.; Amano, T.; Fujiwara, K.; Yoshimura, N. & Ogawa, N. (2005). Effect of color and shape on discrimination of plastic by a portable near-infrared spectrometer. *International Journal of the Society of Material Engineering for Resources*, 17, 33-38, 13479725

Luiken, A. & Bos, P. (2010). Automated identification and sorting of post-consumer textile waste. *Melliand International*, 16, 50-51, 09479163

Madejova, J. (2003). FTIR techniques in clay mineral studies. *Vibrational Spectroscopy*, 31, 1-10, 09242031

Mantsch, H.H. & Chapman, D. (1996). *Infrared Spectroscopy of Biomolecules*. Wiley, 0471021849, New York

Meissl, K.; Smidt, E. & Schwanninger, M. (2007). Prediction of humic acid content and respiration activity of biogenic waste by means of Fourier transform infrared (FTIR) spectra and partial least squares regression (PLS-R) models. *Talanta*, 72, 791-799, 00399140

Meissl, K.; Smidt, E.; Schwanninger, M. & Tintner, J. (2008a). Determination of humic acids content in composts by means of near- and mid-infrared spectroscopy and partial least squares regression models. *Applied Spectroscopy*, 62, 873-880, 00037028

Meissl, K.; Smidt, E. & Tintner, J. (2008b). Reproducibility of Fourier transform infrared spectra of compost, municipal solid waste, and landfill material. *Applied Spectroscopy*, 62, 190-196, 00037028

Michel, K.; Bruns, C.; Terhoeven-Urselmans, T.; Kleikamp, B. & Ludwig, B. (2006). Determination of chemical and biological properties of composts using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 14, 251-259, 09670335

Montanarella, L. (2002). Organic matter levels in European Agricultural soils. *Proceedings of „Biological Treatment of Biodegradable Waste – Technical Aspects“*, Brussels, 8.-10.4.2002, EUR 20517 EN, 2003,

Moron, A. & Cozzolino, D. (2004). Determination of potentially mineralizable nitrogen and nitrogen in particulate organic matter fractions in soil by visible and near-infrared reflectance spectroscopy. *Journal of Agricultural Science*, 142, 335-343, 00218596

Mostbauer, P.; Lenz, S. & Lechner, P. (2008). MSWI bottom ash for upgrading of biogas and landfill gas. *Environmental Technology*, 29, 757-764, 09593330

Naumann, D.; Schultz, C.P. & Helm, D. (1996). What can infrared spectroscopy tell us about the structure and composition of intact bacterial cells?, In: *Infrared Spectroscopy of Biomolecules*, H. H. Mantsch and D. Chapman (Ed.), 279-310, Wiley, 0471021849, New York

Ouatmane, A.; Provenzano, M.R.; Hafidi, M. & Senesi, N. (2000). Compost maturity assessment using calorimetry, spectroscopy and chemical analysis. *Compost Science & Utilization*, 8, 124-134, 1065657X

Polettini, A. & Pomi, R. (2004). The leaching behavior of incinerator bottom ash as affected by accelerated ageing. *Journal of Hazardous Materials*, 113, 209-215, 03043894

Pollanen, K.; Hakkinen, A.; Reinikainen, S.P.; Rantanen, J.; Karjalainen, M.; Louhi-Kultanen, M. & Nystrom, L. (2005). IR spectroscopy together with multivariate data analysis as a process analytical tool for in-line monitoring of crystallization process and

solid-state analysis of crystalline product. *Journal of Pharmaceutical and Biomedical Analysis*, 38, 275-284, 07317085

Reeves, J.B. & Van Kessel, J.S. (2000a). Determination of ammonium-N, moisture, total C and total N in dairy manures using a near infrared fibre-optic spectrometer. *Journal of Near Infrared Spectroscopy*, 8, 151-160, 09670335

Reeves, J.B. & Van Kessel, J.S. (2000b). Near-Infrared spectroscopic determination of carbon, total nitrogen, and ammonium-N in dairy manures. *Journal of Dairy Science*, 83, 1829-1836, 00220302

Reig, F.B.; Adelantado, J.V.G. & Moreno, M.C.M.M. (2002). FTIR quantitative analysis of calcium carbonate (calcite) and silica (quartz) mixtures using the constant ratio method. Application to geological samples. *Talanta*, 58, 811-821, 00399140

Rendek, E.; Ducom, G. & Germain, P. (2006). Carbon dioxide sequestration in municipal solid waste incinerator (MSWI) bottom ash *Journal of Hazardous Materials*, 128, 73-79, 03043894

Smidt, E.; Lechner, P.; Schwanninger, M.; Haberhauer, G. & Gerzabek, M.H. (2002). Characterization of waste organic matter by FT-IR spectroscopy: Application in waste science. *Applied Spectroscopy*, 56, 1170-1175, 00037028

Smidt, E. & Meissl, K. (2007). The applicability of Fourier Transform Infrared (FT-IR) spectroscopy in waste management. *Waste Management* 27, 268-276, 0956053X

Smidt, E.; Meissl, K.; Schmutzer, M. & Hinterstoisser, B. (2008a). Co-composting of lignin to build up humic substances-Strategies in waste management to improve compost quality. *Industrial Crops and Products*, 27, 196-201, 09266690

Smidt, E.; Meissl, K.; Schwanninger, M. & Lechner, P. (2008b). Classification of waste materials using Fourier transform infrared spectroscopy and soft independent modeling of class analogy. *Waste Management*, 28, 1699-1710, 0956053X

Smidt, E.; Meissl, K. & Tintner, J. (2007). Investigation of 15-year-old municipal solid waste deposit profiles by means of FTIR spectroscopy and thermal analysis. *Journal of Environmental Monitoring*, 9, 1387-1393, 14640325

Smidt, E.; Meissl, K.; Tintner, J.; Binner, E. & Lechner, P. (2008c). Influence of composts and digestates on soil properties - Retrospect and outlook. *Proceedings of CODIS 2008 - Compost and digestate: sustainability, benefits and impacts for the environment and plant production*, pp. 299, 9783037360163, CH-Solothurn, Research Institute of Organic Agriculture FIBL, CH-Frick

Smidt, E.; Meissl, K.; Tintner, J. & Ottner, F. (2009). Interferences of carbonate quantification in municipal solid waste incinerator bottom ash: evaluation of different methods. *Environmental Chemistry Letters*, 1-6, 16103653

Smidt, E. & Schwanninger, M. (2005). Characterization of Waste Materials Using FT-IR Spectroscopy – Process Monitoring and Quality Assessment. *Spectroscopy Letters*, 38, 247-270, 00387010

Smidt, E.; Tintner, J.; Meissl, K. & Binner, E. (2008d). Influence of input materials and composting operation on humification of organic matter. *Dynamic Soil, Dynamic Plant*, 2 (Special Issue 1), 50-59, 17496500

Smith, B. (1999). *Infrared Spectral Interpretation*. CRC Press, Boca Raton, 0849324637, London, New York, Washington; DC

Socrates, G. (2001). *Infrared and Raman Characteristic Group Frequencies. Tables and Charts*. John Wiley & Sons Ltd., 0471852988, Chichester

Stewart, D.; Wilson, H.M.; Hendra, P.J. & Morrison, I.M. (1995). Fourier-transform infrared and Raman spectroscopic study of biochemical and chemical treatments of oak wood (Quercus rubra) and Barley (Hordeum vulgare) straw. *Journal of Agricultural and Food Chemistry*, 43, 2219-2225, 00218561

Tan, K.H. (2003). *Humic Matter in Soil and the Environment – Principles and Controversies*. Marcel Dekker, 0824742729, New York Basel

Tesar, M.; Prantl, R. & Lechner, P. (2007). Application of FT-IR for assessment of the biological stability of landfilled municipal solid waste (MSW) during in situ aeration. *Journal of Environmental Monitoring*, 9, 110-118, 14640325

Tintner, J.; Smidt, E.; Böhm, K. & Binner, E. (2010). Investigations of biological processes in Austrian MBT plants. *Waste Management*, 30, 1903-1907, 0956053X

Van Den Broek, W.H.A.M.; Wienke, D.; Melssen, W.J. & Buydens, L.M.C. (1998). Plastic material identification with spectroscopic near infrared imaging and artificial neural networks. *Analytica Chimica Acta*, 361, 161-176, 00032670

van Praagh, M.; Heerenklage, J.; Smidt, E.; Modin, H.; Stegmann, R. & Persson, K.M. (2009). Potential emissions from two mechanically-biologically pretreated (MBT) wastes. *Waste Management*, 29, 859-868, 0956053X

Weber, K.; Lamp, T.; Weidemann, J.; Haren, G.V.; Eisenmann, T.; Mosebach, H.; Gärtner, A. & Bröker, G. (1996). FTIR-langwegabsorptionsspektroskopie zur Messung von Luftverunreinigungen: Grundlagen und ausgewählte Beispiele von Messungen bei Industrieanlagen, im urbanen Umfeld und bei Deponien, In: *VDI Berichte*, 205-255, 00835560

Williams, P. & Norris, K. (2004). *Near-Infrared technology in the agricultural and food industries*. American Association of Cereal Chemists, Inc., 1891127241, St. Paul

Zaccheo, P.; Ricca, G. & Crippa, L. (2002). Organic matter characterization of composts from different feedstocks. *Compost Science & Utilization*, 10, 29-38, 1065657X

Zhang, A.D.; Zeng, W.X.; Niemczyk, T.M.; Keenan, M.R. & Haaland, D.M. (2005). Multivariate analysis of infrared spectra for monitoring and understanding the kinetics and mechanisms of adsorption processes. *Applied Spectroscopy*, 59, 47-55, 00037028

20

Open-Path FTIR Detection of Explosives on Metallic Surfaces

John R. Castro-Suarez[1], Leonardo C. Pacheco-Londoño[1], Miguel Vélez-Reyes[2], Max Diem[3] and Thomas J. Tague, Jr.[4] and Samuel P. Hernandez-Rivera[1]
ALERT DHS Center of Excellence for Explosives
Center for Chemical Sensors Development
[1]Department of Chemistry, University of Puerto Rico, Mayagüez, PR
[2]Department of Electrical Engineering, University of Puerto Rico, Mayagüez, PR
[3]Department of Chemistry, Northeastern University, Boston, MA
[4]Bruker Optics, Billerica, MA 01821
[1,2]Puerto Rico
[3,4]U.S.A

1. Introduction

Defense and security agencies are in constant demand of new ways of detecting chemical and biological threats posed by terrorist organizations. Fundamental and applied research in areas of interest to national defense and security focus in detection of highly energetic materials (HEM), including homemade explosives (HME) that could be used as weapons of mass destruction (Marshal and Oxley, 2009; Yinon and Zitrin, 1996; Schubert and Rimski-Korsakov, 2006). The Department of Homeland Security of the United States of America has even gone a step further and established a university based Center of Excellence in explosives detection, mitigation and response to conduct transformational research, technology and educational development for effective characterization, detection, mitigation and response to the explosives-related threats facing the country and the world. The Awareness and Localization of Explosives Threats (AWARE) is co-lead by Northeastern University (ALERT at NU, Boston, MA) and University of Rhode Island (ALERT at URI, Kingston, RI).

Current detection methods of explosives are based on a wide variety of technologies that focus on either bulk explosives or traces of explosives. Bulk explosives can be detected indirectly by imaging characteristic shapes of the explosive charge, detonators, and wires or directly by detecting the chemical composition or dielectric properties of the explosive material. Trace detection methods rely on detection of vapors emitted from the explosives or on explosive particles that are deposited on nearby surfaces (National Academy of Sciences Committee, 2004). Although there are hundreds of publications about methods of detection of HEM in water, soil, air, clothing, surfaces, etc. and these offer the advantage of providing very low limits of detection at ppb levels (Caron, et al., 2010; Hilmi and Luong, 2000; Yinon, 1996; Szakal and Brewer, 2009; Miller and Yoder, 2010). They require, in the majority of the cases, sampling at the scene followed by a sample preparation step, to be later analyzed by a

particular technique. Sampling and sample preparation are among the main disadvantages in HEM detection, in many cases threatening the health and life of analysts and first responders. Vibrational spectroscopy, in its various modalities, has shown to be useful for detection of dangerous chemicals, among them HEM. Near-infrared or mid-infrared spectroscopies have shown to be powerful techniques for IR vibrational analysis, able to detect organic and inorganic substances in any state: solid, liquid or gas (Gunzler, 2002; Smith, 2000). IR vibrational spectra can to be used for identify and quantify samples in complex matrices because each substance has its own fingerprint spectrum in the mid IR (MIR). This means that IR spectroscopy can be used for discriminant analysis even when the target analyte is in very small quantities (Bangalore, 1997).

Standoff detection is defined as where equipment and operator stay away from the sample while measuring some property of the target (Parmenter, 2004). An area of IR spectroscopy that has increased interest in defense and security applications is standoff IR (SOIR) spectroscopy. In SOIR detection, vibrational signatures can to be recorded from several meters to hundreds of meters in distances between the target and the instrument. Fourier transform infrared (FTIR) standoff detection provides a means of doing real time analysis, in which no sample preparation is needed, no human contact needs to be provided, measurements are typically fast, and chemical information for each explosive can be obtained in high detail which can allow identification and even quantification. This makes the standoff IR detection a powerful technique for sensing of energetic materials at a distance, thus preventing or minimizing the damage caused by terrorist action, in the case that this comes to be detonated.

Open-Path Fourier Transform IR (OP/FTIR) spectroscopy has been used for atmospheric gas analysis and environmental monitoring for over 40 years (Griffiths, et al., 2009). It is one of the two methodologies devised for measuring concentration of gaseous trace components in the atmosphere using infrared spectroscopy: extractive sampling analysis and *in situ* open-path analysis. Although Russwurm and Childers credit Hanst for the initial description of FTIR monitoring of atmospheric pollutants in the atmosphere by OP IR (Russwurm and Childers, 2001; Hanst, 1971) Stephens and his group at the Environmental Protection Agency had already made measurements of ambient concentrations of peroxy acetyl nitrate (PAN) in the Los Angeles city basin before 1969 (Stephens, et al., 1969; Scott, et al., 1957). Aside from the apparently inconsequent controversy (since Hanst was part of the Stephen's group) valid questions on why is OP/FTIR is rarely used nowadays and why its development has been undermined with technological problems remain unanswered. Among possible answers to these questions stands out a limited sensitivity of the technique for atmospheric monitoring: 1-100 ppb by volume (contrasted to the requirements of parts per trillion by volume on many pollutants) and the lack of development of algorithms that can be incorporated into the instruments acquisition and analysis routines (software) that can could make the use of the technique a more amenable and user friendly one. In a recent article by Griffiths and collaborators, the authors point out the difficulties encountering when using OP/FTIR that have led to a slow development of the remote sensing modality (Griffiths, et al., 2009). The clear advantages of wide area sensing and long range capabilities have been overshadowed by hardware and more so, by software limitations. Inadequacy in compensating for variable atmospheric contributions, mainly by water vapor and carbon dioxide has hampered the wide usage of OP/FTIR both in environmental studies as well as in Defense and Security applications. In this study, two types of FTIR standoff detection experiments were carried out: active mode OP/FTIR and passive mode OP/FTIR. The

detection of particle dispersion is also relevant to the forensic community. When an airbag deploys in an accident the lubricant is dispersed on the passenger. Many times the occupants in the vehicle are ejected or end up in a different location in the vehicle. Such a device could definitively place the occupants. Additionally, persons involved in the illicit manufacture of drugs will frequently retain chemical residues on their clothing. In this case, remote detection could be used to link a suspect to the crime scene.

There is a limited number of scientific contributions in the area of SOIR detection of HEM deposited on surfaces. Work by Theriault and colleagues (Theriault, et al., 2004), Van Neste and collaborators (Van Neste, et al., 2009), Blake and co-workers (Blake, et al., 2009) and Pacheco-Londoño and colleagues (Pacheco-Londoño, et al. 2009) have helped to contribute the development of this application of OP/FTIR. Theriault and collaborators made field measurements of liquid contaminants deposited on a number of surfaces at a standoff distance of 60 m using FTIR radiometry (Theriault, et al., 2004). Van Neste and collaborators described standoff detection measurements of trace quantities of surface adsorbed high explosives (Van Neste, et al., 2009). They used two quantum cascade lasers (QCL) operated simultaneously in the MIR, with tunable wavelength windows that match with absorption peaks of nitroexplosives tested. In this important contribution researchers demonstrated a sensitivity of 100 ng/cm^2 and a standoff detection distance of 20 m for surface adsorbed analytes such as explosives and chemical agent simulant tributyl phosphate. The detection of Explosives on metallic substrates is the first step in demonstrating the facility of passive and active open path FTIR detection for general use. Other substrates such as textiles, plastic, wood, and glass are less reflective and present a greater challenge. The emergence of alternative bright sources, such as QCL's, puts the active detection of explosive residues on real life materials over significant pathlengths within reach. Blake and colleagues recorded hyperspectral images of galvanized steel plates, containing cyclotrimethylenetrinitramine (RDX), using a commercial long-wave infrared imaging spectrometer at a standoff range between 14 and 50 m (Blake, et al., 2009). Pacheco-Londoño and collaborators built an active IR standoff detection system by coupling a bench FTIR interferometer to a gold mirror and external cryocooled detector assembly for detection of explosives present as traces on reflective surfaces (Pacheco-Londoño, et al. 2009). Source-target distances in the range of 1 – 3.7 m were studied and limits of detection (LOD) values obtained were 18 and 20 μg/cm^2 for TNT and RDX, respectively. The results of the prototype built were attributed to the use of a modulated MIR beam source that was able to cut down stray light from laboratory illumination.

2. Description of methodology

Open Path Infrared Spectroscopy is a well established technique for atmospheric sensing of gases and condensable vapors. In the current application, after validating the spectroscopic system in detection of gases and condensable vapors, are more challenging application was addressed: detection solid samples deposited as trace contaminants on metallic surfaces were detected by OP/FTIR. Sample preparation is a critical task in the development of any analytical technique. Three steps were performed for standoff detection of explosives and other highly energetic materials deposited on Al plates:

- TNT samples were weighted and dissolved in dichloromethane.
- Mixtures were deposited on the Al plates using a Teflon stub and were then allowed to air-dry.

- OP/FTIR standoff detection experiments were carried out, both in active mode using a mid infrared (MIR) source and passive mode using a thermal excitation (utilizing a tungsten lamp).

These steps are illustrated in Fig. 1.

Background spectra of Al plates with no TNT deposited on them were run for every standoff distance tested in active mode. In the case of passive mode standoff detection, background spectra were acquired at every range value and every plate temperature tested. Then, statistical routines were applied using chemometrics. In particular, partial least squares (PLS) regression analysis was used to perform quantification studies of HEM surface loadings at all standoff distances. Standoff detection of solid samples present as traces on metallic substrates required a sample preparation methodology that would be able to deposit solid samples on a solid substrate, with high coverage uniformity and reproducibility. Due to the size of the substrates, sample smearing technique was used to deposit the HEM TNT at trace amounts on metallic substrates (Primera-Pedrozo, 2008). As shown in Fig. 2a, aluminum plates of areas 30.5 cm × 30.5 cm were used as material support for HEM samples. Dichloromethane was used to clean the aluminum surfaces. Plates were allowed to air-dry before of depositing the desired HEM surface loading. A small amount of dichloromethane was used to dissolve TNT sample to be deposited. A 3 cm × 15 cm Teflon stub was used to smear the HEM sample on the aluminum substrates (Fig. 2b). The amount of HEM that remained on the Teflon stub after sample smearing was negligible. The nominal surface concentrations obtained by the smearing technique were 50, 100, 200, 300, and 400 μg/cm² of HEM. Figs. 2b and 2c show how TNT was deposited on aluminum plates for 50 (Fig. 1b) and 100 μg/cm² (Fig. 1c), respectively.

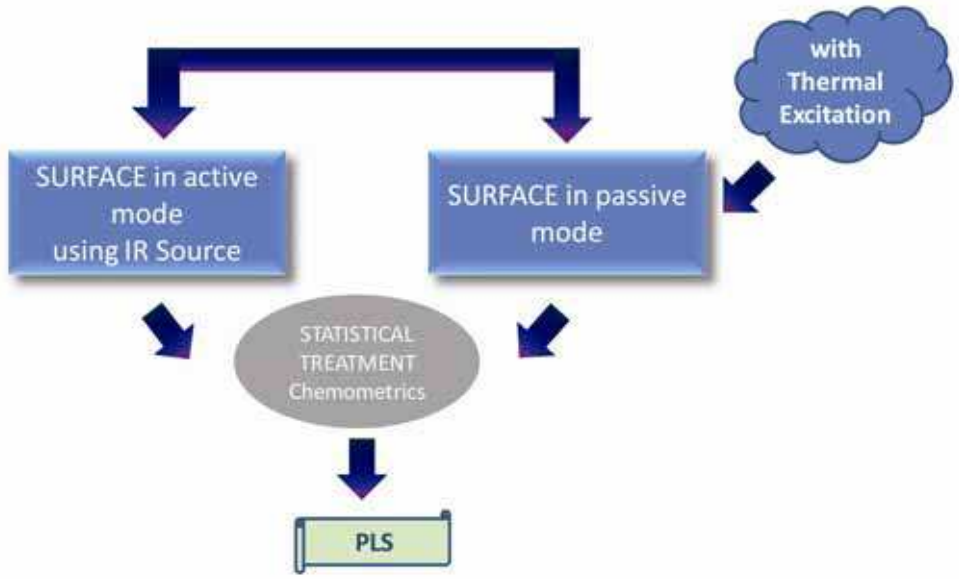


Fig. 1. Steps for remote detection of nitroexplosives and other HEM deposited on surfaces using OP/FTIR

a) b) c)

Fig. 2. Sample preparation: (a) Clean Al plate ready for sample smearing technique; (b) 50 µg/cm² of TNT loading concentration; (c) 100 µg/cm² of TNT surface concentration

A model EM27 (Bruker Optics, Billerica, MA) OP/FTIR spectrometer, was used to obtain the MIR spectral information of TNT samples. Table 1 contains the specifications and technical data of the EM27. Fig. 3 illustrates the difference in operation between a benchtop FTIR spectrometer used in absorption mode (Fig. 3a) and an IR interferometer configured for open-path measurements (Fig. 3b). The optical bench consisted of a compact, enclosed, and desiccated Michelson type interferometer equipped with ZnSe windows, internal black body calibration source, KBr beamsplitter, f/0.9 and a field of view (FOV) of 30 mrad (1.7°). Its main features are: permanently aligned, vibration insensitive, and friction-free mechanical bearing. The system was capable of acquiring 32 spectra per second at 1 cm^{-1} resolution.

PARAMETER	SPECIFICATION
Spectral Range:	700 – 1300 cm-1 (useful for passive measurements) 700 – 4000 cm-1 (useful for active measurements)
Resolution:	1.0 cm-1 (Option: 0.5 cm-1)
NEDT: (Noise Equivalent Delta Temperature):	up to 0.08°C for one scan with a resolution of 1 cm^{-1} and a mirror speed of 40 kHz, depending on the detector
Optical Bench:	compact, enclosed, desiccated, purge interferometer
Beamsplitter:	KBr-substrate with multi-layer coating (Option: ZnSe)
Interferometer:	vibration insensitive, friction-free mechanical bearing; permanently aligned; symmetrical interferogram acquisition at 4 scanning velocities up to 40 kHz
Standard Detector:	MCT-detector (narrow band), liquid nitrogen cooled,
Field of View:	30 mrad (10 mrad with receiver telescope)

Table 1. Specifications of EM27 Open Path FTIR spectrometer

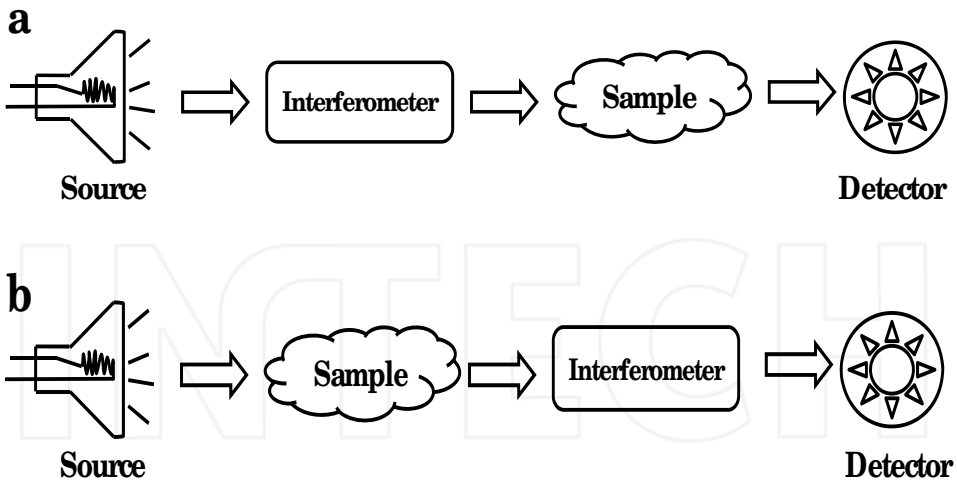


Fig. 3. FTIR interferometer configured as: (a) absorption spectrometer: detects source radiation only; (b) open-path spectrometer: detects source and sample radiation

High sensitivity measurements were achieved by using a high sensitivity closed cycle cryo-cooled photoconductive mercury cadmium telluride (MCT) MIR detector. Both, the MIR source and interferometer were focused on the target by MIR telescopes for active mode measurements (Fig. 3). Telescopic sights with mounts were used to align both source and collector. The transmitter source telescope was a 6 in. diam. F/4, gold coated mirrors with FOV ≥ 7.5 mrad; the receiver IR telescope was also a 6 in. diam. F/3 gold coated reflector with a FOV of 10 mrad. For passive mode measurements, the experimental setup was the same that in active mode, but the telescope coupled MIR source for excitation of the target was not used. Instead, a 500 W tungsten lamp was used to heat the aluminum plates with and without target analytes deposited on them. For both active mode and passive mode, multivariate calibrations were obtained and relevant statistical parameters were calculated and used as criteria to judge the quality of the method.

3. Application of multivariate calibrations to OP/FTIR data analysis

The mathematical modeling and detection of explosives and other threats in a complex environment can be complex. The large spectral bandwidth of a FTIR spectrometer facilitates the analysis. The purpose of calibration techniques is to correlate measured quantities such as the absorption of infrared radiation with properties of the system, for example, the concentration of one component in a multicomponent system. The accepted method of data analysis of gas phase contaminants present in a complex multicomponent mixtures as is the case of atmospheric pollutants present in air is classical least squares (CLS) regression analysis, also termed linear regression analysis or least squares (Russwurm and Childers, 1999). For quantification studies, CLS calibration curves can be generated using two methods: measurement of the absorbance peak heights and integration of areas in the spectral region of interest. Calibration plots using peak areas represents a better choice

for quantitative analysis when compared to peak height analysis (Lavine, 2002; Kramer, 1998). Usually, two steps are required: the calibration of the method and the analysis to determine a value of an unknown sample. It is important to emphasize that measuring surface concentrations using the peak area method is conceptually simple and easy to use, but it has some limitations. The method is univariate (the concentration is determined with a single spectral peak) and depends on a linear correlation between the concentration and the spectral response. The results can, therefore, be undermined by perturbations such as fluctuations caused by detector noise, temperature variations, or molecular interactions.

In a series of recent important papers by Griffiths and collaborators, the authors have demonstrated three important and novel aspects of data analysis for open path FTIR detection at a distance (Hart and Griffiths, 1998; Hart, et al., 200; Griffiths, et al. 2009; Shao, et al. 2010). Specific contributions can be summarized as follows:

- Establishment that multivariate data analysis techniques are required to exploit all the benefits that having a wealth of spectral information immersed in a congested multicomponent spectrum as that contained in SOIR experiments, not only for gas phase measurements but also for solid phase OP/FTIR detection at a distance, as Castro and collaborators have recently demonstrated (Castro, et al., 2010).
- Demonstration that a single background spectrum measured at a fixed source-target distance, temperature, pressure and ambient gases partial pressures can be used for all ranges and combinations of other relevant variables.
- Demonstration that the representation of OP/FTIR spectra in absorbance or percent transmission is equivalent. The output of any FTIR spectrometer is a single-beam spectrum that must be ratioed or compared (subtracted) against a appropriate background spectrum resulting in the transmittance spectrum of the sample, $T(\nu)$. In quantitative measurements, the transmittance should be converted to absorbance, $A(\nu)$, i.e. $-\log10\{(T(\nu)\}$ since the absorbance is a linear function of the concentration of the species absorbing, thus rendering it more amenable to use of chemometrics routines of analysis. For the current application: OP/FTIR detection of solid threat chemicals deposited on surfaces as traces, representation as the spectral difference between the sample spectrum and the background spectrum is even more critical due to the possibility of specular reflectance and scattering from the sample (Castro, et al., 2010).

Multivariate calibrations make use of not only a single spectral point but take into account spectral features over a wide range. Therefore, the analysis of overlapping spectral bands or broad peaks becomes feasible. The information contained in the spectra of the calibration samples will be compared to the information of the concentration values using a PLS regression. The method assumes that systematic variations observed in the spectra are a consequence of the concentration change of the components. However, the correlation between the components concentration and the change in the infrared signal does not have to be a linear one.

Calibrations are typically constructed using chemometrics methods of data analysis such as the partial least squares (PLS) regression algorithm. The PLS algorithm is commonly incorporated in spectroscopic software such as OPUS™, Pirouette™ and Matlab Toolboxes™, among others. The advantages of using chemometrics for the quantification of organic compounds on glass, aluminum and stainless steel and other surfaces have

been discussed in the literature (Mehta, et al., 2003; Hamilton, et al., 2005; Perston, et al., 2007; Primera-Pedrozo, et al., 2004; Primera-Pedrozo, et al., 2005; Primera-Pedrozo, et.al. 2007; Soto-Feliciano, et al., 2006; Primera-Pedrozo, et al., 2008; Primera-Pedrozo, O. M.; 2009).

PLS is a multivariate method that uses spectral features over a wide spectroscopic range. It is a spectral decomposition method that is intimately related to principal component analysis (PCA). In PCA, the spectral matrix is first decomposed into a set of eigenvectors and scores, and then a regression is performed against the concentrations as a separate step. However, PLS uses the concentration information during the decomposition process. In the case of OPUS™ (Bruker Optics), Quant2 software is used to find the best correlation function between the spectral information and the loading concentrations. Quant2 uses a partial least squares-1 (PLS-1) regression method. Calibrations are performed using PLS-1 in which only one component can be analyzed separately, instead of simultaneously analyzing multiple components, as in the PLS-2 routine of chemometrics. Then, cross validations are performed and the root mean square errors of the cross validations (RMSECV) and the root mean square errors of estimations (RMSEE) are used as criteria to evaluate the quality of the correlations obtained. In the standard ''leave-out-one'' cross validations, each spectrum is omitted from the training set and then tested against the model built with the remaining spectra. As illustrated in Tables 1 and 2, some explosives require spectroscopic preprocessing (except mean centering) and more PLS evaluations in order to obtain a good model. In the case of the first derivative, the Savitzky-Golay algorithm is actually used to obtain the derivative. The number of smoothing points used can be adjusted to suppress the effect of noise (Beebe, 1998). Other details of the advantages of using a chemometrics model such as PLS to correlate the loading concentration with IR spectra have been discussed in the literature (Hamilton, et al., 2005).

Multivariate calibrations require a large number of calibration samples and yield a large amount of data. In order to conveniently handle the data, the spectral information and the concentration information are written in the form of matrices, where each row in the spectral data matrix represents a sample spectrum. The concentration data matrix contains the corresponding concentration values of the samples. The matrices are then broken down into their eigenvectors which are called factors or principal components. Only the relevant principal components are used instead of the original spectral data, thus leading to a considerable reduction of the amount of data. A PLS regression algorithm will be developed to find the best correlation function between spectral and concentration data matrix (OPUS™, Bruker Optik, 2006). The OPUS/QUANT software package (OPUS™, Bruker Optik, 2006) is designed for quantitative analysis of spectra consisting of bands showing considerable overlap. The software allows determining the concentration of more than one component in each sample simultaneously. For this purpose, QUANT uses a **p**artial **l**east **s**quares (**PLS**) regression method.

The residual (*Res*) is the difference between the true and the fitted value. Thus the **s**um of **s**quared **e**rrors (**SSE**) is the quadratic summation of these values:

$$SSE = \sum \left[Res_i \right]^2 \tag{1}$$

The coefficient of determination (R^2) gives the percentage of variance present in the true component values, which is reproduced in the regression. R^2 approaches 100% as the fitted concentration values approach the true values:

$$R^2 = \left(1 - \frac{SSE}{\sum (y_i - y_m)^2}\right) \times 100 \quad (2)$$

In the case of a cross validation, the **r**oot **m**ean **s**quare **e**rror of **cross validation** (***RMSECV***) can be taken as a criterion to judge the quality of the method:

$$RMSECV = \sqrt{\frac{1}{M} \times \sum_{i=1}^{M} \left(y_i - y_{(i)}\right)^2} \quad (3)$$

where M is the number of standards in the data set. One method of cross-validation is **l**eave-**o**ne-**o**ut **c**ross-**v**alidation (**LOOCV**). Leave-one-out cross-validation is performed by estimating n calibration models, where each of the n calibration samples is left out one at a time in turn. The resulting calibration models are then used to estimate the sample left out, which acts as an independent validation sample and provides an independent prediction of each y_i value, $y_{(i)}$, where the notation i indicates that the *ith* sample was left out during model estimation. This process of leaving a sample out is repeated until all of the calibration samples have been left out. To obtain the **r**oot **m**ean **s**quare **e**rror of **p**rediction **(RMSEP)**, the validation samples prepared and measured independently from the calibration samples are used. The number of validation samples, p, should be large, so that the estimated prediction error accurately reflects all sources of variability in the calibration method. The **RMSEP** is computed as:

$$RMSEP = \sqrt{\frac{1}{p} \times \sum_{i=1}^{p} \left(y_i - y_{(i)}\right)^2} \quad (4)$$

where p is the number of prediction samples.

Partial least squares (PLS) regression algorithm from Quant2 software of OPUS™ version 6.0 (Bruker Optics) was used to find the best correlation function between the spectral information and the surface concentration. PLS was used for generating a chemometrics model for all analyzed standoff distances. Cross validations were made and the root mean square errors of cross validations (RMSECV) and correlation coefficient (R^2) were used as criteria to judge the quality of the correlations obtained at different standoff distances.

4. OP/FTIR detection of gases and condensable vapors

Griffiths et al. described two general ways in which OP/FTIR can be used for remote sensing measurements (Griffiths, et al., 2009). When the source and the detector are in line of sight with each other and they have separate power sources the operational mode is called *bistatic*. In this setup, the source is non-modulated by the interferometer and as a result stray light contributions cannot be minimized. On the other hand, when all the components reside within the spectrometer, including the MIR source, sharing a common power source, the operational mode is termed *monostatic*. This setup has clear advantages as pointed out by Pacheco-Londoño and collaborators in reducing stray light components, but is limited to source power and cooling restraints (Pacheco-Londoño, et al., 2009). In the active mode employed for remote detection, a bistatic setup in which the source is not modulated was used.

The first task in evaluating the performance of the Bruker Optics EM27 open path interferometer was to use it in detection of gases and condensable vapors. The spectroscopic system was configured for *bistatic* operational mode in active configuration (Fig. 3b) with source and interferometer in line-of-sight of each other. Measurements were done for standoff distance of 1-10 m. The spectra of ambient air and ammonia (NH_3) at 10 m range at room temperature are shown in Fig. 4a-d. In the case of NH_3, spectra were collected at an instrumental resolution of 1 cm^{-1}. The presence of ro-vibrational lines in the remote IR absorption spectrum of ammonia is clearly shown in Figs. 4b and 4c. The intense spectrum obtained for dichloromethane is illustrated in Fig. 4d. Gas phase standoff IR spectra of some high vapor pressure liquids are shown in Fig. 5a-d. Spectra are arranged in increasing order of their room temperature vapor pressure and absorbance of most prominent spectroscopic features (vibrational signals). The presence of spectral contributions from ambient water vapor and carbon dioxide ro-vibrational lines can be seen. These persistent lines were not removed by any of the widely used algorithms since spectral windows for sample identification were available in all cases.

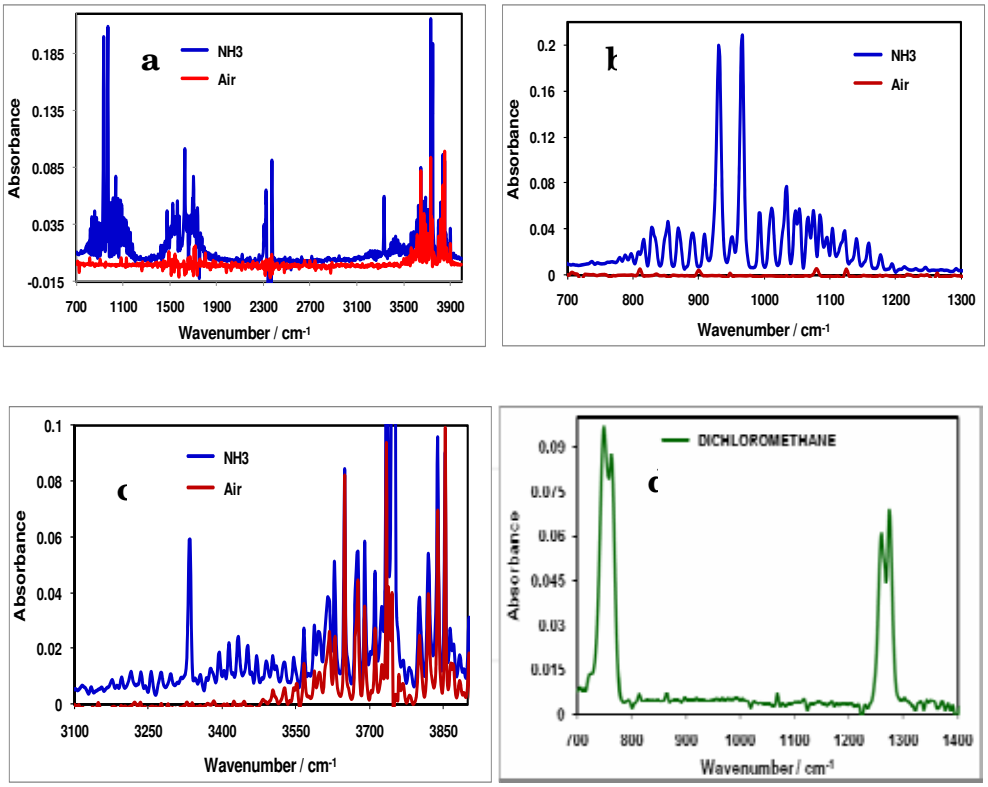


Fig. 4. Active mode OP/FTIR spectra of: (a) air and NH_3 complete spectrum; (b), (c) details of NH_3 spectrum; (d) dichloromethane

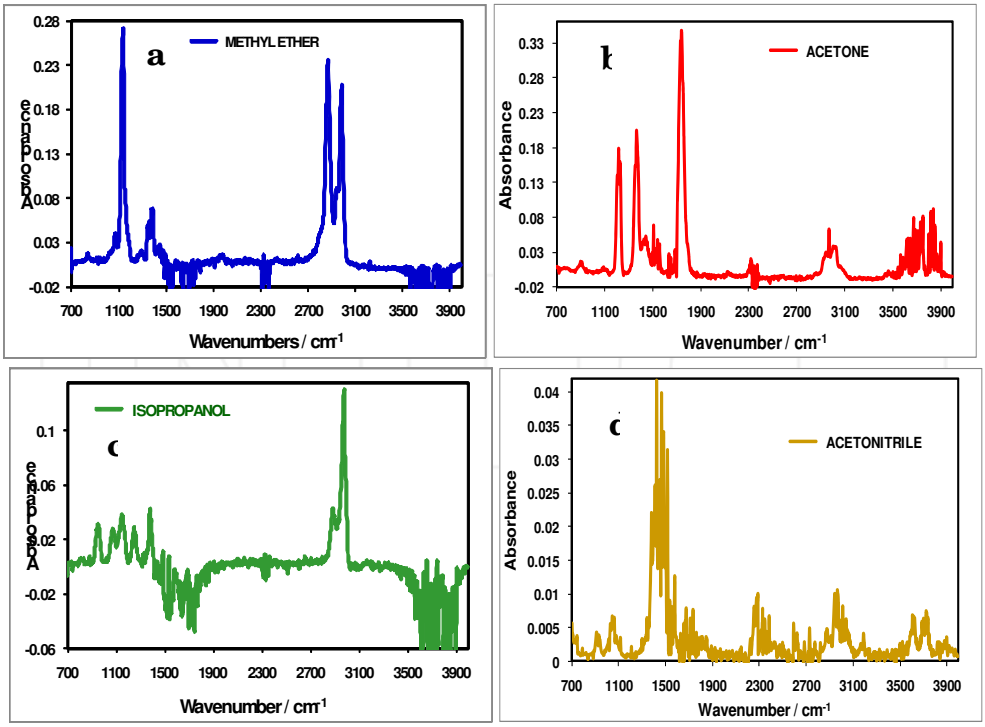


Fig. 5. Active mode remote spectra of condensable vapors: (a) methyl ether; (b) acetone, (c) isopropanol; acetonitrile

The spectral band shapes observed in stand-off detection mode, shown in Figures 4 and 5, are superimposed on a ramp-shaped background, and the bands themselves exhibit strong reflective (dispersive) band profiles. Since these measurements were done on a reflective metal substrate, the distortion of the band profiles is expected; similar effects have been reported in DRIFT (diffuse reflection infrared Fourier Transform) spectroscopy (Chalmers; Mackenzie, 1985). and in microscopically acquired infrared spectra of microspheres (Basan, el at. 2009a). In both cases, the distortion of the absorptive line shapes is due to the fact that within an absorption peak, the reflective index undergoes anomalous dispersion, as shown in Figure 6. In spectroscopic experiments carried out in reflectance mode, a mixing of the absorptive and dispersive line shapes can occur, resulting in bands that have a negative dip at the high wavenumber side of the peak, cf. Figures 4 and 5. This will shift the peak maximum by up to 15 cm^{-1} toward lower values (Basan, el at. 2009b).

For 'real-life' stand-off detection, strong reflectance band distortions, such as those shown in Figures 4 and 5 are not likely, since these are typical for reflectance spectra on metals, and should be much weaker if explosives are distributed on fabric. However, mixing of absorptive and reflective line shapes can also be mediated by scattering effects (Basan, et al. 2009a) and could produce significant band distortions.

Unsupervised correction of the spectral distortions will be necessary since the distortions cause apparent frequency shifts which will confound spectral search and identification algorithms. Although several methods have been developed to correct the dispersive line

shapes observed in biological systems (Basan, et al. 2009a; Bird, et al. 2010) they are not applicable here, since they are based on multivariate methods, and require large number of spectra, as well as undistorted reference spectra. For the spectra reported here, a method originally published in 2005 (Romeo and Diem, 2005), may be more suitable. This method is based on phase correction between real and imaginary spectral contributions which can be obtained by reverse Fourier transform of the contaminated spectra. The original implementation of this method contained a minor logical error, which since has been corrected (Bird, et al. 2010).

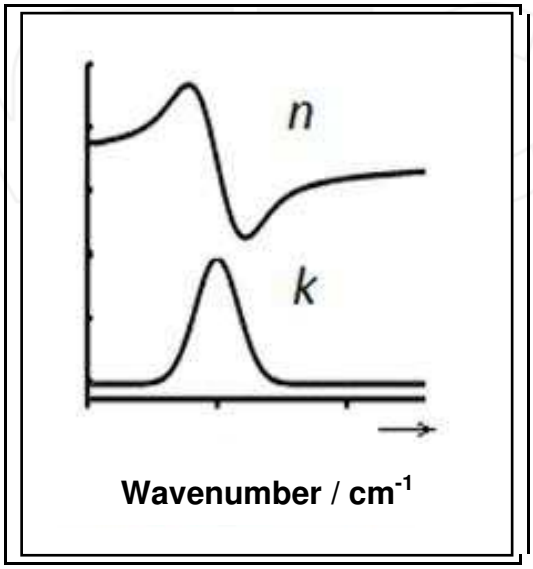


Fig. 6. Anomalous Dispersion of the refractive index (n) in the vicinity of an absorption band

5. Active standoff IR detection of solids deposited on substrates

Active mode standoff IR measurements of solids, including nitroexplosives and other highly energetic materials (HEM) smeared on Al plates at various surface loadings were carried out. Spectra in the fingerprint region were obtained with EM27 spectrometer. In the active mode used in the current application, a bistatic operation setup in which an IR telescope was used to steer a MIR source which was not modulated and sat side by side to the reflective IR telescope collector, operating in back-reflection mode (Fig. 7). Initial experiments were done with solid samples deposited onto aluminum plates at the highest surface loading used: 400 μg/cm^2. Samples were transferred to the metallic test substrates by dissolving in an appropriate solvent and then smearing them on the test plates. A Teflon stub was used to assist in sample smearing. Coated plates were allowed to dry in air at room temperature.

Initial remote IR experiments were designed to optimize experimental considerations, including sample placement and measurement geometry. For the reflectance IR measurements at one meter (1 m) standoff distance, as shown in Fig. 8, the angle of incidence of the MIR radiation (and angle of reflection) was varied from 0° to ~ 30°. TNT signals decreased drastically at high incident angles, becoming nearly unobservable at 27°. Spectra of selected

solid phase compounds deposited as traces on Al substrates are shown in Fig. 9. A fixed standoff distance of 8 m and loading concentration of 400 μg/cm² was used for the SOIR measurements. The first spectrum of Fig. 9 is that of caffeine deposited on Al plate. Fig. 9b and 9c show the spectra of p-benzoic acid and benzoic acid, respectively at 25°C. For the acquisition of the data shown in Fig. 9d the sample plate was heated to 28°C to demonstrate how the emission of vibrational quanta is significantly enhanced by small temperature differences. The remote IR spectrum of aliphatic nitrate ester PETN is shown in Fig. 9e and the corresponding spectra aliphatic nitramine RDX are shown as %T and absorbance in Fig. 9f.
Pacheco and co-workers used a modulated home built setup for remote IR measurements of nitroexplosives from ~ 1 m to ~ 4 m range. At short distances (0.9 m and 1.8 m) the maximum and minimum signal to noise (SNR) values showed high dispersion. They found out that with their uncollimated MIR beam, the problem both was sample and transfer solvent dependent in the low to very low loading concentrations studied. They argued that part of the problem was the lack of uniformity in surface coverage due to nucleation and crystallization phenomena. When the IR beam used was 1-2 cm in diameter or less, sample discontinuities could be detected and this was reflected by the relatively high dispersion in the values of SNR. At longer target-collector distances (1.8, 2.7 and 3.7 m) the maximum and minimum SNR values were very close due to higher sample coverage by a beam spot of ~ 5 - 11 cm in diameter. At a source-target distance of 0.9 m, limit of detection (LOD) value determined was 2 μg/cm² for TNT. According to the authors, LOD values determined could have been influenced by several factors, such as: humidity, alignment of detector, source pointing accuracy, detector efficiency and reflectivity of samples and substrates.

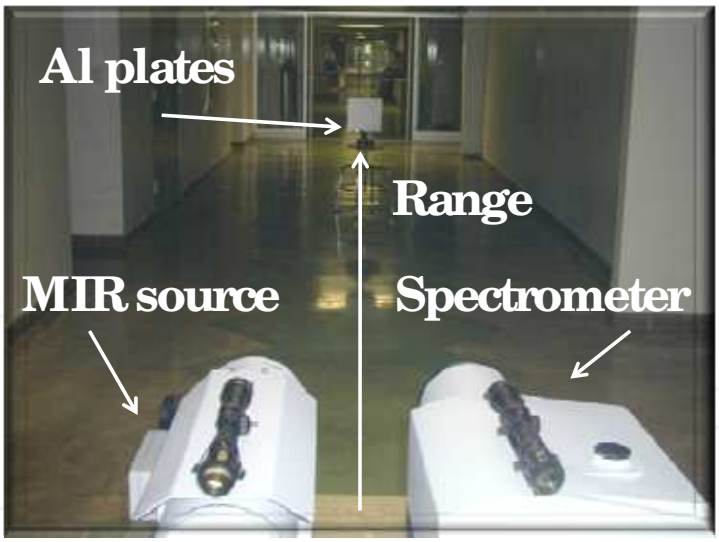


Fig. 7. Experimental setup used for SOIR detection. In active mode operation, both MIR source and FT interferometer were coupled to gold coated mirrors MIR reflective telescopes. In passive mode only the IR spectrometer was used

The reflectivity of sample: substrate and analyte, is mainly determined by how the analyte is deposited and by the solvent used for deposition. If the explosive exists on the surface as a thin layer, the backscattering signal is low but the reflection-absorption infrared (RAIRS)

spectra measured is of high signal-to-noise (SNR) value. When the explosive was present on the surface as discrete particles (crystals), the backscattered SIRS signal improved; correspondingly, the RAIRS signal measured for surface loading validation got worse. Sample smearing was done twice using the Teflon applicator: from left to right and then right to left. When the first pass was done the solvent was allowed to evaporate. Then, the second pass was carried out to induce more sample roughness and particle formation (crystallization) on the surface. TNT deposits on the substrate did not result in a thin layer covering the metallic surface. Instead droplets of a metastable phase were formed on the surface (Manrique-Bastidas, et al., 2004; Vrcelj, et al., 2001; Manrique-Bastidas, et al. (2) 2004). This metastable phase could be easily turned to its crystalline phase by friction, abrasion or even by pressing hard with the Teflon applicator for the sample smearing stage. This effect enhanced the SOIR experiments of TNT because the metastable phase was formed in the first smearing step. When the methanol evaporated and the second smearing stage was performed, crystalline roughness was induced resulting in a lower LOD value for TNT than for RDX. Standoff IR spectra of 400 μg/cm² TNT at 1 m and 14.5 m are shown in Fig. 10. A reference spectrum of neat, microcrystalline sample of TNT (1 mg/100 mg KBr) obtained in the macro sample chamber of a benchtop interferometer (Bruker Optics IFS-66/v) is included for comparison purposes.

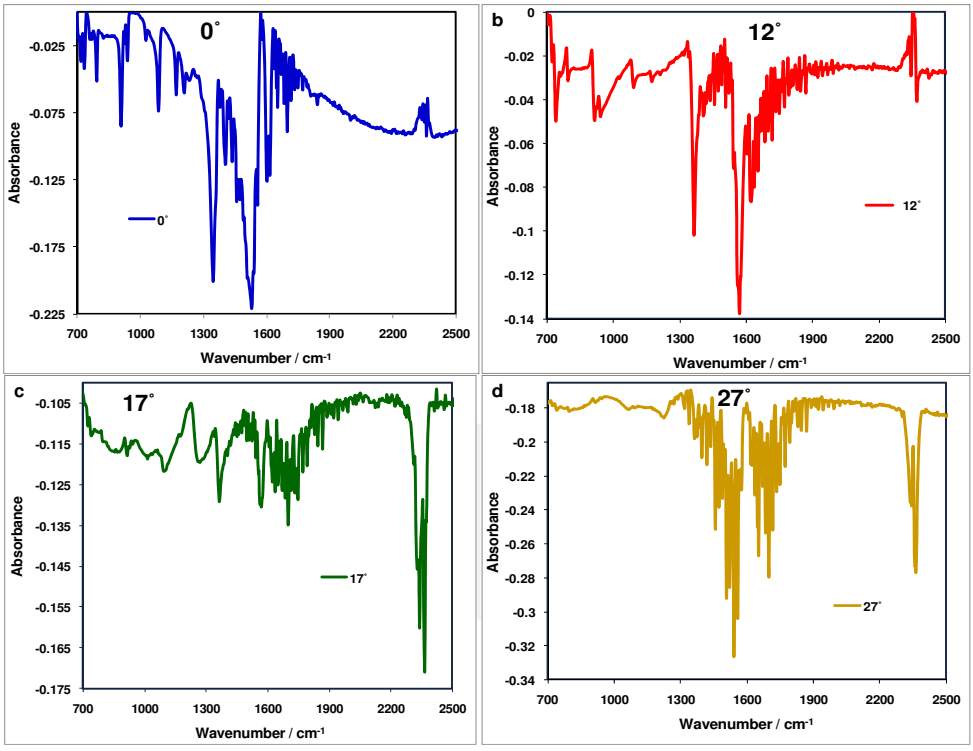


Fig. 8. Active mode OP/FTIR spectra of 400 μg/cm² TNT deposited on Al plate at: (a) 0°; (b) 12°; (c) 17°; (d) 27°. Data shown demonstrates the specular reflectance nature of the IR reflection-absorption (transflection) experiment

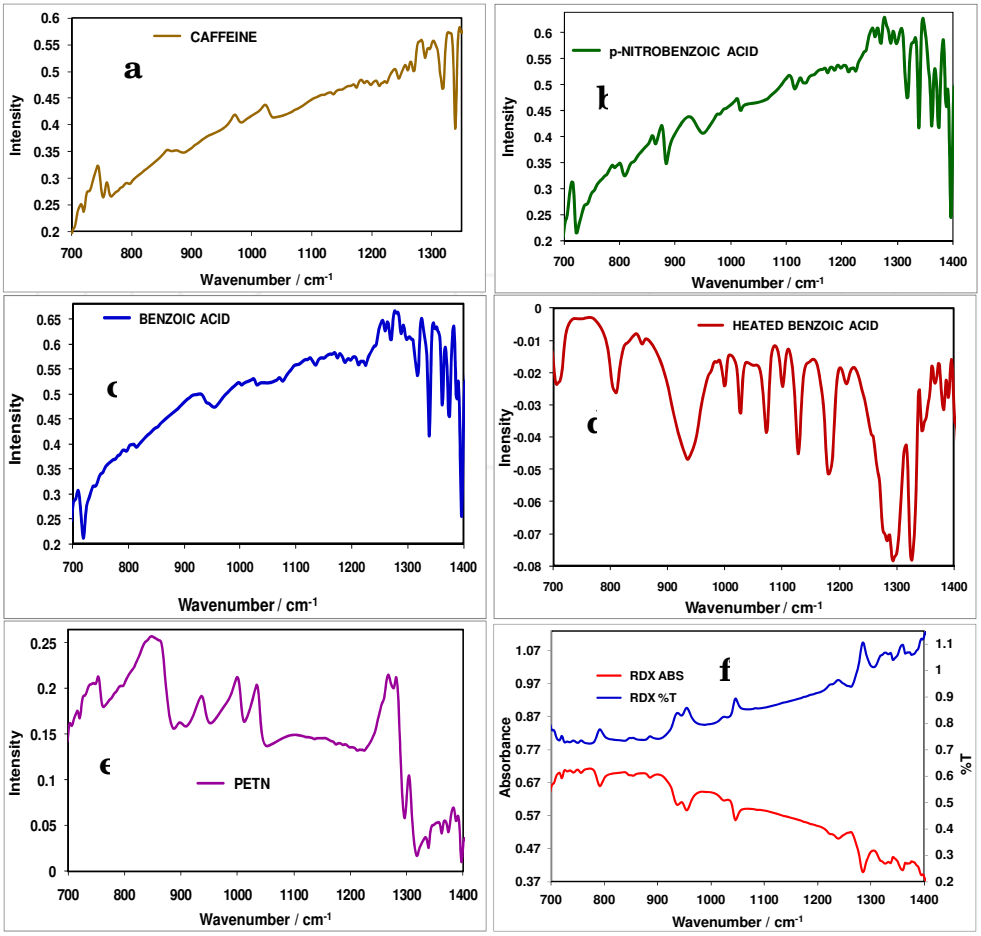


Fig. 9. Remote IR spectra of: (a) caffeine; (b) p-nitrobenzoic acid, (c) benzoic acid at room temperature (25 °C); (d) benzoic acid heated to 28 °C; (e) aliphatic nitrate ester PETN; (f) nitramine RDX in absorption and in %T

Other remote IR detection experiments were done using only TNT as target. Loading concentrations ranging from 50 to 400 μg/cm² of TNT were deposited on Al plates. The targets were carefully aligned to the source and collector and then the SOIR spectra were recorded. The analyzed target-collector distances were 4, 8, 12, 16, 20, 25, and 30 m. A total of 10 spectra were taken for each sample, at 20 scans and 4 cm-1 resolution. Experiments were carried out at room temperature (25°C). Spectra were collected in remote, bistatic active mode detection IR at various surface concentrations at a fixed standoff distance of 8 m. Typical results are shown in Fig.11. These traces were not submitted to any pre-processing routine: offset correction, baseline correction, smoothing, water vapor rotational lines removal, etc. Thus, there is no common baseline for these spectra and some traces exhibit positive intensity ramps to higher wavenumber. However, increase of signal intensity as function of loading surface concentrations is clearly shown without the use of

chemometrics routines. Intense vibrational band about 908 cm^{-1} was tentatively assigned to C-N stretching, vibrational band at 938 cm^{-1} was assigned to C–H out-of plane bend (ring) and symmetric stretch band of the nitro groups appears at 1350 cm^{-1}. Results agree with reported values (Pacheco-Londoño, et al., 2009; Clarkson, et al., 2003).

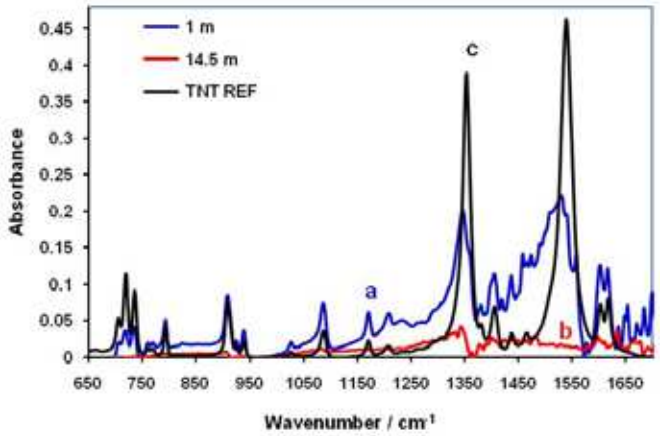


Fig. 10. Comparison of IR spectra of TNT: (a) KBr pellet with 1 mg TNT obtained in a macro sample compartment of a FTIR; (b) active mode SOIR spectrum of 400 $\mu g/cm^2$ deposited on Al plate measured at 1 m range distance; (c) same sample before measured at a source-target distance of 14.5 m. Prominent spectral features are present

When a MIR source was used for carrying out active mode experiments, the intensity of the peaks decreased when the distances increase. This is illustrated in Fig. 11a for spectra of Al plate coated with surface loading of 400 $\mu g/cm^2$ TNT and measured at standoff distances of 4, 8, 12, 16, 20, 25 and 30 m. At standoff distances higher than 25 m it was not possible to visualize clearly some of TNT vibrational signatures. At these distance the density of infrared radiation that gets to the Al plates from the MIR source is low, leading to a smaller number of excited molecules, so that the detector cannot register the low intensity signals emitted. Fig. 11b shows SOIR spectra of TNT as function of loading concentration. Spectra were collected in active mode detection standoff IR at a fixed standoff distance of 8 m.

The statistical treatments with chemometrics using PLS were carried out using the spectral region 700 to 1400 cm^{-1}, where the nitro symmetric stretch and aromatic C-H vibrations are present. Data pre-processing is an important stage in performing a calibration. Thus, the PLS models were built using mean centering as only pre-processing of variables. To ensure the reproducibility of the calibration samples, several spectra of each sample (fixed loading concentration and standoff distance) had to be acquired. If spectra of the same sample are not identical, a data pre-processing procedure must be chosen to bring them in line with each other. Data pre-processing can eliminate variations in offset or different linear baselines. Different treatments data were used, including: **vector normalization, first derivative and second derivative, mean-centering, but no other pre-processing routine was applied,** achieving best results for RMSECV and R^2. Results indicate that the experimental setup has good management of external variables, such as humidity, temperature changes, homogeneity of samples on Al plates, and others at ranges as far as 30 m and surface loadings of 50 $\mu g/cm^2$.

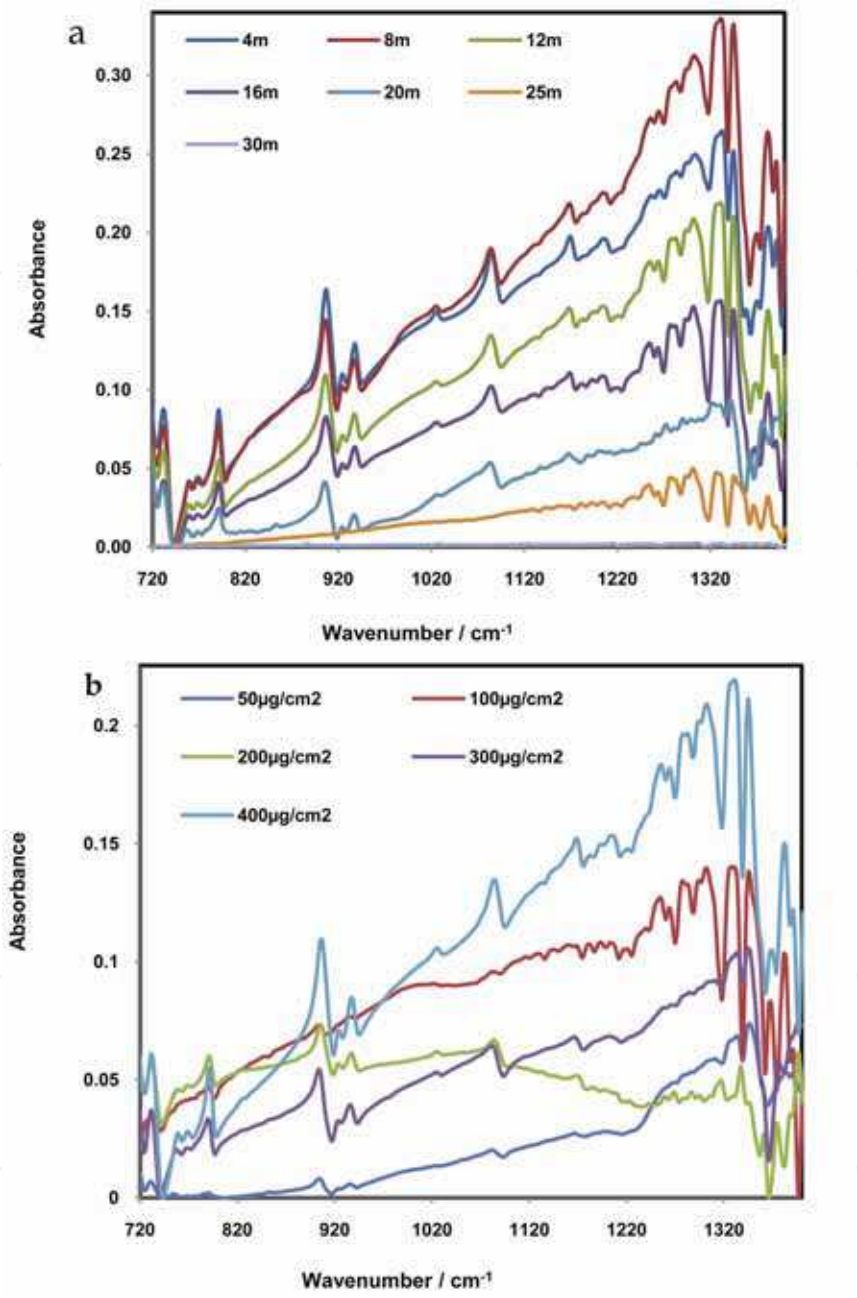


Fig. 11. Active mode FTIR spectra of TNT: (a) at different standoff distances of Al plate coated with surface concentration of 400 µg/cm^2; (b) at several loading concentrations at fixed range of 8 m

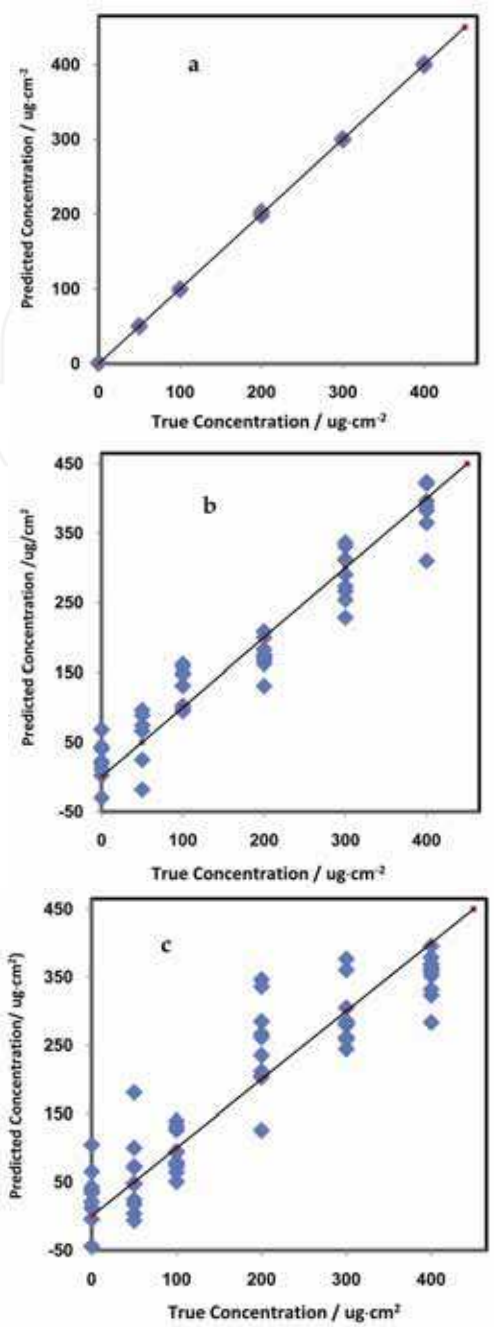


Fig. 12. Predicted loadings vs. true loadings for TNT on Al plates at various ranges using Opus 6.0™ Quant2: (a) 20 m; (b) 25 m; (c) 30 m

Fig. 12 illustrates the results obtained of the cross validation at the analyzed standoff distances: 20, 25 and 30 m and Table 2 summarizes the results of RMSECV and R^2 obtained in the PLS models. In these correlation charts, each point represents ten spectra with a fixed surface concentration (0, 50, 100, 200, 300, or 400 μg/cm²). All PLS correlation charts of predicted surface loading value vs. true surface concentration value for 4, 8, 12 and 16 m were similar to the correlation chart presented for 20 m standoff distance (Fig. 11a). As the standoff distance increases (> 20 m) some of the spectral information is lost, causing the spectra for each sample to be slightly different from the others (within experimental error) thus making it difficult to predict its concentration (see Figs. 11b and 11c). Taking into account the low values of RMSECV and high values of R^2 obtained at maximum distance of 20 m, the correlation (PLS) models are useful tools to determine precisely the surface concentration of TNT unknown samples using OP/FTIR spectroscopy. More precise alignment of both transmitter and receiver MIR telescopes would be required to perform similar correlations for ranges > 20 m.

6. Passive mode standoff IR detection

The setup used for passive mode detection using thermal excitation of the sample is shown in Fig. 13. The MIR source and transmitter telescope were not used in these experiments. Temperature differences tested were 1 to 7°C in one degree interval. Aluminum plates (Fig. 13a) were heated by a 500 W tungsten lamp that was placed on back of the Al plates (Fig. 13b). The standoff distances studied were 8, 16 and 30 m. In passive mode experiments it is not so critical to carefully align the target and detector while recording the spectra. Ten spectra were taken for each sample, at 10 scans and 4 cm^{-1} resolution for passive mode experiment.

The emission from a heated, uncoated with TNT Al plate, used as blank to measure background contributions is shown in Fig.14. The corresponding blackbody spectrum of 400 μg/cm² TNT is also shown overlapping the blank Al plate spectrum. Both traces were measured at a range of 8 m and the plates were maintained at an equilibrium temperature of 32°C by heating with a 500 W tungsten lamp.

Fig. 15 shows the TNT IR vibrational signatures recorded with an EM27 spectrometer using passive mode standoff IR detection of Al plates heated with tungsten lamp to different surface temperatures from 25 to 32°C. These spectra were measured at a standoff distance of 16 m and a TNT surface concentration of 400 μg/cm². Most of the characteristic vibrational signatures of TNT are well defined. The bands that allow identifying TNT were taken in the spectral region 700-1400 cm^{-1}. Most of the persistent bands are observed and the standoff spectrum agrees very well with traditional infrared techniques: sample compartment, KBr pellet, micro-IR, attenuated total reflectance (ATR, both micro and macro) and grazing angle reflectance (GA, both micro and macro). Bands tentative assignments are: 910 cm^{-1} (2,6-NO_2 scissors and C–N stretch); 1087 cm^{-1} (C–H (ring) in-plane bend); 1171 cm^{-1} (C–C in-plane ring trigonal bend, 2,4,6-C–N and C–CH_3 stretch) and 1350 cm^{-1} due to the symmetric stretching vibration of the NO_2 (nitro) group bond. TNT vibrational markers change significantly with surface temperatures. At 25°C (black spectrum, room temperature) TNT vibrational signatures are present, but they can barely be noticed at 16 m standoff distance. For the spectrum at 26°C (green spectrum) there is a significant increment in intensity of TNT vibrational signals.

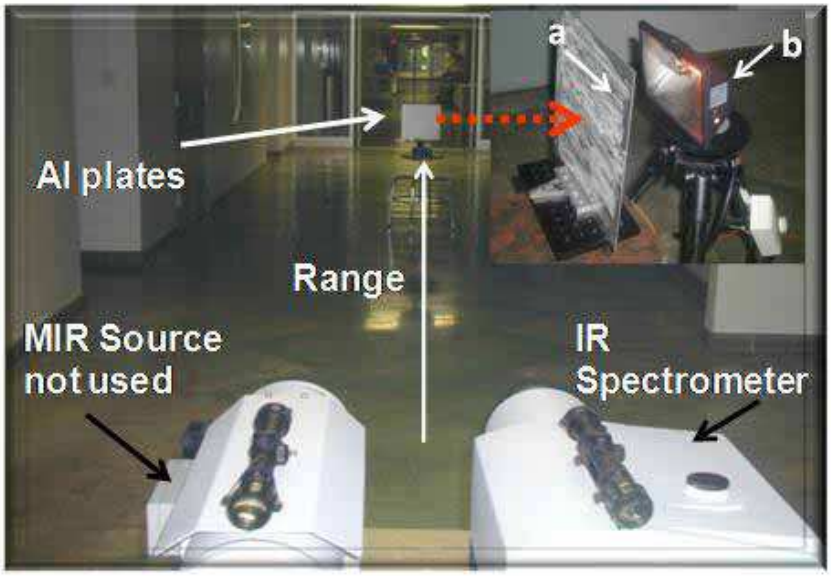


Fig. 13. Passive mode remote IR setup: (a) Al plate with TNT sample smeared-on; (b) 500 W tungsten lamp assembly

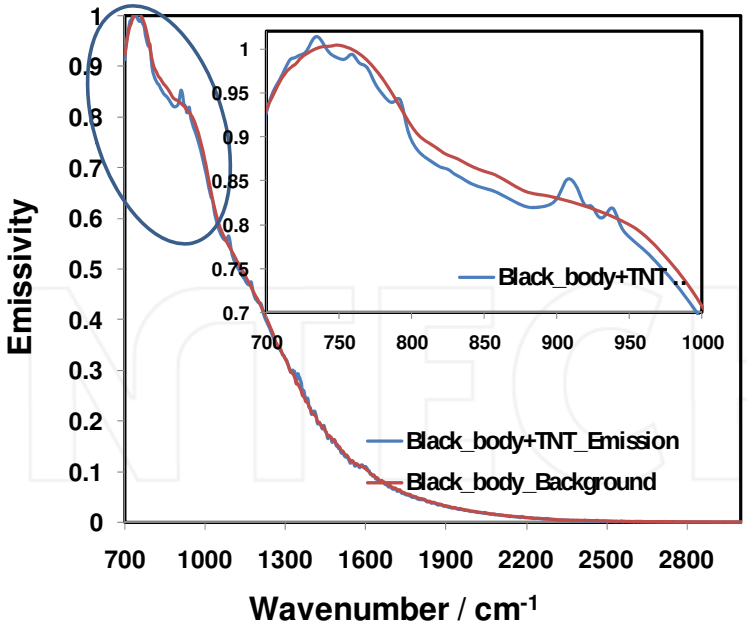


Fig. 14. Overlap of emissivity spectra of background air and air with TNT emissions from heated Al plate with 400 μg/cm² TNT at 8 m range and plate heated to 32°C

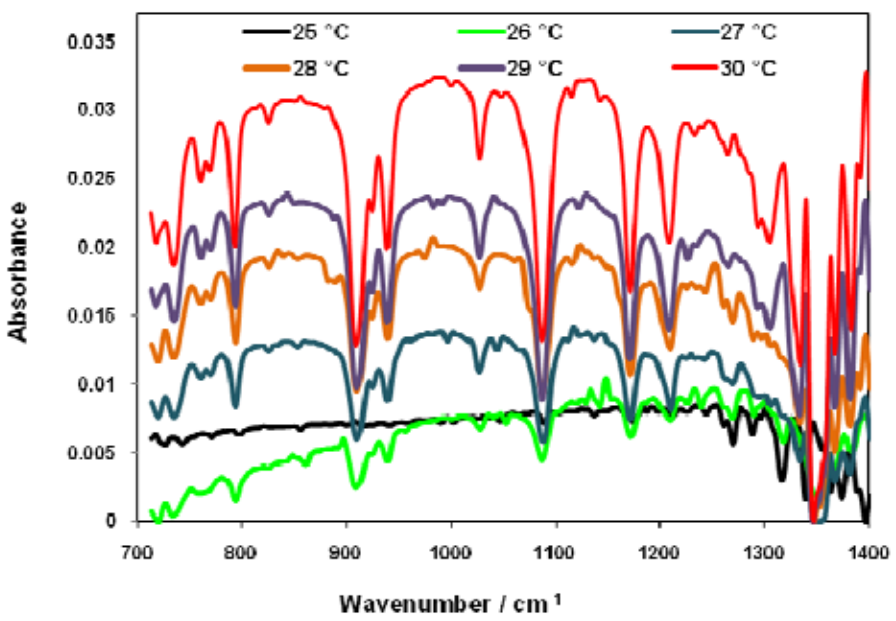


Fig. 15. Passive mode remote IR spectra of 400µg/cm² TNT at several temperatures

The variation in peak areas with temperature for TNT for a surface loading of 200 µg/cm² at two different standoff distances is shown in Fig. 15. The vibrational spectra were measured out to a room temperature (25ºC). For a standoff distance of 8 m (Fig. 11a) peak areas of the vibrational bands at 793 and 1087 cm^{-1} were measured. For a standoff distance of 16 m the bands used for peak areas calculations were 793, 1087 and 1171 cm^{-1} (Fig. 11b). In both cases when TNT was heated to higher temperatures more intense bands were observed. This study shows than the increase of the vibrational signatures has a second order polynomial behavior in all cases, with excellent correlation coefficients squared. These results are very useful for real field standoff detection, because when the target is warmer than room temperature, the vibrational signatures of the explosive are increased significantly.

The results of the effect of standoff distance on the intensity of heated samples are shown in Fig. 16. The spectra were taken at different distances and a specific surface temperature for each one. The tested temperature was always higher that the ambient temperature. Spectra taken at room temperature did not show some of the characteristic TNT vibrational signals (spectra not shown). Fig. 16 shows that the standoff detection in passive mode using thermal excitation is a useful tool for recording IR spectra to maximum range distance of 30 m, under the current experimental conditions.

PLS regression algorithm from Quant2 software for OPUS™, version 6.0 (Bruker Optics, Billerica, MA) was used to find the best correlation function between the IR spectral information and the TNT surface concentration. PLS was used for generating a chemometrics model of analyzed standoff distances at specific temperatures (25 to 32°C). Cross validations were made and RMSECV and R^2 were used as criteria to evaluate the quality of the correlations obtained. The statistical data treatments prepared using chemometrics routines (PLS) were carried out using spectral region 700 to 1400 cm^{-1}, where

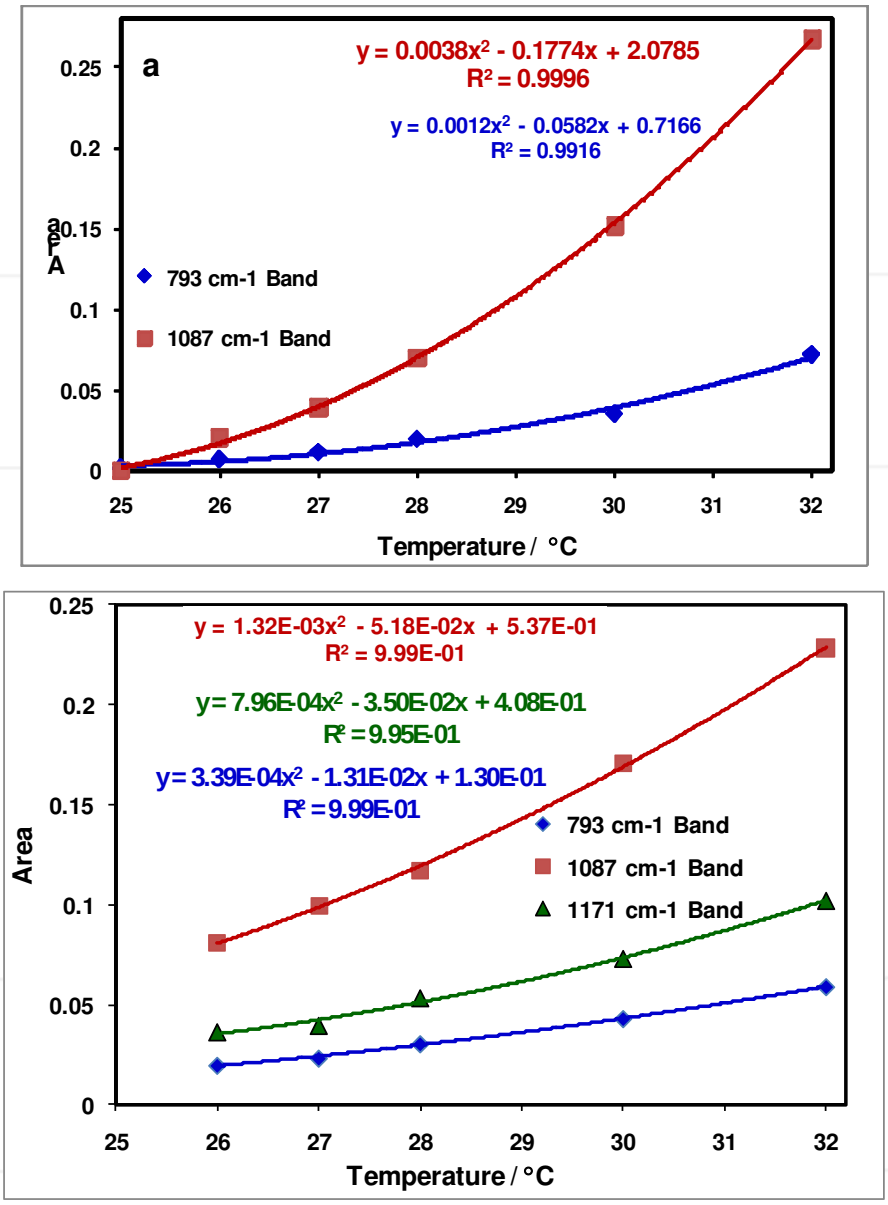


Fig. 16. Effect of temperature on intensity of TNT vibrational signals for 200 μg/cm² at range of: (a) 8 m; (b) 16 m

the nitro symmetric stretch and aromatic C-H and C-C vibrations are present. As before, PLS models were made using mean centering as pre-processing of variables. Data pre-processing included: "vector normalization, first derivative and second derivative but NO pre-processing" achieving best results for RMSECV and R^2. Fig. 14 shows the results obtained

for the cross validations at analyzed standoff distances using different surface temperatures. Table 3 contains the summary of results for RMSECV and R^2 obtained in the PLS models.

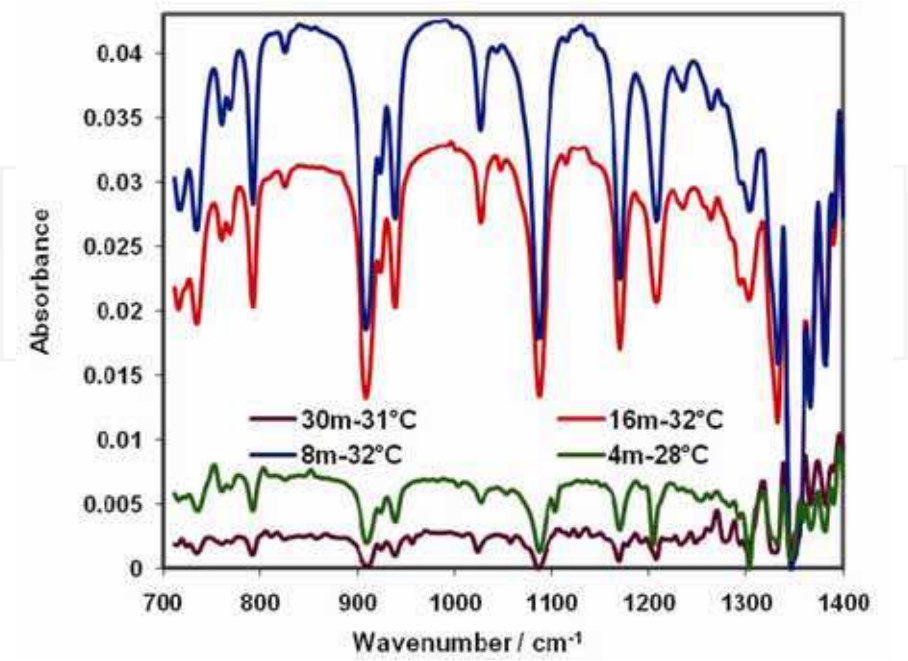


Fig. 17. Effect of the standoff distance on intensity of TNT IR signals. Surface concentration of 400 µg/cm² and 8-16 m range at sample temperatures: 28-32°C

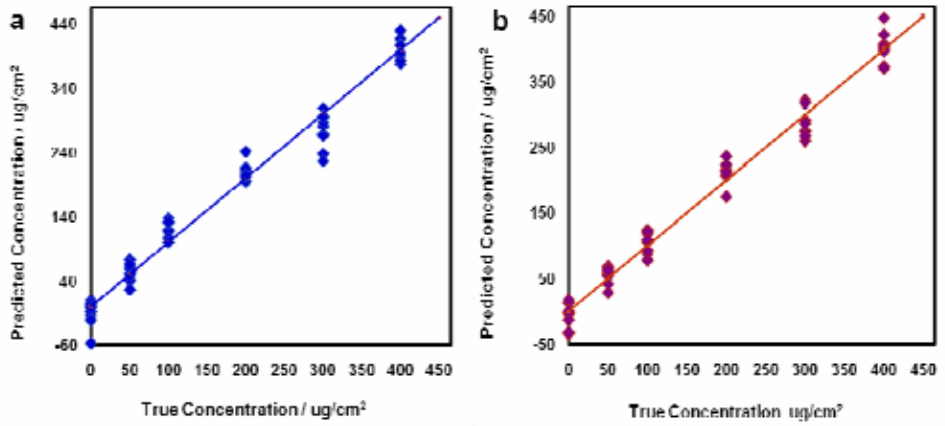


Fig. 18. Predicted vs. true surface concentration for TNT explosives at different standoff distances and temperature in passive mode: (a) 8 m range, 32 °C surface temperature; (b) 16 m range, 32 °C surface temperature

Standoff Distance (m)	Temperature (°C)	R^2	RMSECV	RMSEP	Rank
8	25	0.9461	32.6	28.1	6
8	26	0.9737	22.8	29.2	9
8	27	0.9864	16.4	22.8	3
8	28	0.9692	24.7	24.3	2
8	30	0.9658	26.0	28.1	4
8	32	0.9760	21.8	17.7	6
16	25	0.9555	29.7	24.0	3
16	26	0.9624	27.3	29.0	3
16	27	0.9202	39.7	41.4	3
16	28	0.9567	29.3	27.6	4
16	30	0.9506	31.3	30.5	3
16	32	0.9684	25.0	26.6	3

Table 3. PLS calibration parameters for the different tested standoff distances and temperatures. Spectral range: 700 – 1400 cm^{-1}; no preprocessing

All graphs of predicted vs. true surface concentration for specific distance and temperature (Table 3) have similar behavior to that of Fig. 18a for standoff distance of 8 m and Fig. 18b at a range of 16 m. Each point represents ten spectra with a specific surface concentration (0-400 $\mu g/cm^2$). Taking into account the high values of R^2 (~ 0.96) and relatively low values of RMSECV (around 26.0) obtained for all the tested distances and temperatures these models could be used as a tools to determine the surface concentration of unknown samples of TNT at remote distances using SOIR.

7. Conclusion

A standoff technique using an Open-Path Fourier transform infrared (OP/FTIR) spectrometer has been demonstrated to obtain spectral information of TNT samples deposited on Al plates. The system consisted in an infrared telescope coupled MIR source and an IR coupled EM27 spectrometer manufactured by Bruker Optics. The remote

detection IR system was first tested for the standoff detection of gases and condensable vapors, which is the application that was developed for. High quality measurements were achieved by using a sensitive photoconductive cryo-cooled MCT detector. Standoff detection both in active and passive modes proved to be useful for recording TNT vibrational signatures in the range from 700 to 1400 cm^{-1} of the MIR. Very good results of RMSECV and R^2 were obtained in cross validations for active and passive mode experiments. The active mode standoff detection in worked very well for distances lower than 30 m. Is necessary carefully aligning the target with the detector to be able to measure with high accuracy at ranges higher than 30 m.

For passive mode experiments thermal excitation proved to be a useful tool for enhancing TNT vibrational signatures for standoff detection. Achieving temperature difference of just 1°C between the sample and the spectrometer was enough to bring out spectral information to standoff distances of 8, 16 and 30 m. The increase in intensity of TNT signatures as a function of temperature can be modeled very well with second order polynomials in the temperature range tested above room temperature. In this experiment alignment of sample and detector was not critical as in the standoff active modality configuration. Partial least squares routines of commercial chemometrics statistical routines of spectroscopic analysis were used to enhance the data in multivariate mode.

8. Acknowledgments

Parts of the work presented in this contribution were supported by the U.S. Department of Defense, University Research Initiative Multidisciplinary University Research Initiative (URI)-MURI Program, under grant number **DAAD19-02-1-0257**. The authors also acknowledge contributions from Mr. Aaron LaPointe from Night Vision and Electronic Sensors Directorate, Fort Belvoir, VA, Department of Defense, Dr. Jennifer Becker MURI Program Manager, Army Research Office, DoD and Dr. Stephen J. Lee Chief Scientist, Science and Technology, Office of the Director, Army Research Office/Army Research Laboratory, DoD.

Support from the U.S. Department of Homeland Security under Award Number **2008-ST-061-ED0001** is also acknowledged. However, the views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

9. References

ALERT Department of Homeland Security university based Center of Excellence: at Northeastern University: http://www.northeastern.edu/alert/index.php/

ALERT Department of Homeland Security university based Center of Excellence: at University of Rhode Island: http://energetics.chm.uri.edu/

Akhavan, J. (2004). *The Chemistry of Explosives*. 2nd ed. T J International Ltd, Padstow, Cornwall, UK.

American Society for Testing and Materials (1997) *Standard guide for open-path Fourier-transform infrared (OP/FTIR) monitoring of gases and vapors in air*, E-1865–97. In:

Annual book of ASTM standards, American Society for Testing and Materials, West Conshohocken, PA, vol 03.06.

Bacci, M.; Fabbri, M.; Picollo, M.; Porcinai, S. (2001). Non invasive fibre optic Fourier Transform-infrared reflectance spectroscopy on painted layers identification of materials by means of principal component analysis and Mahalanobis distance. *Anal. Chim. Acta*. 446, 15-21.

Bangalore, A.S.; Small, G.W.; Combs, R.J.; Knapp, R.B.; Kroutil, R.T.; Traynor, C.A.; Ko, J.D. (1997) *Anal. Chem. 69*, 2, 118.

Blake, T.A.; Kelly, J.F.; Gallagher, N.B.; Gassman, P.L.; Johnson, T.J. (2009) *Anal. Bioanal. Chem.* 395, 337.

Beebe, K.; Pell, R.; Beth, M. (1998) *Chemometrics: A Practical Guide*. John Wiley & Sons. N.Y

Caron, T.; Guillemot, M.; Montméat, P.; Veignal, F.; Perraut, F.; Prené P.; Serein-Spirau, F. (2010) *Talanta* 81, 1,543.

Castro-Suarez, J.R.; Pacheco-Londoño, L.C.; Vélez, M.; Diem, M.; Tague, Jr. T.J.; Hernandez-Rivera, S.P. (2010) submitted, *Appl. Spectrosc.*

Clarkson, J.; Smith, W.E.; Batchelder, D.N.; Smith, D.A. Coats, A.M. (2003) *J Molec. Structure*, 648, 203.

Committee on the Review of Existing and Potential Standoff Explosives Detection Techniques, Existing and Potential Standoff Explosives Detection Techniques (2004) National Academies Press: Washington, DC.

Diem, M. (1993). *Introduction to Modern Vibrational Spectroscopy*, John Wiley, New York, NY.

Griffiths, P. R.; de Haseth, J. A. (2007) *Fourier transform infrared spectrometry*, 2nd ed. Wiley Interscience, Hoboken, NJ, pp 466–479.

Griffiths, P.R.; Shao, L.; Leytem, A.B. (2009) *Anal. Bioanal. Chem.* 393, 45.

Gunzler, H.; Gremlich, H.-U. (2002) *IR Spectroscopy: An Introduction*. Wiley-VCH, DE.

Hamilton, M. L.; Perston, B. B.; Harland, P. W.; Williamson, B. E.; Thomson, M. A.; Melling, P. J. *Organic Process Research & Development*. 2005, 9, 337–343.

Hart, B.K., Griffiths, P.R. (1998) Proc 11th Int. Conf Fourier Transform Spectrosc. *Am. Inst. Phys. Conf. Proc*. 430:241–242.

Hart, B. K.; Berry, R. J.; Griffiths, P. R. (2000) *Environ. Sci. Technol.* 34:1346–1351.

Hilmi, A.; Luong, J. (2000) *Environ. Sci. Technol.*, 34, 14, 3046.

Huberty, C.J. (1994) *Applied Discriminant Analysis*. Wiley–Interscience, Hoboken, NJ.

Johnson, R.A.; Wichern, D.W. (1992) *Applied Multivariate Statistical Analysis*. Prentice-Hall: Englewood Cliffs, NJ.

Kramer, R., (1998). *Chemometrics Techniques for Quantitative Analysis*, Marcel Dekker, New York, NY.

Lin-Vien, D.; Colthup, N.B.; Fateley, W.G.; Grasselli, J.G. (1991) *The Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules*. San Diego, CA pp. 179–189.

Manrique-Bastidas, C.A.; Castillo-Chará, J.; Mina, N.; Castro, M.E.; Hernández-Rivera, S.P. (2004a), *Proc. SPIE Int. Soc. Opt. Eng*. Eng. 5415: 1345.

Manrique-Bastidas CA, Primera-Pedrozo OM, Pacheco-Londoño LC, Hernández-Rivera, S.P. (2004b) *Proc. SPIE Int. Soc. Opt. Eng*. 5617: 429.

Mardia, K.V.; Kent, J.T.; Bibbly, J.M. (1979) *Multivariate Analysis*. Academic Press: New York, N.Y.

Marshall, M.; Oxley, J.C. (2009) *Aspects of Explosives Detection*. Elsevier, Amsterdam, The Netherlands.

Mehta, N. K.; Goenaga-Polo, J. E.; Hernández-Rivera, S. P.; Hernández, D.; Thomson, M. A; Melling, P. J. *Spectroscopy*, April, 2003.

Miller, C. J.; Yoder, T.S. (2010) *Sens. Imaging: An International Journal*. 11, 2, 77.

OPUS version 4.2 (2003) *User Manual*. Bruker Optics: Billerica, MA.

OPUS version 6.0 (2006) *Spectroscopic software manual* Bruker Optik GmbH, DE.

Pacheco-Londoño, L.C.; Ortiz, W.; Primera, O.M.; Hernández-Rivera, S.P. (2009) *Anal. Bioanal. Chem.* 395, 323.

Parmeter, J. E. (2004) Proceedings of the 38th Annual 2004 International Carnahan Conference on Security Technology; 355 IEEE: New York.

Perston, B.B.; Hamilton, M.L.; Williamson, B.E.; Harland, P.W.; Thomson, M.A.; Melling, P.J. (2007) *Anal. Chem.*, 79, 1231–1236.

Primera-Pedrozo, O.M.; Pacheco Londono, L.C.; De la Torre-Quintana, L.F.; Hernandez-Rivera, S.P.; Chamberlain, R.T.; Lareau, R.T. (2004) *Proc. SPIE Int. Soc. Opt. Eng.* 5403, 237–245.

Primera-Pedrozo, O.M.; Pacheco-Londono, L.C.; Ruiz, O.; Ramírez, M.; Soto-Feliciano, Y.M.; De La Torre-Quintana, L.F.; Hernandez-Rivera, S.P. (2005) *Proc. SPIE Int. Soc. Opt. Eng.* 5778, 543-552.

Primera-Pedrozo, O. M.; Rodríguez, N.; Pacheco-Londoño, L. C.; Hernández-Rivera, S. P. (2007) *Proc. SPIE Int. Soc. Opt. Eng.*, 6542, 65423J.

Primera-Pedrozo, O.M.; Soto-Feliciano, Y.; Pacheco-Londoño, L.C.; Hernandez-Rivera, S.P. (2008) *Sens. Imaging: An International Journal*. 9, 27–40.

Primera-Pedrozo, O.M.; Soto-Feliciano, Y.; Pacheco-Londoño, L.C.; Hernandez-Rivera, S.P. (2009) *Sens. Imaging: An International Journal*. 10, 1–13.

Soto-Feliciano, Y.; Primera-Pedrozo, O.M.; Pacheco-Londoño, L. C.; Hernandez-Rivera, S.P. (2006) *Proc. SPIE Int. Soc. Opt. Eng.* 6201, 62012H.

Russwurm, G. M.; Childers, J.W (1999) *FTIR open-path monitoring guidance document*, 3rd ed., TR-4423–99–03. ManTech Environmental Technology. Research Triangle Park, NC, USA.

Russwurm, G.M.; Childers, J.W. (2001) *Open-path Fourier transform infrared spectroscopy*. In: Chalmers, J.W.; Griffiths, P. R. (eds.) Handbook of vibrational spectroscopy, vol. 2. Wiley, Chichester, UK.

Yinon, J.; Zitrin, S. (1996) *Modern Methods and applications in analysis of explosives*. John Wiley & Sons Ltd., Chichester, UK.

Sharma, S.P.; Lahiri, S.C. (2008) *Spectrochim. Acta Part A*, 70, 144.

Shao, L.; Griffiths, P.R.; Leytem, A.B. (2010) *Anal. Chem. 82*, 8027–8033.

Smith, B. (2000) *Fundamentals of Fourier Transform Infrared Spectroscopy*. CRC Press, Boca Raton, FL, USA.

Szakal, C.; Brewer, T.M. (2009) *Anal. Chem., 81*, 13, 5257.

Theriault, J.M.; Puckrin, E.; Hancock, J.; Lecavalier, P.; Lepage, C.J.; Jensen, J.O. (2004) *Appl Opt*, 43, 5870.

Urbanski, T. (1964). *Chemistry and Technology of Explosives*. Vol. 1, Macmillan Co.: New York, NY.

Van Neste, C.W.; Senesac, L.R.; Thundat, T. (2009) *Anal. Chem.* 81, 5, 1952.

Yinon, J. (1996) *J Chrom. A*, 742, 1, 205.

Vrcelj, R.M.; Gallagher, H.G.; Sherwood, J.N. (2001) *J Am. Chem. Soc.* 123:2291.

Weber, K.; Fischer, C.; van Haren, G.; Krause, H.; Bunte, G.; Schweikert, W.; Härdle, T. (2006) *Stand-Off Detection of Explosives of Suicide Bombers by Means of Open-Path FTIR Spectroscopy* in Schubert, H.; Rimski-Korsakov, A. *Stand-Off Detection of Suicide-Bombers and Mobile Subjects*, Proceedings of the NATO Advanced Research Workshop on Stand-Off Detection of Suicide Bombers and Mobile Subjects, NATO Security through Science Series B: Physics and Biophysics, Pfinztal, Germany, Springer, DE.

21

Remote Sensing of Atmospheric Trace Gases by Ground-Based Solar Fourier Transform Infrared Spectroscopy

Clare Paton-Walsh
University of Wollongong
Australia

1. Introduction

The changing composition of the earth's atmosphere is a matter of intense scientific research as we strive to understand details of the physical and chemical mechanisms that control our climate. Fourier transform spectroscopy has been applied very successfully to the study of trace gases in the atmosphere by examining terrestrial atmospheric absorption lines in the infrared spectrum from the Sun. In fact many gases were first discovered in the atmosphere during the 1940's from their absorption features in the infrared solar spectrum. These early optical absorption measurements of the atmosphere using the Sun as a source were made with grating spectrometers and examples of atmospheric gases first detected this way include methane and carbon monoxide (*Migeotte*, 1948; 1949).

Continuous or semi-continuous records of infrared solar atmospheric absorption spectra have been made from ground-based Fourier transform spectrometers (FTS) since the late 1970s and early 1980s, when the first ground-based solar-tracking FTS systems were installed at Kitt Peak National observatory in the USA and at the Jungfraujoch Observatory in Switzerland (*Goldman et al.*, 1979; *Murcray et al.*, 1978; *Zander et al.*, 1977). Initially interest was focused on the detection and quantification of stratospheric trace gases (*Rinsland et al.*, 1986; *Zander et al.*, 1986). The discovery of the Antarctic ozone hole (*Farman et al.*, 1985) intensified interest in stratospheric chemistry and helped support the establishment of the Network for Detection of Stratospheric Change (NDSC). This global network of instrument sites became operational in 1991 with ground-based FTS amongst the suite of primary techniques being used. Photographs of the instrument at the NDSC site at Wollongong, Australia are shown for illustrative purposes in figure 1 below. Other NDSC instruments are lidars for ozone, temperature, water and aerosols; microwave instruments for ozone, water and chlorine monoxide; UV/Visible spectrographs for ozone and nitrogen dioxide; Dobson/Brewer spectrophotometers for total column ozone and regular ozone sondes. The establishment of the NDSC resulted in a huge increase in the number of infrared solar absorption measurements being made around the globe during the next few years.

More recently interest in atmospheric chemistry has been focused on tropospheric pollution and anthropogenic emissions of greenhouse gases (*Barret et al.*, 2003; *Jones et al.*, 2009; *Mahieu et al.*, 1995; *Nagahama and Suzuki*, 2007; *Paton-Walsh et al.*, 2008; *Rinsland et al.*, 2000; *Rinsland et al.*, 2001; *Rinsland et al.*, 2002; *Rinsland et al.*, 2008; *Warneke et al.*, 2006; *Zhao et al.*, 2000; *Zhao*

Fig. 1. The ground-based solar Fourier Transform infrared spectrometer at Wollongong, Australia. The sun's radiation is captured by the solar tracker (shown above) and sent to the entrance optics of the high resolution Fourier transform spectrometer in the laboratory below (shown right). (Photography by R Macatangay)

et al., 2002). As a result, the NDSC has changed its emphasis and name to the Network for Detection of Atmospheric Composition and Change (NDACC) – see http://www.ndacc.org/. As well as an ever increasing number of sites in the global network the new millennium has seen an expansion into the near infrared spectra region in

an effort to provide extremely accurate and precise measurements of carbon dioxide and other greenhouse gases. The Total Column Carbon Observing Network (TCCON) was established to help characterise biogenic and oceanic sources and sinks of greenhouse gases to and from the atmosphere and to validate current and future satellite based measurements (http://www.tccon.caltech.edu/).

2. Atmospheric solar infrared Fourier transform transmission spectra

Fourier transform infrared spectroscopy is a powerful tool for monitoring the changing composition of the atmosphere because it can measure so many different gases simultaneously. The ability to measure atmospheric trace gases in the infrared is enhanced by the fact that the major components of the atmosphere (nitrogen, oxygen and argon) have no dipole moment and thus are infrared inactive.

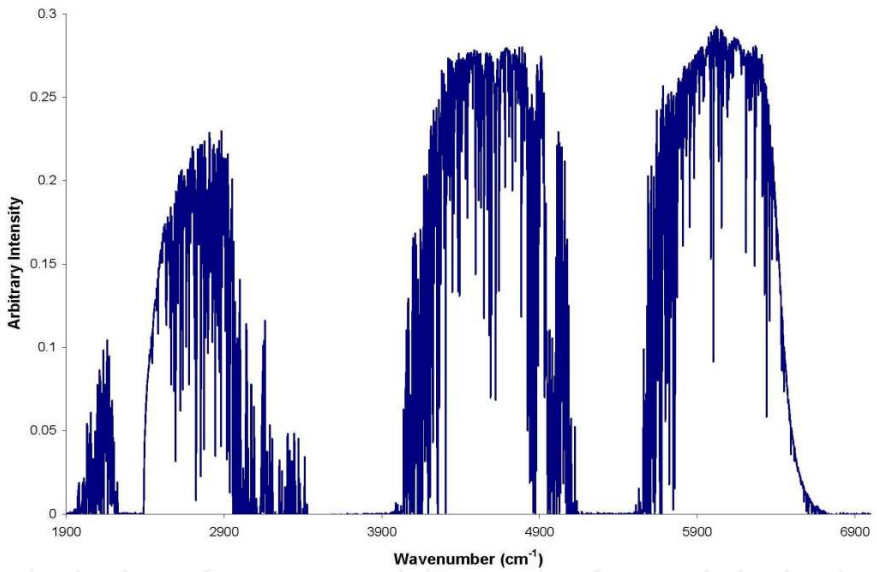


Fig. 2. An example spectrum recorded by a ground-based solar Fourier transform spectrometer

In ground-based solar Fourier transform spectroscopy, the Sun acts as the radiation source and the sample is the atmosphere. The shape of the resulting spectrum depends upon the solar radiation reaching the top of the Earth's atmosphere, absorption by the atmosphere and the optical properties of instrument recording the radiation. The solar radiation reaching the top of the Earth's atmosphere is in essence a blackbody curve at 5800K with emission and absorption lines of gases in the solar atmosphere imposed. Terrestrial atmospheric absorption lines contain information about the species of trace gases present in the atmosphere (line positions), the amounts of each gas present (line depths/areas) and some information about the altitude distribution of each gas (line shapes). In reality there is also a component of radiation as a result of atmospheric emission, but this is so small in comparison to the radiation from the sun that its effects are negligible.

3. Calculating synthetic solar atmospheric transmission spectra

Clearly it is not possible to measure a background spectrum or a set of calibration spectra when the atmosphere is the sample. For this reason the analysis of these spectra requires the calculation of a synthetic spectrum using a database of absorption line parameters such as the HITRAN database (*Rothman et al.*, 1998; *Rothman et al.*, 2003; *Rothman et al.*, 2005) and a model of the atmospheric conditions such as pressure, temperature and gas concentrations that all vary with altitude. The calculation must also take into account instrumental effects such as line shape and resolution.

The HITRAN (HIgh resolution TRANsmission) database contains calculated quantum mechanical parameters that describe the vibrational-rotational transitions of the most common atmospheric molecules. Parameters include the line positions (frequencies of absorption lines), line strengths at a reference temperature and pressure, lower state energy and pressure broadening and shift parameters.

3.1 Atmospheric absorption line positions

The characteristic absorption features seen in Figure 2 are caused by molecules absorbing radiation at frequencies that correspond to the allowed transitions between different vibrational and rotational states. The frequencies at which molecules absorb are determined by the allowed transitions between energy levels of the molecule and provide a unique identifier of trace gases in the atmosphere. Vibrational energy states are much more widely spaced than rotational energy states and so the line position is determined mainly by the change in vibrational state with small frequency differences depending on the accompanying change in rotational energy.

The energy of allowed states of a simple diatomic molecule can be approximated by Equation 1:

$$E = \hbar\sqrt{\frac{k}{\mu}}(v + 1/2) + BJ(J+1) \tag{1}$$

(if it is assumed that molecule behaves like a simple harmonic oscillator and a rigid rotor) and where:

- $\hbar$ is Planck's constant divided by 2π,
- k and μ are the force constant and the reduced mass of the molecule respectively,
- $\boldsymbol{B}$ is the rotational constant ($B = \hbar / 2I$, where **I** is the moment of inertia, the product of the reduced mass and the square of the radius of rotation)
- and v and $\boldsymbol{J}$ are the vibrational quantum number and rotational quantum number of the state respectively.

For diatomics the allowed transitions between states are $\Delta v = \pm1, \pm2, \pm3, \ldots$ and $\Delta J = \pm1$.

A vibration-rotation transition with $\Delta v = -1$ results in an *emission* spectrum and a transition with $\Delta v = +1$ an *absorption* spectrum, and both types of spectra may be accompanied by either a gain in rotational energy $\Delta J = +1$ or a loss in rotational energy $\Delta J = -1$. This produces a spectrum with two branches, one either side of the pure vibrational frequency, v_0. The R branch corresponds to a vibrational transition accompanied by a gain in rotational energy $\Delta J = +1$ and consists of a set of lines spaced approximately 2B apart to the high frequency side of the pure vibrational frequency, v_0, becoming more closely packed as the rotational energy increases further from the band centre, (assuming $B_1 < B_0$). The P branch corresponds

to a vibrational transition accompanied by a loss in rotational energy ΔJ= -1 and consists of a set of lines spaced approximately 2B apart to the low frequency side of the pure vibrational frequency, v_0, becoming more widely spaced as the rotational energy increases away from the band centre. Linear polyatomic molecules can also vibrate in a manner such that the dipole moment is changed perpendicular to the principal axis of rotational symmetry and in this case the selection rules also allow ΔJ = 0, i.e. the vibrational energy change can occur without an accompanying change in rotational energy. In such spectra the Q-branch appears at the band centre between the P and R branches (at the pure vibrational frequency, v_0). The Q-branch is a relatively intense feature because the vibrational transitions occur from all existing rotational states with approximately the same energy change and hence the same frequency.

3.2 Line broadening and line shapes

As a consequence of Heisenberg's uncertainty principle, absorption lines are never infinitely narrow. The uncertainty in the energy of a state multiplied by the uncertainty in time (the lifetime of a state) must be greater or equal to **ħ**, ($\Delta E \Delta t > \hbar$). So the shorter the lifetime, the larger the uncertainty in a state's energy and the broader the absorption or emission line (as the energy uncertainty manifests itself as an uncertainty in the frequency of the line). Uncertainty broadening due to the radiative or intra-molecular lifetime of the state in isolation is always present but for vibrational-rotational states is usually very small (< 10^{-6} cm^{-1}). Another form of uncertainty broadening, which dominates at atmospheric pressures, is collisional broadening (also called pressure broadening) and occurs when the collisions of atoms, ions or gas molecules shorten the lifetime of states. In gases it is proportional to pressure and this means that absorption lines from spectra taken through the whole atmosphere will have different shapes depending upon the distribution of the absorbing gas in the atmosphere. A gas that is mainly located in the troposphere (such as methanol) will display broad absorption lines because of the high pressure whilst predominantly stratospheric gases (like ozone) will produce much narrower lines since the pressure is low. Uncertainty broadening leads to a Lorentzian line shape contribution at a given wavenumber **υ**:

$$f_L(v) = \frac{\alpha_L / \pi}{(v - v_0)^2 + {\alpha_L}^2} \tag{2}$$

where:

- v_0 is the absorption line frequency in wavenumbers, and
- α_L is the Lorentzian half-width at half height, which is proportional to the total pressure.

Absorption lines due to atmospheric gases are also subject to Doppler broadening. Doppler broadening occurs because molecules that travel with different velocities with respect to the light source, absorb at different wavelengths, just as light from stars accelerating away from the Earth is red-shifted. Doppler broadening produces a Gaussian line shape due to the Gaussian distribution of molecular velocities:

$$f_G(v) = \frac{1}{\alpha_G \sqrt{\pi}} \exp\left(-\frac{(v - v_0)^2}{{\alpha_G}^2} \right) \tag{3}$$

where

- α_G is the Gaussian half-width at half height and is given by:

$$\alpha_G = \frac{v_l}{c}\sqrt{\frac{2kT}{m}} \tag{4}$$

where

- **k** is the Boltzman constant,
- **T** is the Kelvin temperature and
- **m** is the molecular mass.

Pressure broadening (with Lorentzian lineshape) dominates in the troposphere[1], but its effects drop off rapidly with altitude as the pressure drops. Doppler broadening (with Gausssian line shape) is temperature dependent but its variation through the atmosphere is much smaller than pressure broadening. Stratospheric gas lines are primarily Doppler broadened and the two types of broadening become equally significant at around 25 km. The convolution of Lorentzian and Gaussian line shapes produces a Voigt line shape. This variation of the shape and width of absorption lines with the pressure of the absorbing gas means that spectra of atmospheric gases contain information about the altitude of the absorbing gas as well as the total number of absorbing molecules in the path.

3.3 Atmospheric absorption line intensities

The integrated line strength of each absorption line of each molecule is determined by the transition probability, the population and degeneracy of the initial state and the number of absorbing molecules in the path. In local thermodynamic equilibrium the population of states is determined by the Boltzmann distribution, which is dependent upon the temperature of the absorbing molecules and for this reason line strengths are temperature dependent.

At a given wavenumber (v), the intensity of radiation (the radiative power) that reaches the ground $I(v)$, is related to the radiative power at the top of the atmosphere I_0(v) by Equation 5:

$$I(v) = I_0(v)e^{-m\tau(v)} \tag{5}$$

where

- $\tau(v)$ is the *optical depth* of the atmosphere and
- m is the *airmass factor* - a geometrical factor accounting for the slant path through the atmosphere.

In order to calculate a synthetic spectrum all molecules that absorb in the spectral region being simulated must be considered together. At any given wavenumber (v), the contribution to the optical depth (τ) for every absorption line **k** of each molecule **i**, is given by

$$\tau_i^k(\nu) = \sigma_i^k(\nu).a_i \tag{6}$$

where

- σ_i^k **(υ)** is the absorption coefficient or cross section at

[1] Typical values for a mid-size molecule are $\alpha_L \sim 0.15$ cm^{-1} atm^{-1} and $\alpha_G \sim 0.004$ cm^{-1}.

- **υ**, for each absorption line k of each molecule i, (typically in units cm^2 $molec^{-1}$)
- a_i is the amount of molecule i, in units of molec cm^{-2}.

The absorption coefficient σ_i^k **(υ)** is the convolution **(⊗)** of the integrated line strength, S_i^k and the pressure broadening and Doppler broadening line-shapes:

$$\sigma_i^k(\nu) = S_i^k \otimes \left[f_L(v)\right]_i^k \otimes \left[f_G(v)\right]_i^k \tag{7}$$

The HITRAN 2004 database contains line parameters for 39 atmospheric gases including the wavenumber of each absorption line, the integrated line strengths at 296K, the lower state energy E_0, the air-broadened Lorentzian half-width at atmospheric pressure and its temperature dependence.

The integrated line strength is temperature dependent because of the temperature dependence of the population of the lower-state energy and a small contribution from the spontaneous emission. The integrated line strength at any given temperature **(S_T)** may be calculated using the integrated line strengths at 296K and other parameters given in HITRAN with Equation 8:

$$S(T) = S(296)\frac{Q(296)}{Q(T)}.\frac{\exp(-\frac{c_2E_0}{T})}{\exp(-\frac{c_2E_0}{296})}.\frac{(1-\exp(-\frac{c_2v_0}{T}))}{(1-\exp(-\frac{c_2v_0}{296}))} \tag{8}$$

where

- **Q(296)** and **Q(T)** are the partition functions at these temperatures and
- c_2 is the second radiation constant, (c_2=*hc/k* = 1.439 cm K) (*Griffith*, 1996).

3.4 Calculating synthetic solar atmospheric transmission spectra

The dry atmosphere is composed of mainly nitrogen (78%) and oxygen (21%) with minor and trace gases making up the other 1%. The normal modes of vibration of nitrogen and oxygen are infrared inactive because the symmetry of the molecules is such that vibrations do not cause a change in the dipole moment, so there is nothing for incoming infra-red radiation to interact with. Thus the main constituents of the atmosphere transmit infra-red radiation and the infrared spectrum of the atmosphere as shown in Figure 2 is dominated by minor constituents and trace gases such as water, carbon dioxide and methane. This is the reason that these species are important greenhouse gases.

The temperature, pressure and concentrations of trace gases in the atmosphere change continuously with altitude. In order to calculate a synthetic spectrum the atmosphere is modelled as a series of homogeneous layers each with a defined temperature, pressure and gas composition[2]. To model the radiative transfer through the atmosphere the path length of radiation from the sun through each layer of the modelled atmosphere must be calculated using a ray-tracing algorithm that takes into account the solar zenith angle at the time of the

[2] The amounts of atmospheric trace gases are often measured in mole ratios, defined as the number of moles of the substance of interest per mole of air. For trace gases the usual units are micro moles per mole (μmol.mol^{-1}) – sometimes expressed as parts per million (ppm), e.g. 5 ppm CO means 5 molecules of carbon monoxide per million molecules of air). For ultra-trace gases the units often used are nano moles per mole (nmol.mol^{-1}) or picomoles per mole (pmol.mol^{-1}).

measurement and the effects of atmospheric curvature and refraction. At least thirty different atmospheric layers are required in order to do the radiative transfer calculation without errors in this calculation becoming a dominant uncertainty (*Meier et al.*, 2003).

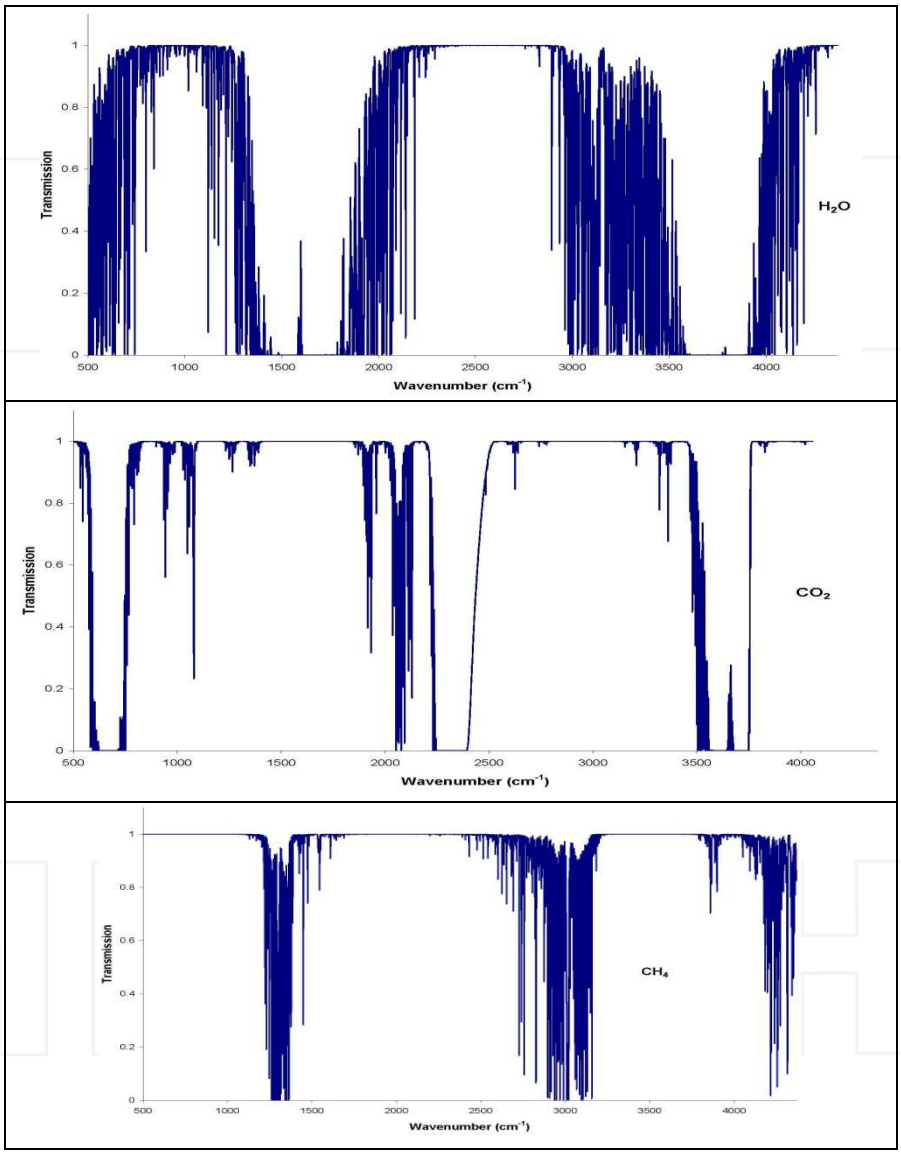


Fig. 3. The infrared transmission spectrum of water (top panel), carbon dioxide (middle panel) and methane (bottom panel) in the atmosphere from 500 cm^{-1} to 4400 cm^{-1} as simulated using the HITRAN 2000 database for a solar zenith angle of 70° and concentration profiles taken from US standard atmosphere (*Meier et al*, 2004)

Following on from Equation 6, when all the absorption lines of all molecules are considered, the total optical depth at any given frequency, **τ(ν)**, is the sum of the contribution of all the absorption lines of all molecules, for all homogeneous layers:

$$\tau(v) = \sum_{layers} \sum_{i} \sum_{k} \tau_i^k(v) \tag{9}$$

The true atmospheric transmittance spectrum, $T(v)$ is the ratio of the intensity at the ground, $I(v)$, and the intensity at the top of the atmosphere, $I_0(v)$:

$$T(v) = \frac{I(v)}{I_0(v)} = \exp(-\tau(v)) \tag{10}$$

The calculation of the synthetic spectrum must also include the effects of instrumental parameters such as the resolution and instrumental line-shape including phase error and wavelength shift. The spectrum that will actually be measured is a convolution of the intensity at the ground, $I(v)$ and the instrumental line-shape. The instrumental line-shape (ILS) may be derived using measurements of low-pressure gas cells (*Hase et al.*, 1999), or a theoretical ideal ILS can be calculated from the instrument's field of view, resolution and the apodisation function[3].
Example simulations calculated in this manner are shown in Figure 3 (*Meier et al.*, 2004). In these cases the infrared transmission spectra that result from just one component atmospheric gas at a time are given, namely water, carbon dioxide and methane. The spectra are simulated from 500 cm^{-1} to 4400 cm^{-1} using the HITRAN 2000 database for a solar zenith angle of 70° and concentration profiles taken from US standard atmosphere.

4. Theoretical basis for the retrieval of trace gas amounts from atmospheric solar infrared Fourier transform transmission spectra

4.1 Development of analysis algorithms

Over the years several different analysis algorithms have been developed to perform the retrieval of trace gas amounts from solar FTIR spectra. All the techniques calculate a synthetic spectrum using a "forward model" as described above, that includes a model of the instrumental effects and a layered model of the atmosphere with assumptions about environmental parameters such as the pressure, temperature and composition of each layer. The synthetic spectrum is compared to the measured spectrum and adjustments made to the forward model until a best fit with the measured spectrum is obtained.
Different analysis algorithms allow different adjustments to the forward model in order to achieve the best fit to the measurement.

- "SFIT1" (*Rinsland et al.*, 1982; *Rinsland et al.*, 1984) and "GFIT" (*Washenfelder et al.*, 2006) allow only a scaling of the *a priori* concentration profile of the absorbing gases to achieve best fit. This means that the distribution of the absorbing gas is defined entirely

[3] An apodisation function is a mathematical function that is applied to the interferogram that may give greater weight to the information around zero path difference compared to information at greater optical path differences. In the analyses presented in this thesis a boxcar apodisation function was used that gives equal weight to the information throughout the interferogram.

by the *a priori* assumptions – the fitting process for example cannot put more gas in the troposphere and less in the stratosphere but must multiply the concentration in each of the modelled layers by the same multiplicative factor.

- Later development "SFIT2" (*Pougatchev et al.*, 1995) using optimal estimation techniques (*Rodgers*, 1990) aimed to extract the limited spectral information on the vertical distribution of the target gas, which is contained in the shape of the absorption features as a result of pressure broadening by the surrounding atmospheric gases (*Hase et al.*, 2004; *Rinsland et al.*, 1998).
- More recently another analysis code ("PROFITT") has been developed using optimal estimation techniques that allows the temperature profile and concentration profiles to be adjusted (*Hase et al.*, 2004).

4.2 Inverse modelling and optimal estimation

In contrast to GFIT, SFIT2 allows the volume mixing ratio profile of the absorbing gas in the simulated spectrum to be adjusted so that the shape of the absorption line can best fit the measured spectrum. The principle difficulty in this technique is that the radiative transfer model requires at least thirty atmospheric layers to achieve a reasonable model of the transmission of solar radiation through the atmosphere, but the shape of an absorption feature typically contains between one and five independent pieces of information. Thus the problem is mathematically underdetermined and much of the information must still be provided by the *a priori* concentration profile in the forward model.

SFIT2 uses an inverse modeling technique (*Rodgers*, 1990; *Rodgers*, 2000) to extract the volume mixing ratio profile of the gases of interest from the measured spectrum. The volume mixing ratios of the gas of interest at each of the modeled layers are the variables of interest (called state variables or together the state vector **x,** with n elements). The measured spectrum (a series of observed radiances at different frequencies) is represented by the observation vector **y** (with m elements), and the forward model, **F**, describes the relationship between the observation vector **y** (the spectrum) and the state vector **x**, (the volume mixing ratios of the gases of interest).

$$y = F(x, b) + \varepsilon \tag{11}$$

where **b** is a *parameter vector* including all model variables that are not to be optimized (also called the *model parameters*), and **ε** is the error vector including errors in the observations, in the forward model, and in the model parameters.

Inverting Equation 11, **x** may be obtained given **y**, but due to the error term **ε** the best that can be achieved is a statistical estimate. As stated before the problem is mathematically underdetermined and in optimal estimation **x** is weighted by our prior (*a priori*) knowledge of the state vector $\mathbf{x_a}$ (the *a priori* volume mixing ratio profile). The optimal solution of **x** including this *a priori* knowledge is called the "*optimal estimate*" or the '*retrieval*' (also sometimes called the "*a posteriori* solution").

The *a priori* estimate has its own error: $x_a = x + \varepsilon_a$ and the key to solving the optimal estimation problem is weighting the error statistics of ε and ε_a (*Rodgers*, 2000).

The inverse problem lends itself to the use of matrix algebra, and the *Jacobian matrix*, ***K***, is a linearization of the forward model that represents the sensitivity of the observation variables **y** to the state variables **x**, assembled in matrix form (*Jacob*, 2007):

$$K = \nabla_x F = \frac{\partial y}{\partial x} \tag{12}$$

K changes with **x** and so it is calculated initially for the *a priori* value $\mathbf{x_a}$ and then re-calculated iteratively until the solution converges. The algorithm iterates until the cost function ***J(x)*** is minimised by solving for $\nabla_x J(x) = 0$ as shown in Equation 13:

$$\nabla_x J(x) = 2S_a^{-1}(x - x_a) + 2K^T S_\varepsilon^{-1}(Kx - y) = 0 \tag{13}$$

where S_a and S_ε are the *a priori* error and observational error covariance matrices respectively (the matrix equivalents of ε_a and ε). Note that observational error includes errors in the forward model as well as spectral noise and that in many instances it is the forward model error that dominates the S_ε matrix.
The solution to Equation 13 yields the *optimal estimate* or *retrieval* $\hat{x}$ and is given by Equation 14:

$$\hat{x} = x_a + G(y - Kx_a) \tag{14}$$

where G is known as the *gain matrix* and describes the sensitivity of the retrieval to the observations, i.e. $G = \frac{\partial \hat{x}}{\partial y}$, and is given by Equation 15:

$$G = (K^T S_\varepsilon^{-1} K + S_a^{-1})^{-1} K^T S_\varepsilon^{-1} \tag{15}$$

4.3 Uncertainties in retrieved trace gas amounts and the "averaging kernel"

The ability of the measured spectrum to constrain the volume mixing ratio profile of the gas of interest, (the state vector, x) is given by the *averaging kernel matrix* $A = \frac{\partial \hat{x}}{\partial x}$, which represents the sensitivity of the retrieval $\hat{x}$ to the true state vector x.

The *averaging kernel matrix,* **A=GK**, i.e. it is the product of the gain matrix $G = \frac{\partial \hat{x}}{\partial y}$ and the Jacobian matrix $K = \frac{\partial y}{\partial x}$.

The information content may be defined in terms of the degrees of freedom for signal which is the trace of the averaging kernel. Mathematically this is the sum of the diagonal elements of the averaging kernel matrix (*Rodgers*, 2000).
Using the averaging kernel matrix leads to an alternative expression for the *optimal estimate* or *retrieval* $\hat{x}$ **:**

$$\hat{x} = Ax + (I_n - A)x_a + G\varepsilon \tag{16}$$

where I_n is the identity matrix of dimension n. There are three terms on the right hand side of Equation 16. The first term, Ax represents the contribution of the true state x to the solution. The second term $(I_n - A)x_a$ represents the contribution of the *a priori* assumptions. The third term $G\varepsilon$ is the contribution from random observational error. An ideal measurement would have an averaging kernel matrix that was an n dimensional identity matrix, $A = I_n$ in which case the 2nd term is zero (*Jacob*, 2007; *Rodgers*, 2000).

The second term in Equation 16, $[(In - A)x_a]$ is called the *smoothing uncertainty* (since it results in a smoothing of the retrieval towards the *a priori* values). The third term, $G\varepsilon$, is known as the *retrieval uncertainty* (or *signal-to-noise uncertainty*) but again it should be noted that this includes not only measurement noise but also errors in the forward model such as errors in the HITRAN database and that these factors often dominate the spectral noise. There may also be errors in the forward model that are correlated with one another and are therefore not random, potentially leading to uncharacterised biases in the retrievals. One other error that can be estimated is the *temperature uncertainty*, due to errors in the assumed temperature profile in the forward model. These lead to errors in the line strengths and resulting retrieved trace gas amounts that may be calculate using Equation 8.

5. A summary of past successes for ground-based solar Fourier transform infrared spectroscopy of atmospheric composition

In the late 1970's there were two groups of researchers actively engaged in making solar infrared measurements of atmospheric composition from the ground. The scientific focus at the time was on understanding the chemistry of the stratosphere and many measurements were made using balloon-borne grating spectrometers (*Murcray et al*., 1975; *Zander*, 1976). *Bradford et al* (1976) outlined the feasibility of monitoring atmospheric trace gases using ground-based high resolution infrared spectroscopy. Within a couple of years the first scientific results from ground-based stations had been published, confirming the presence of hydrogen fluoride in the stratosphere from ground-based measurements from the Jungfraujoch in Switzerland, (47°N, 8.0°E) (*Zander et al*., 1977) and identifying atmospheric absorption features of ammonia in spectra from Kitt peak in the USA, (32°N, 112°W) (*Murcray et al*., 1978). In 1979 *Goldman et al*, published their "New Atlas of Infrared Solar Spectra" which included telluric absorption features of carbon dioxide, water, methane, carbon monoxide, nitrous oxide, ozone, carbonyl fluoride, chlorine nitrate, difluorochloromethane (CFC-22), dichlorodifluoromethane (CFC-12), sulphur hexafluoride and nitric acid as well as a number of absorption features from carbon monoxide and hydroxyl radicals in the solar atmosphere.

As the capabilities increased more trace gases were added to the list of measureable species including hydrogen cyanide (*Rinsland et al*., 1982), nitric oxide (*Rinsland et al*., 1984) and chlorine nitrate (*Zander et al*., 1986). After the discovery of the Antarctic ozone hole (*Farman et al*., 1985) and the establishment of the Network for Detection of Stratospheric Change (NDSC), the number of ground-based Fourier transform spectrometers making solar atmospheric absorption measurements increased dramatically. During the first half of the 1990's Fourier transform solar remote sensing spectrometers were installed at a number of new sites including Lauder, New Zealand (45°S, 170°E) in 1990 (*Jones et al*., 1994); Mauna Loa, Hawaii (20°N, 156°W)(*David et al*., 1993) & Arrival Heights, Antarctica (78°S, 167°E)(*Kreher et al*., 1996; *Wood et al*., 2004) in 1991, Ny Alesund, Spitzbergen (79°N, 12°E) in 1992 (*Notholt and Schrems*, 1994) and Harestua, Norway (60°N, 11°E) in 1994 (*Mellqvist et al*., 2002). In 1994 the first measurements using the moon as a light source were reported from Spitzbergen, enabling measurements of the Arctic atmosphere during polar night (*Notholt*, 1994).

In 1995 three additional sites were added to the network at Zugspitze, Germany (47°N, 11°E) (*Sussmann and Schafer*, 1997), Rikubetsu, Japan (44°N, 144°E)(*Zhao et al*., 2000) and Wollongong, Australia (34°S, 151°E) (*Rinsland et al*., 2001). The following year saw a further

three sites established at Kiruna, Sweden, (68°N, 20°E) (*Blumenstock et al*., 1997), Eureka, Nunavit (80°N, 86°W) (*Batchelor et al*., 2009; *Donovan et al*., 1997) and Moshiri, Japan (44°N, 142°E) (*Zhao et al*., 2000). Measurements at Tsukuba, Japan (36°N, 140°E) started in 1998, followed by Thule, Greenland (77°N, 69°W), Poker Flat, Alaska (65°N, 147°W) and Izana, Tenerife Island (28°N, 16°W) all in 1999.

In the meanwhile there were a number of mobile instruments operating. The Jet Propulsion Laboratory's "Mark IV" instrument was making regular balloon flights from Alaska interspersed with ground-based measurements from Esrange (68°N, 21°E), Fairbanks (65°N, 148°W), Mount Barcroft (38°N, 118°W) or Table Mountain (34°N, 118°W) (*Toon et al*., 1999b). The Alfred Wegner Institute had a ship-borne spectrometer onboard the Polarstern (*Notholt et al*., 1995; *Notholt et al*., 2000), and the National Physical Laboratory had a mobile instrument that made a series of side-by-side instrument intercomparisons with a number of other spectrometers in the network as well as a number of campaign measurements at Are, Sweden, Aberdeen, Scotland and Calar Alto, Spain (*Bell et al*., 1994; *Bell et al*., 1998; *Paton-Walsh et al*., 1997). In addition to instrument intercomparisons, there were a number of algorithm intercomparison exercises and a standard procedure for characterising the instrument line-shape was developed (*Hase et al*., 1999; *Hase et al*., 2004). *Apriori* profiles were most commonly based upon measurements from the balloon-based MarkIV instrument (*Toon et al*., 1999a) or from the ATMOS instrument that was flown on the Space Shuttle (*Gunson et al*., 1996).

Early in the new millennium the NDSC changed its name to the Network for Detection of Atmospheric Composition Change (NDACC) to highlight the change in scientific focus from stratospheric chemistry to changing tropospheric composition and greenhouse gases. The NDACC database (see http://www.ndacc.org/) contains standard gases for most stations equipped with a remote sensing FTIR spectrometer including ozone, nitric acid, hydrogen chloride, chlorine nitrate, hydrogen fluoride, nitrous oxide, carbon dioxide, carbon monoxide, methane, ethane and hydrogen cyanide. Many stations also provide other gases such as CFCs, nitrogen dioxide, nitrogen oxide and acetylene. Further stations were added including Toronto, Canada (44°N, 80°W) in 2002 and Bremen, Germany (53°N, 9°E) in 2004, whilst campaign measurements have been made at sites including Reunion Island, (22° S, 56°E) (*Senten et al*., 2008), Paramaribo, Suriname (6°N, 55°W)(*Petersen et al*., 2010) and Addis Ababa (9°N, 39°E). Further gases have also been identified in spectra from solar remote sensing Fourier transform spectrometers included chlorine monoxide (*Bell et al*., 1996), formic acid, (*Rinsland et al*., 2004), ethylene(*Rinsland et al*., 2005) and methanol (*Paton-Walsh et al*., 2008). In addition there have been a number of studies that examine different isotopes (*Frankenberg et al*., 2009; *Goldman et al*., 2000; *Goldman et al*., 2002; *Haverd et al*., 2005; *Irion et al*., 1996; *Meier and Notholt*, 1996; *Notholt et al*., 2010). Trends in both stratospheric gases eg (*Rinsland et al*., 2003) and tropospheric gases eg (*Jones et al*., 2009) have been established from NDACC remote sensing FTIR spectrometers. Recently there has been significant interest in the ability of this technique to characterise free tropospheric water vapour and its isotopes (*Palm et al*., 2010; *Schneider et al*., 2010). The NDACC continues to provide ground-validation for a number of retrieved products from several different satellite-based sensors (*Dils et al*., 2006; *Dupuy et al*., 2009; *Mahieu et al*., 2008; *Payan et al*., 2009; *Strong et al*., 2008; *Vigouroux et al*., 2007; *Wolff et al*., 2008; *Yurganov et al*., 2008).

In 2004 the Total Carbon Column Observing Network (TCCON) was established with the aim of making very accurate measurements of greenhouse gases in the near-infrared spectral region (see https://tccon-wiki.caltech.edu/). The primary function of TCCON was

to act as a network of ground-thruthing stations for the Orbiting Carbon Observatory (*Crisp et al.*, 2004), however TCCON also had a remit to provide independent constraints to models of the carbon cycle and provide validation data to other satellite-based sensors. New sites were established at Park Falls, Wisconsin, USA (46°N, 90°W)(*Washenfelder et al.*, 2006) and Darwin, Australia (12°S, 131°E) (*Deutscher et al.*, 2009) whilst in a number of existing NDACC sites have been upgraded to allow for the extended spectral coverage and improved precision demanded for TCCON. The failure of the launch of the Orbiting Carbon Observatory has meant that TCCON has yet to fulfil its primary function however a re-launch is now planned and in the meanwhile TCCON has started to elucidate details of the carbon cycle itself (*Yang et al.*, 2007). More TCCON stations are being set-up and the data is available for the validation of other satellite instruments such as SCIAMACHY, AIRS, IASI and GOSAT(*Bosch et al.*, 2006).

6. Concluding remarks

Remote sensing of atmospheric trace gases by ground-based solar Fourier transform infrared spectroscopy has developed rapidly since its origins in the 1970s. It has had major successes in characterising the composition of the atmosphere and trends in both stratospheric and tropospheric gases. Whilst the technique has greater uncertainties than ground level *in situ* measurements of atmospheric composition, it characterises the total atmospheric column amount. Interpretation of ground level measurements is often complicated by the effects of vertical transport and changing boundary layer height, but total column measurements are less sensitive to this problem because the measurement is integrated over the whole atmosphere. A significant drawback to these remote sensing techniques is that they are numerically ill-posed and thus some *apriori* information is required to derive total column amounts from the spectra. In future atmospheric models are likely to be constrained by use of combined *in situ* and remotely sensed data such that the total uncertainties are minimised.

Remote sensing Fourier transform spectrometers play a vital role in our efforts to understand the changing composition of the atmosphere. The networks of ground-based instruments are continuing to expand, with improving precision & accuracy. These measurements will continue to improve our understanding of atmospheric composition and chemistry and provide ground validation for a new generation of satellite-based instruments.

7. References

Barret, B., M. De Maziere, and E. Mahieu (2003), Ground-based FTIR measurements of CO from the Jungfraujoch: characterisation and comparison with in situ surface and MOPITT data, *Atmospheric Chemistry And Physics*, *3*, 2217-2223.

Batchelor, R. L., K. Strong, R. Lindenmaier, et al. (2009), A New Bruker IFS 125HR FTIR Spectrometer for the Polar Environment Atmospheric Research Laboratory at Eureka, Nunavut, Canada: Measurements and Comparison with the Existing Bomem DA8 Spectrometer, *Journal of Atmospheric and Oceanic Technology*, *26*(7), 1328-1340.

Bell, W., N. A. Martin, T. D. Gardiner, et al. (1994), Column Measurements of Stratospheric Trace Species over Are, Sweden in the Winter of 1991-1992, *Geophysical Research Letters*, *21*(13), 1347-1350.

Bell, W., C. PatonWalsh, T. D. Gardiner, et al. (1996), Measurements of stratospheric chlorine monoxide (ClO) from groundbased FTIR observations, *Journal of Atmospheric Chemistry*, *24*(3), 285-297.

Bell, W., C. Paton-Walsh, T. D. Gardiner, et al. (1998), Ground-based FTIR measurements of stratospheric trace species from Aberdeen during winter and spring 1993/94 and 1994/95 and comparison with a 3D model, *Journal of Atmospheric Chemistry*, *30*(1), 119-130.

Blumenstock, T., H. Fischer, A. Friedle, et al. (1997), Column amounts of $ClONO_2$, HCl, HNO_3, and HF from ground-based FTIR measurements made near Kiruna, Sweden, in late winter 1994, *Journal of Atmospheric Chemistry*, *26*(3), 311-321.

Bosch, H., G. C. Toon, B. Sen, et al. (2006), Space-based near-infrared CO2 measurements: Testing the Orbiting Carbon Observatory retrieval algorithm and validation concept using SCIAMACHY observations over Park Falls, Wisconsin, *Journal Of Geophysical Research-Atmospheres*, *111*(D23).

Bradford, C. M., F. H. Murcray, J. W. Vanallen, et al. (1976), GROUND LEVEL DETECTION AND FEASIBILITY FOR MONITORING OF SEVERAL TRACE ATMOSPHERIC CONSTITUENTS BY HIGH-RESOLUTION INFRARED SPECTROSCOPY, *Geophysical Research Letters*, *3*(7), 387-390.

Crisp, D., R. M. Atlas, F. M. Breon, et al. (2004), The orbiting carbon observatory (OCO) mission, in *Trace Constituents in the Troposphere and Lower Stratosphere*, edited by J. P. Burrows and A. M. Thompson, pp. 700-709, Pergamon-Elsevier Science Ltd, Kidlington.

David, S. J., S. A. Beaton, M. H. Anderberg, et al. (1993), Determination of Total Ozone over Mauna-Loa Using Very High-Resolution Infrared Solar Spectra, *Geophysical Research Letters*, *20*(19), 2055-2058.

Deutscher, N. M., D. W. T. Griffith, G. W. Bryant, et al. (2009), Total column CO2 measurements at Darwin, Australia - Site description and calibration against in situ aircraft profiles, *Journal of Geophysical Research*.

Dils, B., M. De Maziere, J. F. Muller, et al. (2006), Comparisons between SCIAMACHY and ground-based FTIR data for total columns of CO, CH4, CO2 and N2O, *Atmospheric Chemistry And Physics*, *6*, 1953-1976.

Donovan, D. P., H. Fast, Y. Makino, et al. (1997), Ozone, column ClO, and PSC measurements made at the NDSC Eureka observatory (80 degrees N, 86 degrees W) during the spring of 1997, *Geophysical Research Letters*, *24*(22), 2709-2712.

Dupuy, E., K. A. Walker, J. Kar, et al. (2009), Validation of ozone measurements from the Atmospheric Chemistry Experiment (ACE), *Atmospheric Chemistry And Physics*, *9*(2), 287-343.

Farman, J. C., B. G. Gardiner, and J. D. Shanklin (1985), Large Losses of Total Ozone in Antarctica Reveal Seasonal Clox/Nox Interaction, *Nature*, *315*(6016), 207-210.

Frankenberg, C., K. Yoshimura, T. Warneke, et al. (2009), Dynamic Processes Governing Lower-Tropospheric HDO/H2O Ratios as Observed from Space and Ground, *Science*, *325*(5946), 1374-1377.

Goldman, A., R. D. Blatherwick, F. H. Murcray, et al. (1979), NEW ATLAS OF IR SOLAR SPECTRA, *Applied Optics*, *18*(5), 604-605.

Goldman, A., M. T. Coffey, T. M. Stephen, et al. (2000), Isotopic OCS from high-resolution balloon-borne and ground-based infrared solar absorption spectra, *Journal of Quantitative Spectroscopy & Radiative Transfer*, *67*(6), 447-455.

Goldman, A., C. P. Rinsland, A. Perrin, et al. (2002), Weak ozone isotopic absorption in the 5 mu m region from high resolution FTIR solar spectra, *Journal of Quantitative Spectroscopy & Radiative Transfer*, *74*(1), 133-138.

Griffith, D. W. T. (1996), Synthetic calibration and quantitative analysis of gas phase infrared spectra, *Applied Spectroscopy*, *50*(1), 59-70.

Gunson, M. R., M. M. Abbas, M. C. Abrams, et al. (1996), The Atmospheric Trace Molecule Spectroscopy (ATMOS) experiment: Deployment on the ATLAS Space Shuttle missions, *Geophysical Research Letters*, *23*(17), 2333-2336.

Hase, F., T. Blumenstock, and C. Paton-Walsh (1999), Analysis of the instrumental line shape of high-resolution Fourier transform IR spectrometers with gas cell measurements and new retrieval software, *Applied Optics*, *38*(15), 3417-3422.

Hase, F., J. W. Hannigan, M. T. Coffey, et al. (2004), Intercomparison of retrieval codes used for the analysis of high-resolution, ground-based FTIR measurements, *Journal of Quantitative Spectroscopy & Radiative Transfer*, *87*(1), 25-52.

Haverd, V., G. C. Toon, and D. W. T. Griffith (2005), Evidence for altitude-dependent photolysis-induced O-18 isotopic fractionation in stratospheric ozone, *Geophysical Research Letters*, *32*(22), 4.

Irion, F. W., M. R. Gunson, C. P. Rinsland, et al. (1996), Heavy ozone enrichments from ATMOS infrared solar spectra, *Geophysical Research Letters*, *23*(17), 2377-2380.

Jacob, D. J. (2007), Lectures on Inverse Modeling, edited.

Jones, N. B., M. Koike, W. A. MAtthews, et al. (1994), Southern Hemisphere Seasonal Cycle in total column Nitric Acid, *Geophys. Res. Lett.*, *21*(7), 593-596.

Jones, N. B., K. Riedel, W. Allan, et al. (2009), Long-term tropospheric formaldehyde concentrations deduced from ground-based fourier transform solar infrared measurements, *Atmospheric Chemistry And Physics*, *9*(18), 7131-7142.

Kreher, K., J. G. Keys, P. V. Johnston, et al. (1996), Ground-based measurements of OClO and HCl in austral spring 1993 at Arrival Heights, Antarctica, *Geophysical Research Letters*, *23*(12), 1545-1548.

Mahieu, E., C. P. Rinsland, R. Zander, et al. (1995), Vertical Column Abundances of HCN Deduced from Ground-Based Infrared Solar Spectra - Long-Term Trend and Variability, *Journal of Atmospheric Chemistry*, *20*(3), 299-310.

Mahieu, E., P. Duchatelet, P. Demoulin, et al. (2008), Validation of ACE-FTS v2.2 measurements of HCl, HF, CCl3F and CCl2F2 using space-, balloon- and ground-based instrument observations, *Atmospheric Chemistry And Physics*, *8*(20), 6199-6221.

Meier, A., and J. Notholt (1996), Determination of the isotopic abundances of heavy O-3 as observed in arctic ground-based FTIR-spectra, *Geophysical Research Letters*, *23*(5), 551-554.

Meier, A., A. Goldman, P. Manning, et al. (2003), Improvements to Air Mass Calculations for Ground-Based Infrared Measurements, *J. Quant. Spectrosc. Radiat. Transfer*, *doi:10.1016/S0022-4073(02)00018-3*.

Meier, A., G. Toon, C. Rinsland, et al. (2004), Spectroscopic Atlas of Atmospheric Microwindows in the Middle Infra-red, Institutet for Rymdfysik, Kiruna, Sweden.

Mellqvist, J., B. Galle, T. Blumenstock, et al. (2002), Ground-based FTIR observations of chlorine activation and ozone depletion inside the Arctic vortex during the winter of 1999/2000, *Journal Of Geophysical Research-Atmospheres*, *107*(D20).

Migeotte, M. V. (1948), Spectroscopic evidence of methane in the earth's atmosphere, *Phys Rev*, *73*, 519-520.

Migeotte, M. V. (1949), The fundamental band of carbon monoxide at 4.7 microns in the solar spectrum, *Phys Rev*, *75*, 1108-1109.

Murcray, D. G., F. S. Bonomo, J. N. Brooks, et al. (1975), DETECTION OF FLUOROCARBONS IN STRATOSPHERE, *Geophysical Research Letters*, *2*(3), 109-112.

Murcray, D. G., A. Goldman, C. M. Bradford, et al. (1978), IDENTIFICATION OF NU-2 VIBRATION-ROTATION BAND OF AMMONIA IN GROUND LEVEL SOLAR SPECTRA, *Geophysical Research Letters*, *5*(6), 527-530.

Nagahama, Y., and K. Suzuki (2007), The influence of forest fires on CO, HCN, C2H6, and C2H2 over northern Japan measured by infrared solar spectroscopy, *Atmospheric Environment*, *41*(40), 9570-9579.

Notholt, J. (1994), The Moon as a Light-Source for Ftir Measurements of Stratospheric Trace Gases During the Polar Night - Application for Hno3 in the Arctic, *Journal Of Geophysical Research-Atmospheres*, *99*(D2), 3607-3614.

Notholt, J., and O. Schrems (1994), GROUND-BASED FTIR MEASUREMENTS OF VERTICAL COLUMN DENSITIES OF SEVERAL TRACE GASES ABOVE SPITSBERGEN, *Geophysical Research Letters*, *21*(13), 1355-1358.

Notholt, J., I. Beninga, and O. Schrems (1995), SHIP-BORNE FT-IR MEASUREMENTS OF ATMOSPHERIC TRACE GASES ON A SOUTH (33-DEGREES-S) TO NORTH (53-DEGREES-N) ATLANTIC TRAVERSE, *Applied Spectroscopy*, *49*(10), 1525-1527.

Notholt, J., G. C. Toon, C. P. Rinsland, et al. (2000), Latitudinal variations of trace gas concentrations in the free troposphere measured by solar absorption spectroscopy during a ship cruise, *Journal Of Geophysical Research-Atmospheres*, *105*(D1), 1337-1349.

Notholt, J., G. C. Toon, S. Fueglistaler, et al. (2010), Trend in ice moistening the stratosphere - constraints from isotope data of water and methane, *Atmospheric Chemistry And Physics*, *10*(1), 201-207.

Palm, M., C. Melsheimer, S. Noel, et al. (2010), Integrated water vapor above Ny Alesund, Spitsbergen: a multi-sensor intercomparison, *Atmospheric Chemistry And Physics*, *10*(3), 1215-1226.

Paton-Walsh, C., W. Bell, T. Gardiner, et al. (1997), An uncertainty budget for ground-based Fourier transform infrared column measurements of HCl, HF, N2O, and HNO3, deduced from results of side-by-side instrument intercomparisons, *Journal Of Geophysical Research-Atmospheres*, *102*(D7), 8867-8873.

Paton-Walsh, C., S. R. Wilson, N. B. Jones, et al. (2008), Measurement of methanol emissions from Australian wildfires by ground-based solar Fourier transform spectroscopy, *Geophysical Research Letters*, *35*(8).

Payan, S., C. Camy-Peyret, H. Oelhaf, et al. (2009), Validation of version-4.61 methane and nitrous oxide observed by MIPAS, *Atmospheric Chemistry And Physics*, *9*(2), 413-442.

Petersen, A. K., T. Warneke, C. Frankenberg, et al. (2010), First ground-based FTIR observations of methane in the inner tropics over several years, *Atmospheric Chemistry And Physics*, *10*(15), 7231-7239.

Pougatchev, N. S., B. J. Connor, and C. P. Rinsland (1995), Infrared measurements of the ozone vertical-distribution above Kitt Peak, *Journal Of Geophysical Research-Atmospheres*, *100*(D8), 16689-16697.

Rinsland, C. P., M. A. H. Smith, P. L. Rinsland, et al. (1982), Ground-based infrared spectroscopic measurements of atmospheric hydrogen-cyanide, *Journal Of Geophysical Research-Oceans and Atmospheres*, *87*(NC13), 1119-1125.

Rinsland, C. P., R. E. Boughner, J. C. Larsen, et al. (1984), Diurnal-variations of atmospheric nitric-oxide - ground-based infrared spectroscopic measurements and their interpretation with time-dependent photochemical model-calculations, *Journal Of Geophysical Research-Atmospheres*, *89*(ND6), 9613-9622.

Rinsland, C. P., R. Zander, L. R. Brown, et al. (1986), Detection of Carbonyl Fluoride in the Stratosphere, *Geophysical Research Letters*, *13*(8), 769-772.

Rinsland, C. P., N. B. Jones, B. J. Connor, et al. (1998), Northern and Southern Hemisphere Ground-Based Infrared Spectroscopic Measurements of Tropospheric Carbon Monoxide and Ethane, *Journal of Geophysical Research*, *103*, 28,197-128,218.

Rinsland, C. P., E. Mahieu, R. Zander, et al. (2000), Free tropospheric CO, $C2_{H6}$, and HCN above central Europe: Recent measurements from the Jungfraujoch station including the detection of elevated columns during 1998, *Journal Of Geophysical Research-Atmospheres*, *105*(D19), 24235-24249.

Rinsland, C. P., A. Meier, D. W. T. Griffith, et al. (2001), Ground-based measurements of tropospheric CO, C_2H_6, and HCN from Australia at 34 degrees S latitude during 1997-1998, *Journal Of Geophysical Research-Atmospheres*, *106*(D18), 20913-20924.

Rinsland, C. P., A. Goldman, E. Mahieu, et al. (2002), Ground-based infrared spectroscopic measurements of carbonyl sulfide: Free tropospheric trends from a 24-year time series of solar absorption measurements, *Journal Of Geophysical Research-Atmospheres*, *107*(D22), -.

Rinsland, C. P., E. Mathieu, R. Zander, et al. (2003), Long-term trends of inorganic chlorine from ground-based infrared solar spectra: Past increases and evidence for stabilization, *Journal Of Geophysical Research-Atmospheres*, *108*(D8), 21.

Rinsland, C. P., E. Mahieu, R. Zander, et al. (2004), Free tropospheric measurements of formic acid (HCOOH) from infrared ground-based solar absorption spectra: Retrieval approach, evidence for a seasonal cycle, and comparison with model calculations, *Journal Of Geophysical Research-Atmospheres*, *109*(D18).

Rinsland, C. P., C. Paton-Walsh, N. B. Jones, et al. (2005), High spectral resolution solar absorption measurements of ethylene (C_2H_4) in a forest fire smoke plume using HITRAN parameters: Tropospheric vertical profile retrieval, *Journal Of Quantitative Spectroscopy and Radiative Transfer*, *96*(2), 301.

Rinsland, C. P., L. Chiou, E. Mahieu, et al. (2008), Measurements of long-term changes in atmospheric OCS (carbonyl sulfide) from infrared solar observations, *Journal of Quantitative Spectroscopy & Radiative Transfer*, *109*(16), 2679-2686.

Rodgers, C. D. (1990), Characterization and error analysis of profiles retrieved from remote sounding measurements, *Journal of Geophysical Research*, *95*, 5587-5595.

Rodgers, C. D. (2000), *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Scientific Publishing Co. Pte. Ltd, New Jersey.

Rothman, L. S., C. P. Rinsland, A. Goldman, et al. (1998), The HITRAN molecular spectroscopic database and HAWKS (HITRAN Atmospheric Workstation): 1996 edition, *Journal of Quantitative Spectroscopy & Radiative Transfer*, *60*(5), 665-710.

Rothman, L. S., A. Barbe, D. C. Benner, et al. (2003), The HITRAN molecular spectroscopic database: edition of 2000 including updates through 2001, *Journal of Quantitative Spectroscopy & Radiative Transfer*, *82*(1-4), 5-44.

Rothman, L. S., D. Jacquemart, A. Barbe, et al. (2005), The HITRAN 2004 molecular spectroscopic database, *Journal of Quantitative Spectroscopy & Radiative Transfer*, *96*(2), 139-204.

Schneider, M., K. Yoshimura, F. Hase, et al. (2010), The ground-based FTIR network's potential for investigating the atmospheric water cycle, *Atmospheric Chemistry And Physics*, *10*(7), 3427-3442.

Senten, C., M. De Maziere, B. Dils, et al. (2008), Technical note: New ground-based FTIR measurements at Ile de La Reunion: observations, error analysis, and comparisons with independent data, *Atmospheric Chemistry And Physics*, *8*(13), 3483-3508.

Strong, K., M. A. Wolff, T. E. Kerzenmacher, et al. (2008), Validation of ACE-FTS N2O measurements, *Atmospheric Chemistry And Physics*, *8*(16), 4759-4786.

Sussmann, R., and K. Schafer (1997), Infrared spectroscopy of tropospheric trace gases: Combined analysis of horizontal and vertical column abundances, *Applied Optics*, *36*(3), 735-741.

Toon, G. C., J. F. Blavier, B. Sen, et al. (1999a), Comparison of MkIV balloon and ER-2 aircraft measurements of atmospheric trace gases, *Journal Of Geophysical Research-Atmospheres*, *104*(D21), 26779-26790.

Toon, G. C., J. F. Blavier, B. Sen, et al. (1999b), Ground-based observations of Arctic O_3 loss during spring and summer 1997, *Journal Of Geophysical Research-Atmospheres*, *104*(D21), 26497-26510.

Vigouroux, C., M. De Maziere, Q. Errera, et al. (2007), Comparisons between ground-based FTIR and MIPAS N2O and HNO3 profiles before and after assimilation in BASCOE, *Atmospheric Chemistry And Physics*, *7*, 377-396.

Warneke, T., J. F. Meirink, P. Bergamaschi, et al. (2006), Seasonal and latitudinal variation of atmospheric methane: A ground-based and ship-borne solar IR spectroscopic study, *Geophysical Research Letters*, *33*(14).

Washenfelder, R. A., G. C. Toon, J. F. Blavier, et al. (2006), Carbon dioxide column abundances at the Wisconsin Tall Tower site, *Journal Of Geophysical Research-Atmospheres*, *111*(D22).

Wolff, M. A., T. Kerzenmacher, K. Strong, et al. (2008), Validation of HNO3, ClONO2, and N2O5 from the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS), *Atmospheric Chemistry And Physics*, *8*(13), 3529-3562.

Wood, S. W., R. L. Batchelor, A. Goldman, et al. (2004), Ground-based nitric acid measurements at Arrival Heights, Antarctica, using solar and lunar Fourier transform infrared observations, *Journal Of Geophysical Research-Atmospheres*, *109*(D18).

Yang, Z., R. A. Washenfelder, G. Keppel-Aleks, et al. (2007), New constraints on Northern Hemisphere growing season net flux, *Geophysical Research Letters*, *34*(12), 6.

Yurganov, L. N., W. W. McMillan, A. V. Dzhola, et al. (2008), Global AIRS and MOPITT CO measurements: Validation, comparison, and links to biomass burning variations and carbon cycle, *Journal Of Geophysical Research-Atmospheres*, *113*(D9).

Zander, R. (1976), HIGH-RESOLUTION INFRARED SOLAR OBSERVATIONS BY BALLOON, *Infrared Physics*, *16*(1-2), 125-127.

Zander, R., G. Roland, and L. Delbouille (1977), CONFIRMING PRESENCE OF HYDROFLUORIC-ACID IN UPPER STRATOSPHERE, *Geophysical Research Letters*, *4*(3), 117-120.

Zander, R., C. P. Rinsland, C. B. Farmer, et al. (1986), Observation of Several Chlorine Nitrate ($ClONO_2$) Bands in Stratospheric Infrared-Spectra, *Geophysical Research Letters*, *13*(8), 757-760.

Zhao, Y., Y. Kondo, F. J. Murcray, et al. (2000), Seasonal variations of HCN over northern Japan measured by ground-based infrared solar spectroscopy, *Geophysical Research Letters*, *27*(14), 2085-2088.

Zhao, Y., K. Strong, Y. Kondo, et al. (2002), Spectroscopic measurements of tropospheric CO, C_2H_6, C_2H_2,and HCN in northern Japan, *Journal Of Geophysical Research-Atmospheres*, *107*(D18), -.

22

Earth Scientist's Guide to Discrete-Time Power Spectrum Analysis

Bjørn-Gustaf J. Brooks
Center for Climatic Research, University of Wisconsin-Madison
U.S.A.

1. Introduction

Fourier methods are commonplace in the Earth Sciences and have greatly enhanced our understanding and forecast capabilities for cyclical phenomena that recur on interannual (e.g. El Niño, Pacific Decadal Oscillation) to millennial scales (e.g. Milankovitch cycles). Nowadays most low level programming languages (C, Fortran) have math libraries that include the fast Fourier transform algorithm (FFT) and nearly all abstract programs (Python, Octave/Matlab, IDL, R) provide an array of Fourier functions for scripting sophisticated signal processing routines. Whether your interest as a practicing Earth scientist is in Fourier transformation for efficient data manipulation, or for problems where the Fourier transform or its power spectrum is needed for direct analysis, you have probably found no shortage of relevant literature. Nonetheless, you may also have found some difficulty in making sense of *which* Fourier methods to implement for your particular analysis idea, and *how* to appropriately apply them. This chapter will serve you as a basic guide for unraveling some of the complicated implementations of discrete-time power spectrum analysis using direct language and supplementary Matlab/Octave routines using both observed and modeled data.

This chapter assumes that you have a certain task to accomplish, and therefore it is designed to teach you how to set up an approach appropriate for Fourier analysis, and also to advise you of potential pitfalls and limitations in Fourier analysis. The reader need not have prior exposure to signal processing methodologies, but should have a solid base in mathematics, probability theory and more importantly the issues related to your analysis data so that the significances of cycles within your data can be rationally interpreted. If you find yourself lost by the terminology I recommend you familiarize yourself with basic treatments of discrete-time systems, for example Press et al. (1992) or Cadzow (1973).

The body of this chapter is split into three sections, *Preprocessing data*, *Single Series Spectrum Analysis*, and *Multiseries Spectral Analysis*. Step-by-step examples are given on the analysis of a variety of freely accessible earth science datasets covering atmospheric science, biosphere-atmosphere carbon cycling, climate modeling, and paleodiversity as well as some example implementations of Markov chain Monte Carlo routines for computing statistical significances. Each section contains direct explanations with ready to deploy example code that you are free to use for your own investigations. Supplementary code can be accessed online from ftp://ftp.climatemodeling.org/pub/esg/.

2. Preprocessing data: Identifying the pertinent information

An often overlooked aspect in the early stages of data analysis and exploration is the preliminary processing of time series. No matter what your particular topic is it is often desirable to filter measurements that confound your analysis or do not contribute useful information. A thorough exploration of the many methods for reducing noise and removing trends prior to computing the Fourier power spectrum may seem like a lot of work, but it will be effort well-rewarded because it will increase confidence in your results. Even data that has already been heavily processed, for example satellite products like leaf area index (LAI), may still require treatment beyond, or in stead of, monthly averaging, seasonalizing, or annualizing. This section discusses filtering, subsetting, and detrending topics and provides example code for deploying this in your own data.

Within any dataset there are data that contribute useful information as well as data that do not, or even confound your analysis. One example comes from the study of atmospheric tracers of photosynthesis such as carbon dioxide or carbonyl sulfide. Atmospheric CO_2 and OCS have diurnal and annual cycles that are strongly driven by biotic processes, primarily reflecting the uptake of atmospheric CO_2 by plant photosynthesis and CO_2 release through heterotrophic and autotrophic respiration. Measurements of these tracers should show stronger annual or diurnal cycles during years when precipitation is substantially influenced by El Niño. But strong cycles may be difficult to observe unless the observations are preprocessed or filtered. This is because a series of tower measurements of atmospheric trace gasses will include both locally representative observations made when turbulent mixing is low and winds are calm, and regionally representative observations made when the boundary layer is well-mixed. If your goal were to observe smaller influences caused by interannual ENSO cycles among much larger biotic influences you may have to detrend from a fitted polynomial (section 2.1) filter the noise (section 2.2), or subset the data to remove locally representative measurements (section 2.3), which would result in non-uniform series that can be analyzed using Lomb-Scargle (section 3.2).

2.1 Detrending

Within geological time series, from ice cores to cyclostratigraphy, it is not uncommon to find long term and systematic trends. If your objective is to understand something other than these trends it will be necessary to detrend the data. An example of the importance of detrending was observed by Cornette (2007) who showed that the prominence of a 62-million-year cycle in extinctions recorded in the geologic rock record was strongly affected by detrending the original diversity time series. This is because the most significant variability was caused by a large non-linear biodiversification trend (Figure 1A) representing three major phases of biodiversification. Extinction cycles were only apparent in the Fourier power spectrum after detrending the time series. Likewise if a very long lived and significant harmonic or trend exists in your data detrending should be considered. Pseudo-code showing how this fossil diversity time series was detrended appears in Procedure 1 (and also in supplementary code dftps.m).

2.2 Noise and smoothing filters

Noise adds considerable clutter that can confound Fourier power spectrum (PS) analysis and reduce the strength of important cycles. Noise is given a broad definition here to refer to variability in any series that cannot be resolved or directly tied to a predictable physical cause.

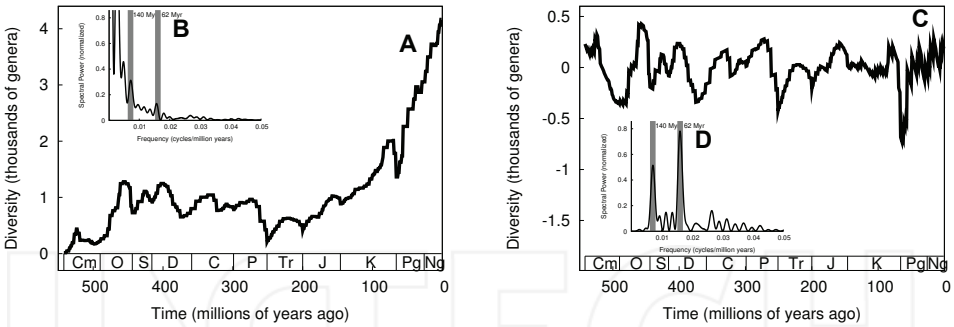


Fig. 1. Detrending effects. Subplots A and B show diversity without detrending– C and D are detrended. A) Time series of Sepkoski's marine fossil diversity covering the past 542 million years, as reproduced from Rohde & Muller (2005). B) The Fourier power spectrum (PS) of that series. C) Time series of diversity residuals obtained from A by detrending by a cubic polynomial (*cf.* Cornette, 2007). D) The PS of residual diversity showing significant cycles corresponding to the 62 and 140 million year cycles. These show that in non-detrended fossil diversity nearly all the power of the 62 and 140 million year cycles is subsumed by the larger trend in A.

Procedure 1 General procedure for detrending a series by a polynomial curve of order `N`. The diversity data in Figure 1A was detrended by a polynomial of `N` = 3 to produce Figure 1C. Series x is comprised of *nterms* measured at increments of t.

input: data series, $\{(t_k, x_k)\}_{k=1}^{nterms}$
output: data series of residuals, $\{r_k\}_{k=1}^{nterms}$

```
p = poly_fit(t,x,N)
pv = poly_val(p,t)
r = x - pv
```

It is also important to keep in mind that noise may simply reflect events that are not well represented by the particular measurement system or point of observation being used. Noise generally has a limited systematic effect and should be filtered if at all possible.

A consideration when filtering noise is to choose a filter appropriate to the type of noise within your data (*i.e.* white, red, blue, gray). White noise can give the Fourier power spectrum plots (like those in Figure 1B and 1D) broad spectrum noise, which appears as peaks (all of similar magnitude) dispersed across a range of frequencies in the power spectrum. Least squares filters including Savitzky-Golay are often used to reduce this kind of broad spectrum noise. In Matlab/Octave such filtering is achieved by calling the `sgolayfilt` function. In IDL calling `savgol` in conjunction with `convol` achieves the same purpose.

Confounding variability can also come from red noise, which is not uncommon in paleoclimate datasets and some long term atmospheric and ecosystem records. Power spectrum plots of series containing red noise have characteristic slopes that diminish with increasing frequency (toward the right of the spectral plot), resulting in power spectrum plots that appear cluttered on left near longer cycles (*i.e.* where red wavelengths of the

visible spectrum would appear). Red noise can be caused by quasi-stationary processes, for example a precipitation record with repeated and prolonged multi-year droughts that cause precipitation to fall below the mean. Red noise can be smoothed by filters that use lower order polynomials, for example in IDL using `savgol` with a degree of 2.
Blue noise causes clutter around shorter cycles toward the right of the power spectrum, and can be filtered using higher order polynomials. Grey noise is identified by its bimodal noise peaks at both the 'red' and the 'blue' end of the power spectrum. Several test cases are provided in the supplementary Matlab/Octave code noise_filt.m, and you should use them to test the effectiveness of the Savitzky-Golay filter using various filter configurations.
If however you find that it is not possible to effectively filter noise without reducing the information content of your time series, you may instead evaluate the power spectrum of the unfiltered data against a stringent statistical test such as a Markov chain Monte Carlo test (see the random walk test for red noise in section 3.3).

2.3 Subsetting and subsetting filters

Where noise filtering is not tenable or where it is desirable to search for the sources of cycles you should consider partitioning your time series based on the relative contributions of various groups (*subsetting*) or applying a *subsetting filter* to reject entire observations. Here I use the term subsetting to indicate partitioning of each value, as opposed to subset filtering, which I use to indicate rejecting entire observations that exceed some cutoff. Subset filtering would be relevant for analyses such as in atmospheric observation where it is not appropriate to smooth a biased measurement because ties to the actual observations are required.
A common issue with observational data is that not all measurements communicate useful information about the processes you are interested in, therefore it may be necessary to partition the complete set of observations in some way and to compare the significances of cycles between subsets. One example subsetting strategy is to partition each value in the series based on the relative contributions from different groups. Given that you also possessed meta-data, *i.e.* the annotations describing each specimen, you could begin to address some very interesting questions about the nature of cycle. As it happens, the diversity values used to construct Figure 1A can be partitioned by groups such as deep water fossils vs. shallow water fossils, or hard-shelled fossils vs. all others. Now we might ask, 'Is any one group of organisms responsible for the cycles in fossil diversity seen in Figure 1D?'
Figure 2 uses a similar but more comprehensive dataset of fossil diversity downloaded from the Paleobiology Database (PD, 2008) to subset the total diversity into multiple curves representing the contributions from each phylogenetic group, such as mammals, birds, and reptiles. Figure 2A compares the diversity curve of several dominant shelled fossil animals (gastropods, bivalves, and articulate brachiopods or GBA) against all other phyla. Figure 2B shows the Fourier power spectra of those same subsets and suggests that the 62 million year cycle in this dataset is largely driven by the GBA group.
On the other hand meta-data annotating each observation may not be available to you, and you may have to create a subsetting filter. Here you would make a decision, for example using a statistical basis to reject observations from a subset. For example, given a dataset of wind speed measurements from a meteorological station we might use a subset filter to reject all measurements of wind speeds less than 4 meters per second, leaving us with a new subset of unevenly spaced measurements. Non-uniformly sampled data cannot be analyzed using

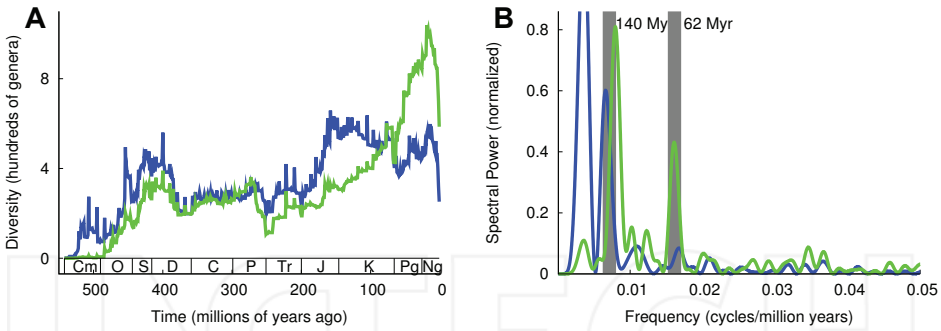


Fig. 2. Using metadata to subset by groups. A) shows two time series of fossil abundances derived from the Paleobiology Database (paleodb.org) and ranging from 542 million years ago to present. The first subset in green is comprised of gastropod, bivalve, and articulate brachiopod fossils (GBA). Their complement (non-members of that group) appear in black. B) shows the Fourier power spectrum of those same curves when detrended. Notice that a rather significant peak appears in the GBA group near the 62-million-year cycle frequency, which is not present in the non-GBA group. This shows that by using the annotations of your data it is possible to investigate the sources of cycles.

standard Fourier decomposition, but can be processed using the Lomb-Scargle method, which will be addressed later in Section 3.2).
Subsetting filters utilize *a priori* knowledge to reject outliers, which means you decide based on your analysis of the dataset what an acceptable cutoff should be. One way to avoid criticisms on the subjectivity of your cutoff would be to process your data using a series of cutoffs that range from very inclusive to very specific, and to do spectral analysis on each differently filtered subset, and compare.

2.4 Normalizing

A final but critical note on preprocessing is about the importance of normalizing your time series to a standard deviation of 1 prior to Fourier decomposition. Whether you apply filters to your data or not it is always necessary to normalize the variance of your original time series to make it comparable to other series. This is simple to do. Divide each observation by the standard deviation of the time series. See lines 21–22 in supplementary code dftps.m.

3. Single series spectrum analysis

Once your series of uniformly sampled values has been normalized (perhaps even filtered or subset) Discrete-time Fourier transform power spectrum analysis (DFTPS) can be used as a powerful tool for determining the relative strength of cycles within your time (or spatial) series. You should be aware that Fourier transformation works by decomposing a series into its complex conjugate of *real* ($\Re$) and *imaginary* ($\Im$) signal parts. The spectral power is computed as the square of the signal power from $\Re$, which produces a series of amplitudes across a range of frequencies that describe the strength of cycles. $\Im$ is used to determine the phase of the cycles, which can tell you when in your time series the peaks of cycles of different frequencies should occur. These methods can be applied to any series of uniformly sampled

data. If your interest lies in the theory behind spectral analysis of discrete time systems a good place to start is Cadzow (1973) or Bloomfield (1976). More sophisticated implementations of signal processing designed for specific purposes (*e.g.* Brooks, 2009) are often useful when computational power is limited. But if you have access to a high performance machine or cluster then you will probably be able to solve big data problems easier using distributed memory parallelism.

3.1 Computing the power spectrum

As opposed to continuous analog signals, discrete time series are segmented and require a special step in treatment before transformation by the fast Fourier transform algorithm (FFT). You must pad your series of normalized values with zeros so that the transformed signal of your discretely sampled values is sufficiently long to allow for measurement of its peaks across a range of frequencies (see lines 18-19 of dftps.m).

Zero padding is particularly important when examining cycles with longer periods. Consider that you have 100 years of monthly means from a model and you want to know how well the model is able to reproduce interannual variability, caused for example by El Niño–Southern Oscillation (ENSO, 2-7 yr. period) or the Pacific Decadal Oscillation (PDO, 20-30 yr. period). Despite having cycle lengths that are roughly 20 years apart (ENSO: ∼5 years, PDO:∼25 years), these cycles will show up very near each other in frequency on the power spectrum plot. Since frequency is the inverse of time $f_{ENSO} = \frac{1}{12}$ (months) $\times 5$ (years) $= 0.0167$, which is near $f_{PDO} = \frac{1}{12} \times 25 = 0.0033$. In order to resolve such small frequency differences between longer cycles it is necessary to pad the signal with zeros. To better understand this you might try commenting-out line 18 of dftps.m and substituting it with: `lps=2^(8);` (you will also have to adjust the y-axis range). The subsequent plot will be very coarsely sampled on the left toward longer frequencies because you reduced the padding. Zero padding will not change the amplitude of your peaks but it will help you better resolve the frequency and timing of cycles. Fortunately Matlab, Octave, and Python provide handy methods for padding the series and computing the Fourier transform in one step using the `fft` function (Procedure 2). The computation of uniform increments of f to plot your power spectrum peaks against is also straight-forward (see line 35 of dftps.m).

Procedure 2 General FFT procedure for Matlab, Octave, Python. Given the series `x` of discrete values x_k the function `fft` is used to compute its complex conjugate meanwhile 'stretching' the signal to length `ne`, provided that `ne` > *nterms*.

input: time series, $\{x_k\}_{k=1}^{nterms}$
output: x_ft (complex conjugate with $\Re$ & $\Im$ parts), sp (spectral power), f (frequency), p (period)

```
x_n = x/std_dev(x)          # Normalize to variance of 1
x_ft = fft(x_n,ne)          # Transform and pad to length ne
sp = x_ft .* conj(x_ft)     # Compute spectral density curve. sp is
                            # complex and contains real and imag parts
f = flt_intgen(ne)/ne       # Generate sampling frequencies for sp
p = 1/f                     # Cycle periods in original units
```

Remember that the FFT algorithm will give you complex output: a real part ($\Re$), which you will use to make your power spectrum plot, and an imaginary part ($\Im$), which you might use to determine the phase-synchronization. The imaginary part is not to be overlooked. In order to determine causality relationships for example, the phase data ($\Im$) of the independent and dependent cycles should coincide in a logical way. Computation of the phase is discussed in Section 3.4 and Procedure 3.

3.2 Non-uniformly sampled data

The FFT algorithm requires uniform time steps, and cannot be used to analyze non-uniform series such as in Figure 3A. Fortunately there are alternative ways for dealing with unevenly sampled data including the Lomb-Scargle (LS) method. The LS approach is a common implementation, but you should be aware that estimating the period of cycles from non-uniformly sampled data is not a trivial issue, and precision can vary depending on the frequency estimators used. An in-depth treatment of spectral analysis for unevenly sampled data appears in Press et al. (1992), which should be referred to for a detailed explanation.

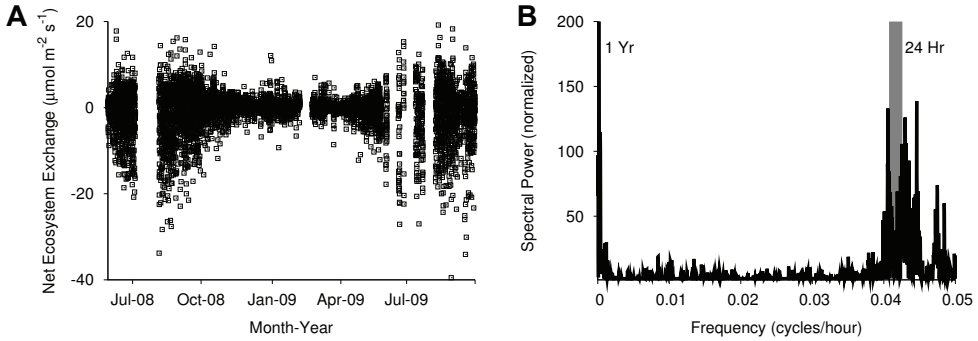


Fig. 3. Series that are not uniformly sampled can be transformed using Lomb-Scargle. A) Non-uniformly sampled time series of CO_2 exchange between the atmosphere and biosphere (Net Ecosystem Exchange) measured at the Park Falls, Wisconsin tall for June, 2008-Oct., 2009 (UW-Madison, 2010). B) The spectral power of that data when the LS method is used.

LS implementations (such as the example in supplementary code lsps.m) work by computing the spectral power across an increasing set of frequencies (usually controlled by a variable called `ofac`), up to a frequency limit (`hifac`). The resulting power spectrum represents an oversampling of the data (suitable results can be obtained by oversampling rates where `ofac` ≥ 4). `hifac`, which is related to the Nyquist frequency, sets the limit of frequencies to be explored. If, for example, you are not interested in frequencies larger than $f = 0.05$, then you should conserve computational time by setting `hifac` = 0.1.

Some considerations should be kept in mind when specifying your `ofac` resampling rate. If you are investigating two narrowly spaced cycles and your series has relatively few broad gaps, then you would probably be justified in using higher values for `ofac` that will allow you to resolve closely spaced cycles. On the other hand if your series is subject to substantial gaps, large values for `ofac` are probably not a good idea.

Figure 3B is a power spectrum plot, the same as in Figures 1B, 1D and 2B, except that this one was computed using the LS method. Figure 3B illustrates an important issue common to

all power spectra: the dispersion of peaks reflecting a cycle with a period that slightly varies throughout the time series. Minor variations in shorter frequencies, such as those to the right ($f > 0.03$) in Figure 3B, can result in either a broader peak or multiple peaks near the frequency of interest. The primary diurnal cycle should occur at $f = 0.0416$ in Figure 3B. Instead there are actually three peaks of nearly equal amplitude, which indicates that the timing of the diurnal NEE cycle pictured in Figure 3A varies throughout the time series, probably due to changes in daylight length throughout the year as well as synoptic frontal passages and other variability that can affect the exchange of carbon (NEE). For example, small shifts in timing of peak NEE by 1 hour in either direction (*e.g.* 23-hour/25-hour spacing) would result in large spreads in frequency ($f = 0.04$ to 0.045).

In spectral plots longer cycles have fewer opportunities to occur in a limited dataset and therefore the certainty about their significance is generally less than a shorter cycle of similar amplitude. The next section will discuss a method for dealing with uncertainty in the significance of longer-term cycles using Monte Carlo methods.

3.3 Estimating uncertainty using Markov chain Monte Carlo trials

By this point in the chapter you already know how to identify whether or not a cyclical pattern exists within your series. If you have already used a script to produce a power spectrum plot (*e.g.* dftps.m, lsps.m), you probably asked yourself: 'How do I know if these cycles are significant?' There are a variety of ways to estimate the statistical significance of cycles within your data and many of them involve Monte Carlo trials. Markov chain Monte Carlo methods (MCMC) are a widely used class of algorithms that iteratively and randomly resample (permute) time series, and can be used to test the significances of cycles by comparing them to randomly derived cycles. MCMC tests allow you to compare the original time series to many randomized versions of that same data, and to examine whether or not cycles of equal or greater magnitude exist in the randomized versions. It is important to understand, however, that Monte Carlo significances do not rule-out bias in your sampling protocol. Significance tests of this kind can only inform you about how likely a cycle is to occur given a particular collection of values.

As a starting point for understanding Markov chain Monte Carlo methods I present an example called random walk trials, sometimes referred to as random step or drunkard's walk. Random walk trials can be used to determine statistical significances and are particularly useful for evaluating low frequency cycles. Many random walk implementations exist, and here we will keep things simple by using just one implementation, which appears in supplementary code dftps_mcmc.m.

If you have experimented with the dftps_mcmc.m routine you may have wondered how it computes significances. The significance calculation used in dftps_mcmc.m is described as follows: For a given peak of height h at a frequency f, the significance of a peak can be computed as the fraction, p, of N randomly generated sequences of the original series, from which the spectrum at f exceeds h. Here you might report the significance of a cycle as the fraction of trials in which the height of the peak in the original series exceeded the height from the randomly generated series. Since random walk trials work by randomly permuting the original series before Fourier decomposition, 10,000 MCMC trials would represent the null hypothesis that cycles within your data are merely coincidental and can be reproduced given sufficient randomization.

Uncertainty can be represented not only as a percentage but also in your power spectrum plot by a line, for example representing the mean spectral power of the MCMC iterations. Supplementary code dftps_mcmc.m outputs both the significance estimates and a power spectrum plot with the MCMC random walk significance curve. Note that all peaks that occur at or below the random walk significance curve should be interpreted as insignificant, whereas peaks with amplitudes greater than the curve are significant to the degree indicated by the statistics of the random walk trials.

3.4 Phase synchronization and phase shifted cycles

Two cycles (from different time series) with exactly the same period might seem to be related. However, their phases could indicate that their peaks in the original time series are asynchronous, which could indicate a lag in the cause-and-effect relationship, or they may not be related at all. Let's say that you have two time series of daily means, one for air temperature and another for the temperature of a nearby lake. Both records will have strong annual cycles, near $f = \frac{1}{365} = 0.0027$, but because of the specific heat capacity of water, the peaks in lake temperature will lag behind air temperature. If you examined the imaginary part ($\Im$) of the Fourier transform at $f = 0.0027$ you would see the phase shift of the cycle in the range $-\pi : \pi$ radians. (For example find the index number of the element in `f` on line 29 of dftps.m that is closest to the frequency 0.0027 and then find the corresponding element by its index in the imaginary part of `xn_fft` on line 19 of dftps.m) A good way of visualizing this is to plot the original series along with a sine wave corresponding to the cycle of interest with the appropriate phase shift as determined from the imaginary part of the Fourier transform. This is readily done in Matlab/Octave using `angle`, and Python using `phase`, and IDL using `atan`, which compute the phase angle shift in radians ($[-\pi : \pi]$). Pseudocode describing the construction of the phase shifted sine wave appears in Procedure 3. Note also that this can be used to extrapolate the continuation of a cycle beyond the observational data, which might be useful when predicting future cycle peaks.

An example that should give us no trouble comes from modeled temperature data from the Parallel Climate Model (Washington et al., 2000). Because the complex transform corresponds perfectly to the magnitude and phase, we need only to convert the real and imaginary parts back into physical coordinates according to Procedure 3 in order to produce Figure 4.

Procedure 3 General procedure for determining the phase shift and computing the sine wave corresponding to the period and phase of the cycle of interest. This procedure uses the imaginary data at index `k` to compute the phase shift (`ps`) for constructing the sine wave.

input: $\{x\_ft\}_{i=1}^{nterms}$ (complex conjugate); p (period); t (time increments in physical units); a (sine wave amplitude in physical units)
output: ps_k, p_k

```
ps = atan(x_ft[k])
sw = a * sin(2 * (pi/p) * t + ps)  # the phase shifted sine wave
```

4. Multiseries spectrum analysis

Cause and effect relationships are common, but sometimes difficult to verify. One way to test whether cycles with the same period from two different time series are related is

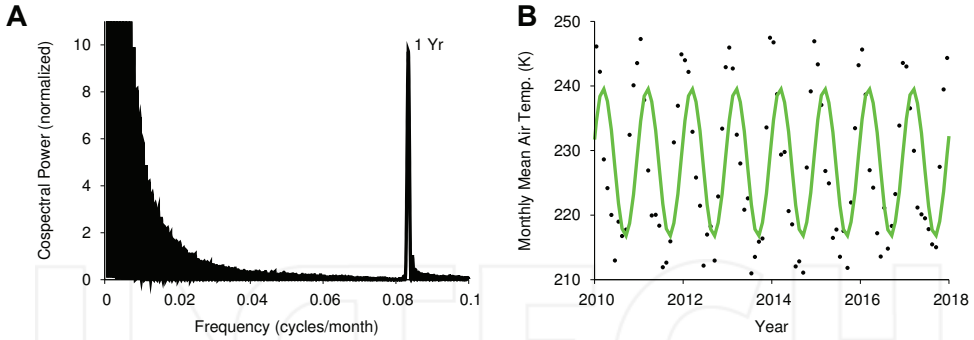


Fig. 4. Power spectrum plot (A) and time series with phase shifted wave (B). The phase shift of the annual cycle in A is applied to the green sine wave in B, which shows where the peaks and troughs of the annual cycle should occur according to their average occurrence over 100 years of model data.

through multiseries spectrum analysis. Cross-spectrum analysis is a common implementation used to compare how well cycles from two different series covary in period and phase synchronization. Think of cross-spectrum analysis as being the Fourier transform equivalent of cross correlation. For example, you would expect that net radiation measured at a meteorological station would correlate well with air temperature. Cross-spectrum analysis should reveal diurnal and annual cycles that are strong and nearly synchronized.

As opposed to the results of single series spectrum analysis the cross-spectrum density represents the covariance between two series. Cross-spectrum functions in abstract languages like Matlab typically compute spectrum density using the Fourier method or Welch's method, and their power spectra can be interpreted in the same way, except that the phase data takes on a heightened importance. Figure 5 was produced using supplementary code csps.m again using model data for years 2000-2099 from the Parallel Climate Model (Washington et al., 2000). Figure 5A shows the cross-spectrum density of soil moisture and precipitation for one arctic land surface grid cell, which signals a very strong correlation in both series of the annual cycle. Figure 5B represents the phase shift across all frequencies. The frequency of the annual cycle is located by the vertical gray band, which indicates that the two annual cycles are nearly π radians (180°) out of phase, meaning that precipitation is high when soil moisture is low. This seems surprising at first, but remember that in the arctic a majority of the precipitation comes as snow, which has a delayed release into soils or the soils may be frozen (permafrost) for most of the year.

5. Summary

There are many worthwhile uses for discrete-time Fourier power spectrum analysis methods, as shown in this chapter including: 1) correlating the amplitude and phase of similar cycles from different time series, 2) estimating the statistical significance of cycles, and 3) investigating the sources of cycles using subsets of the complete set of data. However, by now you have probably also realized a few of the limitations of power spectrum analysis such as: 1) very long cycles are difficult to detect given time series of limited length, 2) inappropriately preprocessing/filtering your data prior to Fourier decomposition

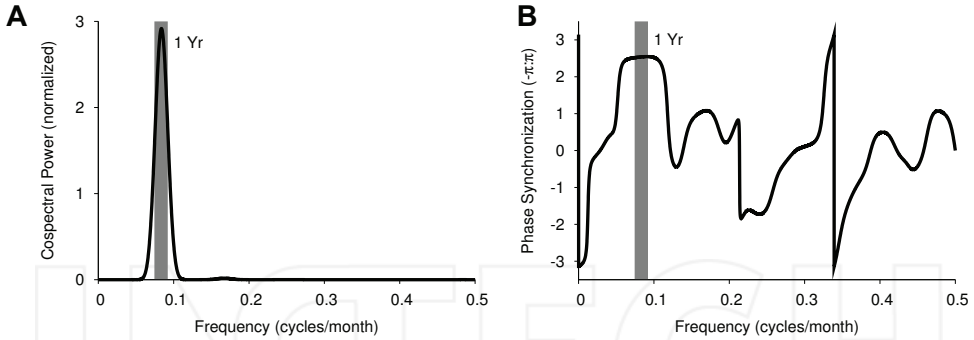


Fig. 5. Cospectrum and phase synchronization. A) shows the total power of the cospectrum representing a strong cycle in both soil moisture and precipitation. The gray band locates the annual cycle. B) shows that the phases of the annual cycles are nearly 180° (π radians) out of phase, which underscores the importance of considering the phase data in cross-spectrum analysis.

can lead to erroneous conclusions, 3) choosing a noise filter is subjective and can lead to different results, and 4) failing to specify the correct null hypothesis to test against possible cause-and-effect relationships between cycles (*i.e.* type iii error) can lead to false conclusions. Understanding the strengths and weaknesses of Fourier power spectrum analysis as you have been introduced to here will help you to place the correct emphasis on signal processing results in your study. If you do find room for such analysis in your work feel free to implement the supplementary code presented here without reservation and in any way you like.

6. References

Bloomfield, P. (1976). *Fourier analysis of time series: An introduction*, John Wiley & Sons Inc.

Brooks, B.-G. J. (2009). *Applying Wavelet and Fourier Transform Analysis to Large Geophysical Datasets*, Vol. 5545/2009, Springer Berlin, Heidelberg, pp. 426–434. doi: 10.1007/978-3-642-01973-9_47.

Cadzow, J. A. (1973). *Discrete–Time Systems: An Introduction with Interdisciplinary Applications*, Prentice-Hall, Inc.

Cornette, J. L. (2007). Gauss-vaníček and fourier transform spectral analyses of marine diversity, *Computing in Science and Engineering* 9(4): 61–63. doi: 10.1109/MCSE.2007.76.

PD (2008). The Paleobiology Database, http://www.paleodb.org/. Data were downloaded on 8 August, 2008. Ichnotaxa and open nomenclature genera were excluded. The collections reflect 8,821 publications. Lithological categories can be found online at http://paleodb.org/public/tips/lithtips.html.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (eds) (1992). *Spectral Analysis of Unevenly Sampled Data*, Cambridge University Press, 2 edition, Cambridge, p. 992.

Rohde, R. A. & Muller, R. A. (2005). Cycles in fossil diversity, *Nature* 434: supplementary material 1. http://www.nature.com/ nature/journal /v434/n7030/extref /nature 03339-s3.xls.

UW-Madison (2010). University of Wisconsin Park Falls tall tower NEE data, Website. Data were downloaded on 24 January, 2011 from http://flux.aos.wisc.edu/data/wlef/flux/2008/prefnee_2008.txt and http://flux.aos.wisc.edu/data/wlef/flux/2009/prefnee_2009.txt.

Washington, W. M., Weatherly, J. W., Meehl, G. A., Semtner, Jr., A. J., Bettge, T. W., Craig, A. P., Strand, Jr., W. G., Arblaster, J., Wayland, V. B., James, R. & Zhang, Y. (2000). Parallel climate model (PCM) control and transient simulations, *Climate Dynamics* 16: 755–774. doi: 10.1007/s003820000079.

23

Imaging Fourier Transform Spectroscopy for Astronomy

Laurent Drissen[1], Anne-Pier Bernier[1], Maxime Charlebois[1], Alexandre Alarie[1], Frédéric Grandmont[2] and Julie Mandar[1,2]
[1]*Université Laval, Québec*
[2]*ABB Inc., Québec*
Canada

1. Introduction

Unlike most other scientists, astronomers do not have direct access to the objects they are studying. As a matter of fact, except for a few cases (solar wind and neutrinos, lunar samples, meteorites, cosmic rays), all information originating from the Universe is transmitted to us by light. Because it has the ability to interact with matter, it keeps a lasting impression of the environment where it was born or has had interaction with. One of the greatest challenges in astronomy is thus to extract, using methods ever more clever, the maximum information from photons that crossed through space over thousands, if not billions, of years. A giant step forward was taken nearly 400 years ago when Galileo Galilei pointed a modest telescope towards the sky. Technological developments have since considerably increased the dimension and visual acuity of telescopes (segmented mirrors, adaptive optics), the quantum efficiency of detectors (often close to 90%), the detectable wavelength range (from radio waves to gamma rays), as well as all the specific measurement techniques such as photometry, spectroscopy and polarimetry.

There are basically two traditional approaches to obtaining spectral information on extended astrophysical objects: narrow-band imagery and integral field dispersive spectroscopy. Imagery with filters allows the observer to map a target in selected wavelength ranges and to extract the required physical information by comparing the relative flux of the sources in these bands. This technique is used to obtain color-magnitude diagrams of star clusters or resolved galaxies (SLOAN ugriz broad-band filters for example), or to map abundance gradients in nebulae or gas-rich galaxies (using narrow-band interference filters centered on specific emission lines such as Hα 656.3 nm, [NII] 658.4 nm or [OIII] 500.7 nm). Images of the targets in the different band passes must be obtained one after the other with a CCD detector, rejecting each time all photons excluded by the selected filters (up to 99.8%). Moreover, narrow-band imagery does not provide a high enough spectral resolution to determine the gas velocity.

Dispersive spectroscopy with slits allows a much finer spectral resolution ($R = \lambda/\Delta\lambda \sim 10^3 - 10^5$) at the expense of spatial information on the targets. Extensively used since the mid-19th century to obtain the spectrum of individual stars or small slices of extended objects, dispersive spectroscopy has been transformed by the advent of multi-object spectrographs

(MOS) in the 1990's: multiple slitlets or optical fibers are positioned at the location of the targets in a wide field of view, the light of which is then sent to a disperser and recorded on a CCD. Major breakthroughs have also been obtained in astronomical instrumentation over the last decade by combining imagery and spectroscopy into a single experimental observation technique that produces cubes of data. These are typically referred to as Integral Field Unit (IFU) instruments. Given the limitations of modern array detectors, different instrument concepts convey different trades that each enhances the possibility of discovery for a given science program category. The ability to cover a greater area than the classical spectrometer slit has often been the driving motivation behind these 3-D instrument developments (Monnet 2009). The use of MOS [Sloan digital sky survey (Stoughton et al. 2002) or 2dF (Colles et al. 2001)] and integral field spectrographs [GMOS-IFU on Gemini (Allington-Smith et al. 2002) , or VIMOS-IFU (Sanchez et al. 2004)] on large telescopes has revolutionized data collection by allowing respectively to obtain spectra of a large number (up to a few hundred) of objects dispersed in a large field or to spatially sample relatively small (of the order of 10 arcseconds) objects. An integral field spectrograph allowing observations across a relatively large field field of view (41 x 33 arcseconds at a spectral resolution R ~ 1000), SAURON (Bacon et al. 2001), has revolutionized the study of late-type galaxies, and a similar, but much more complex, instrument, MUSE, is being built for the VLT (Bacon et al. 2010).

The vast majority of imaging spectrometers used on telescope to date however build on dispersive approaches which must "sacrifice" detector pixels to retrieve the spectral content instead of scene elements. Typical ratio of distinct scene elements (pixels) to available detector pixels is on the order of 1/1000. A pure imager would have a ratio of 1 but its spectral capability is limited by the width of the filters used to select specific spectral wavebands.

A variety of concepts now propose different balances between field and spectral elements in terms of both coverage and resolution in order to make the best use of their detector pixels. By using non-dispersive approaches such as interferometric ones, one can hope to use all detector pixels for imagery thus prioritizing spatial coverage, resolution or both. The compromise then usually shifts to the spectral or temporal side as spectra for scene elements must be acquired in the time domain using multiple exposures. The most familiar instrument of this kind is probably the Faby-Perot interferometer in which spectral slices of the final data cube are acquired one by one while mechanically changing the central wavelength transmitted by the etalon. Although unrivaled for field coverage in the IFU group, this instrument faces an important waveband width limitation imposed by the multiple orders transmitted by the etalon, which must be filtered out optically to permit unambiguous retrieval of the spectral information. Fabry-Perot are mostly used to obtain high resolution (R ~ 20 000) spectra of individual lines such as Hα, [OIII] 500.7 nm or the [SII] doublet at 671.7, 673.1 nm (Hernandez et al. 2008, Lagrois & Joncas 2010).

We present in this paper a very brief historical review, as well as the most recent developments, of another approach , imaging Fourier transform spectroscopy (FTS), which has been given a strong boost during the past decade, mostly because of enormous improvements in digital imaging capabilities, computer power and servo control systems. A large number of research programs would benefit from an instrument capable of simultaneously obtaining spatially resolved, high quality spectra on extended areas (of the order of 10 arcminutes) and with a resolution up to R ~ 10^4. Imaging Fourier transform

spectroscopy is very promising in that regard. Based on the principle of the Michelson interferometer, Fourier Transform Spectrometers (FTS) are extremely efficient because all photons are collected and analyzed. Moreover, by using appropriate optical configurations, it is possible to transform the traditional one-pixel FTS into a truly integral field spectrometer. In this chapter, we will :

a. Present a brief historical review of the use of FT spectrographs in astronomy;
b. Introduce the concept of an imaging FTS aimed at observing astronomical sources;
c. Use our experience with our instrument, SpIOMM, to illustrate the technical challenges that must be overcome to ensure the efficiency of such an instrument;
d. Present some of the most interesting scientific results obtained with SpIOMM ;
e. Discuss future developments of wide-field imaging FTS on large ground-based and space-based telescopes.

2. A very brief history of the imaging FTS in astronomy

Although FTS are most widely used for military and chemical applications, they have also been very successful in planetary exploration (on board the Mariner, Voyager and more recently Cassini spacecrafts; Flasar et al. 2004) and in the analysis of the Earth's atmosphere (a recent example being the ACE-FTS instrument on board the SCISAT-1 remote sensing Canadian satellite; Bernath et al. 2005). The use of FTS in astronomy is not widespread, mostly because of the technical difficulties in building such instruments, but some examples need to be mentioned. The FTS at Kitt Peak's Mayall telescope was used the 1970's and 1980's to provide exquisite spectra of late-type stars (Scoville et al. 1979, Ridgway et al. 1984). At the Canada-France-Hawaii Telescope (CFHT), the high-resolution FTS was widely used on a large variety of planetary and stellar programs (Chalabaev & Maillard 1985, Maillard et al. 1987). Made able to work on an imaging mode in the early 1990's, the CFHT FTS was renamed BEAR (Maillard & Simons 1992); it provided integral field spectra of a variety of objects such as planetary nebulae, massive star clusters and star-forming regions in a 24 arcsecond field of view (Paumard et al. 2004). Other examples include the FTS built by D. Naylor (University of Lethbridge) on the James Clerk Maxwell submillimeter telescope (Naylor et al. 2004, Friesen et al. 2005), an FTS for SPIRE, one of three instruments to fly on ESA's Herschel Space Observatory (Naylor et al. 2010, White et al. 2010), a far-infrared FTS on the Japanese satellite AKARI, a mid-IR FTS (CIRS) on the Cassini spacecraft and for the nar-IR, PFS on Mars Express with a copy on Venus Express. We would also like to mention another imaging FTS prototype, working in the visible part of the spectrum that was built at the Laurence Livermore Lab and tested at the 3.5-m Apache Point Observatory telescope (Wurtz et al. 2002a, 2002b), in which one of us (FG) was involved, but which development ceased a few years ago. The development of this instrument was a major step forward to demonstrate the ability of an imaging FTS to acquire hyperspectral images in the visible band. The advantages and disadvantages of the imaging FTS technique, as well as the relative merit of different approaches to 3-D imagery are discussed by Ridgway & Brault (1984) and, more recently, by Bennett (2000).

The development of imaging FTS in astronomy was given a strong incentive during the early definition phases of the NGST (now known as the James Webb Space Telescope, a 6.5-m segmented mirror infrared telescope to be launched at the Sun-Earth L2 point in 2014): astronomers supported by the three participating space agencies (NASA, ESA and the Canadian Space Agency) presented studies of imaging FTS at the NGST Instrumentation

meeting in Hyannis in 1999 september (Graham 2000, Morris et al. 2000, Posselt et al. 2000) . None of these concepts however were included in the final instrument suite of the telescope. More recently, Boulanger et al (2008) proposed the design of a 1.2-m space telescope, H2EX, equipped with a wide-field imaging FTS specifically aimed at studying molecular hydrogen in the universe.
A recent review of the imaging FTS concept, with some historical perspective and technical details not discussed in the present paper, is presented by Maillard et al (2011).

3. The IFTS concept

An astronomical imaging Fourier transform spectrometer (IFTS) is basically a Michelson interferometer inserted into the collimated beam of an astronomical camera system, equiped with two detectors. Contrary to conventional integral field dispersive spectrographs, all detector pixels are used for imagery as with the Fabry-Perot, but instead of acquiring spectral slices one by one (thus rejecting every other wavelength) to cover the entire waveband of interest, an "all in one" approach is used. Moreover, while the Fabry-Perot is limited to a very narrow wavelength range (typically 1 or 2 nm), an IFTS has no waveband limitations other than the sensitivity of the detectors and the transmitance and reflectance properties of its optics. Schematically, the core of an IFTS is a Michelson interferometer consisting of a beamsplitter used to separate the incoming beam into two equal parts; two mirrors on which the two halves of the original beam are reflected back; a moving mechanism to adjust the position and orientation of one of the mirrors (the other mirror is fixed); and a metrology system to monitor the mirror alignement. All wavelengths from the field are simultaneously transmitted to either one or both of the interferometer outputs in which the array detector sits. The interferometer is configured to modulate the scene intensity between the two outputs instead of spectrally filtering it. This configuration results in a tremendous light gathering power since no light is lost except through items common to any optical design (substrate transmission, coatings efficiency, quantum efficiency of detector). All photons from the scene can hence be recorded at each exposure provided that both complementary outputs of the interferometer are recorded (see below).
Hence, instead of corresponding to a particular slice of the expected spectral data cube, each exposure populates what is called the "interferogram cube": a series of broadband images of different intensities. The key to an unambiguous spectral information recovery lies in the calculation of a Discrete Fourier Transform (DFT or FFT) on each pixel recordings through the interferogram cube. The vector composed of such a pixel recording is called an interferogram and is uniquely determined by the spectral content of the light shined on the pixel. The interferogram cube can at any time during acquisition be turned into a spectral cube since each acquired broadband image contains information covering the whole waveband. The inclusion of additional exposures to an interferogram cube simply refines the meshing of the output spectra (spectral resolution, see section 3.2) and has no effect on the waveband which is determined by scanning parameters and optics transmission (including a filter introduced in the optical path to reduce the width of the waveband, if necessary).
In order to properly generate the interferogram cube, a moving component of the interferometer must be precisely positioned at predetermined sequential interference position before each exposure can be recorded. This operation is referred to as scanning through the interference patterns. For the IFTS, the scanning parameter is the Optical Path

Difference (OPD) existing between both arms of the interferometer and the moving component a mirror. The DFT calculation assumes that all data points of the interferogram vector are acquired at equidistant OPD intervals. Deviations from this assumption result in an increased noise level or artifacts in the resulting spectra. The performance of an IFTS instrument is thus tightly linked with the performance of its OPD scanning system, which can be very challenging in the visible band (350 - 850 nm). The technical challenges associated with building an efficient Michelson interferometer in this wavelength range are partly responsible for the absence of a widespread use of IFTS in astronomical observatories today. In practice however, a lot of the recent IFUs require increasingly complex data processing software such that this apparent distance to the final data tends to even out among 3D capable instruments. The main IFTS limitation that remains in the classic IFU selection trade-off is the balance between spectral resolution desired and acquisition time (or number of exposures required). Another potential hurdle is, we think, conceptual. While dispersive spectroscopy is very intuitive, to a point where most highschool students have experienced the use of a prism or a dispersive grating, and understood the process giving rise to a rainbow, Fourier transforms are non-trivial mathematical concepts standing between the data acquisition and the desired spectrogram. The fact that a given pixel recording cannot be directly related to a given spectral point typically leaves observers with a somewhat less tangible feeling for the data which must be addressed by the careful design of a comprehensive user interface.

3.1 Spectroscopy with a Fourier transform spectrometer

To explain how a Fourier transform spectrometer works, we first assume that the interferometer at the core is a classical Michelson, that the light coming to the telescope is monochromatic, such as a laser beam , and that the signal is recorded on a single-pixel detector (see Fig. 1). The incoming beam is first split into two equal parts by a beamsplitter.

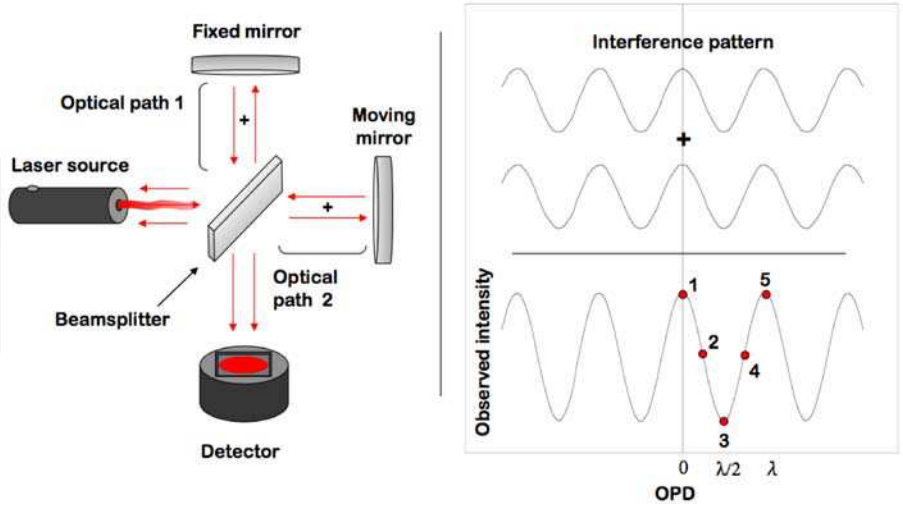


Fig. 1. Optical configuration of a classical Michelson interferometer at the core of a Fourier transform spectrometer

Half the light is transmitted through the beamsplitter, bounces back on a moving mirror and interferes, in the beamsplitter, with the other half-beam which has, in the meantime, been reflected by the beamsplitter to a fixed mirror and bounced back. Initially, the optical path travelled by the two beams are the same; we are at the Zero Path Difference (ZPD) position. The two beams are in perfect phase when they combine and the interference is completely constructive: the detector receives the sum of the two beams while none of the light goes back to the source. The moving mirror is then slightly displaced (typical displacements vary between 175 nm and a few micrometers, depending on the wavelength range and the desired spectral resolution), creating a small offset in optical paths between the beams. These not being in phase anymore, the interference is not entirely constructive and the detector receives a bit less light while the difference goes back to the source.

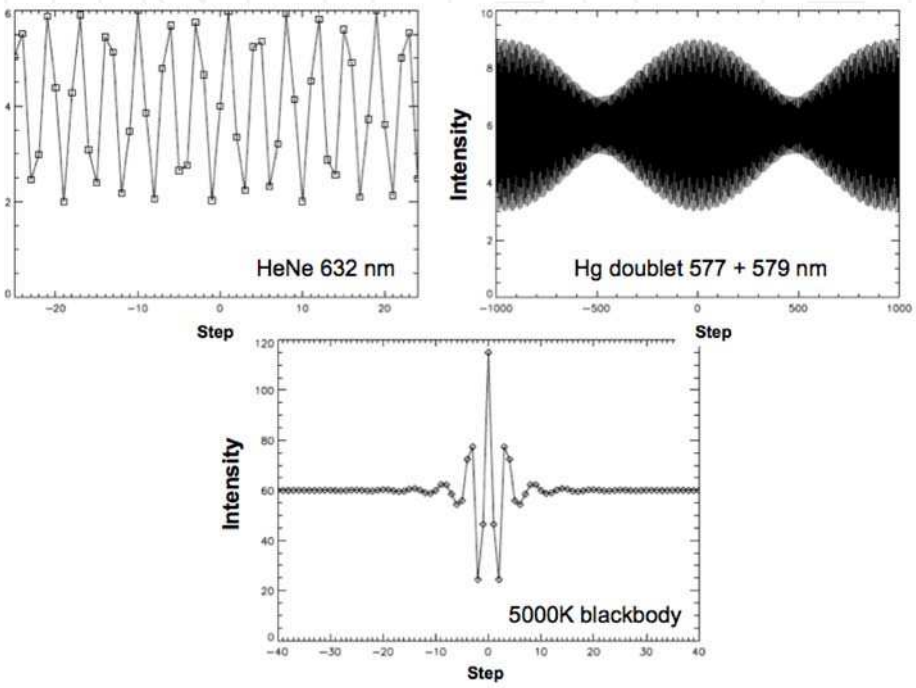


Fig. 2. Simulations of interferograms from different light sources, where step 0 corresponds to the ZPD. Upper left: single He-Ne (632 nm) laser, sampled with mirror displacement steps of 175 nm; the sinusoidal pattern is clearly seen. Upper right: mercury doublet (577 nm and 579 nm); the beating between the two frequencies is obvious. Lower graph: a continuum 5000 K blackbody source; most of the action occurs near the ZPD

Eventually, after a sufficiently large mirror displacement, the interference will be totally destructive: no light will be recorded on the detector, all the beam will go back to the source. In practice, an interferogram is obtained by first moving the mirror to a predetermined position away on one side of the ZPD, moved by equidistant steps to the same position on the other side of the ZPD. Since the incoming light source is monochromatic, the temporal

signal pattern received by the detector after a large number of mirror displacements, called an interferogram, will be sinusoidal. The original frequency, or wavelength, of the laser light beam is thus recovered by calculating the Fourier transform of the time-dependent signal recorded by the detector. An incoming beam which includes two or more emission lines will produce a more complex interferogram, while the interferogram pattern of a continuum source will be mostly concentrated near the ZPD. Astrophysical objects, such as planetary nebula, ionized gaseous nebulae or galaxies present complex spectroscopic features (many emission lines produced by ionized gas superimposed on a stellar continuum) which produce very diverse types of interferograms. Again, a Fourier transform of these interferograms will recover the original spectra, in terms of both frequency and intensity. Fig. 2 illustrates the shape of typical interferograms from different artificial sources.

3.2 Instrument line shape

What does a spectrum obtained in such a way look like? In principle, the Fourier transform of a continuous, infinite interferogram will provide exactly the same frequencies and amplitude as the incoming beam. If the light beam to be analyzed is monochromatic, the interferogram will be a sine wave. Assuming that this interferogram is continuous and infinite, its Fourier transform will be a delta function. However, in real life, the interferogram is sampled at discrete, and in principle equal, intervals of optical path difference: the intensity (or, in the case of an imaging FTS, an image) of the source is obtained, then the mirror is moved, another image is obtained, and so on. Moreover, the signal is observed for a limited time and thus the interferogram is sampled with a finite number of data points. For example and as we shall see later, for a typical astronomical scene observed in the 650 – 680 nm range, the interferogram is sampled with a mirror step of about 5 micrometers and 300 steps for a total optical path difference of ~ 3 mm (the optical path corresponds to twice the mirror translation). Therefore, the recovered line shape will be the "ideal" incoming line shape, degraded by a function corresponding to the "imperfect" instrument. In mathematical terms, the limited spectral resolution of the final spectrum is caused by the convolution of the original line shape with a function that depends on the properties of the instrument and the sampling technique. Using the definition of the Fourier transform and its properties (seen elsewhere in this book), we see that:

$$f(x): \text{perfect, infinite interferogram}$$

$$\text{g(x): instrumental function}$$

$$FT\left[f(x)\cdot g(x)\right] = \int_{-\infty}^{\infty} f(x)\cdot g(x)\, e^{-i\omega x} dx = F(\omega)\otimes G(\omega)$$

$$\text{where } F(\omega) = FT\left[f(x)\right] \text{ and } \mathrm{G}(\omega) = FT\left[g(x)\right]$$

Let's first have a look at the effect of a finite interferogram on the instrument line shape. In this case, the observed interferogram can be seen as the product of an infinite interferogram (a sine wave for a monochromatic light source) with a square box function, $g(x) = sb(x)$, having an intensity of 1 between –d and +d (the maximum optical path difference on both sides of the ZPD) and 0 everywhere else. The instrument line shape (ILS) will thus be determined by the Fourier transform of the square box:

$$ILS = FT\left[sb(x)\right] = \int_{-\infty}^{\infty} sb(x)e^{i\omega x}dx = \int_{-\infty}^{\infty} e^{i\omega x}dx = \frac{e^{i\omega x}}{i\omega}\Big|_{-d}^{d}$$

$$= \frac{1}{i\omega}\left[e^{i\omega d} - e^{-i\omega d}\right] = \frac{1}{i\omega}\left[\left\{\cos(\omega d) + i\sin(\omega d)\right\} - \left\{\cos(\omega d) - i\sin(\omega d)\right\}\right]$$

$$= \frac{2i\sin(\omega d)}{i\omega} = 2d\frac{\sin(\omega d)}{\omega d}$$

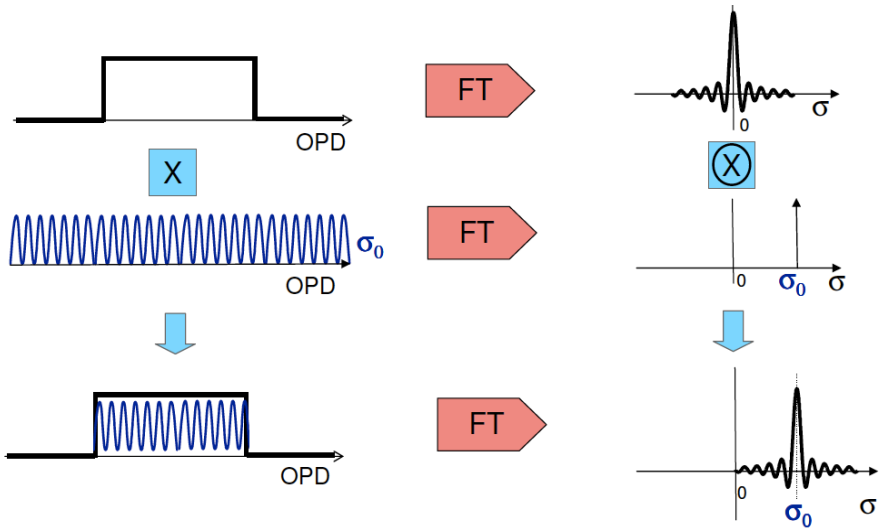


Fig. 3. FTS instrument line shape of a monochromatic source for which the interferogram was truncated by a limited observing time. The recovered spectrum is the convolution of a delta function with the sinc function, which is the Fourier transform of a square box

The instrument line shape generated by the fact that the interferogram is bounded is thus a sinc function. If the observed light source is monochromatic, the recovered spectrum will be the convolution of a delta function $F(\omega)$ with the sinc function $G(\omega)$ (see Fig. 3): a sinc function centered at the wavelength of the incoming light source. Let's now determine the width of this function, which dictates the spectral resolution attainable with an FTS, by calculating the half width at half maximum of the ILS:

$$2d\frac{\sin(\omega_{hwhm}d)}{\omega_{hfhm}d} = 0.5 \times 2d\frac{\sin(\omega_{\max}d)}{\omega_{\max}d}$$

Using the fact that $\frac{\sin x}{x}; 1 - \frac{x^2}{6} + \frac{x^4}{120} - \ldots$, we see that the maximum of this function, 1, occurs at $\omega_{\max} = 0$. Therefore,

$$\sin(\omega_{hwhm}d) = 0.5 \Rightarrow \omega_{hwhm} = \frac{1.9}{d}$$

The larger d is, the thinner the ILS becomes. Spectral resolution of an FTS is thus directly proportional to the total optical path difference sampled by the interferometer. The maximum spectral resolution attainable with an FTS is then set by the maximum displacement of the moving mirror within the interferometer. In an imaging FTS, other factors, such as image quality (defined by the instrument optics or, more likely, the seeing disk blurred by convection in the atmosphere) or image sampling can decrease the theoretical spectral resolution of the instrument. Figure 4 illustrates the typical line shape of a monochromatic source obtained with an imaging FTS. It is possible to modify the shape of the ILS a posteriori by multiplying the interferogram by a function that would attenuate the presence of the "side lobes" associated with the sinc function; one then talks about apodization. One of the most widely used apodization functions is the Gaussian, since it is the only function whose frequency content is the same as the function itself; the Fourier transform of a Gaussian is a Gaussian. The counterpart of apodizing the interferogram however is a loss of resolution. In our typical astronomical applications, the ILS sidelobes disappear in the noise after an apodization with a Gaussian that decreases the spectral resolution by about 25%.

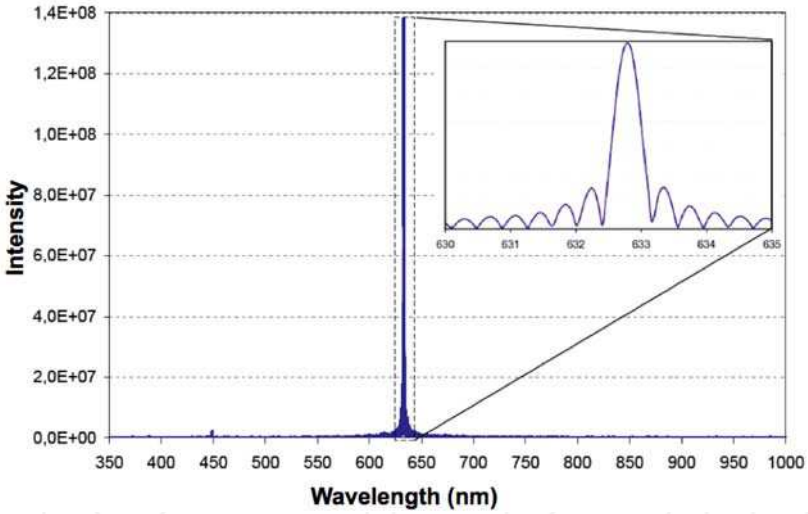


Fig. 4. Spectrum of a He-Ne laser (632 nm) obtained with the imaging FTS SpIOMM, showing the typical instrument line shape

3.3 Modulation efficiency

Even if a good transmission is achieved in the optical design, contributing positively to the image's signal-to-noise ratio, it does not necessarily translate in a good performance for spectroscopy. In order to perform well on this aspect, a good modulation efficiency is also required. The performance of an FTS is thus characterized by its modulation efficiency (ME), i.e. the capability of the interferometer to modulate the incident light:

$$ME = \frac{I(\text{modulated light})}{I(\text{incoming beam})}$$

This parameter can be viewed as an analog to the grating efficiency in dispersive spectrographs. In the worst case scenario, where the modulation efficiency is zero, the light from the source is recorded on the detector but the interferogram is a straight line and no spectral information can be extracted from it.

This efficiency depends on a multitude of factors, the most important being the following:

1. The capability of the beamsplitter to separate the incident beam into two beams of equal intensity over the whole wavelength range (ME $\propto$ 4RT, where R and T refer to the reflectance and transmission of the beamsplitter). The modulation efficiency is however relatively permissive to deviation from the 50% perfect case. For example, a 60% - 40% R-T beamsplitter can still generate modulation efficiency near 96% while a 70% - 30% one would limit the performance to 84%.This is usually not a big problem over a small wavelength range, but it becomes a challenge if the FTS covers, for example, the entire visible range (from 350 to 900 nm).
2. The surface quality of the optical components in the interferometer (mirrors and beamsplitter). Three conclusions are obvious from Fig. 5, where we present the influence of the surface quality on the ME for different wavelengths and optical configurations. At a given wavelength, ME is lowered by a decreased surface quality; it is more and more difficult to obtain a good ME as we move from the infrared to the ultraviolet (most FTS commercially available indeed work in the infrared); and the number of reflections within the interferometer plays a major role in the global ME. Mirrors with a surface quality of $\lambda/20$ (peak-to-valley) are commercially available for a reasonable price, but large $\lambda/30$ mirrors must be custom made and are therefore much more expensive. Moreover, even if the mirror substrate is of high enough quality, any error in the coating deposit or any tension caused by the mechanical parts used to maintain the mirror within the interferometer can ruin the initial surface figure and dramatically reduce the modulation efficiency, especially in the blue part of the spectral range.
3. Homogeneity of the refraction index within the interferometer cavity. Any convection or temperature inhomogeneity in one or both arms can alter the OPD and therefore lower the ME, since the velocity of light depends on the refraction index of the material in which it travels.
4. The mirror alignment. In order for the beams from the two arms to interfere properly, the two mirrors need to be very well aligned. The smallest deviation, in any direction, from a right angle between the two mirrors reduces the spatial coherence (interference) of the two beams as they recombine. Again, this effect is more obvious at small wavelengths. A deviation of only 1.5 microradian from perfect alignement can decrease the ME by up to 25% at 350 nm.
5. Stability of the OPD during an exposure. The optical distance between the two mirrors must be kept constant during an exposure at a given step. An OPD jitter with a standard deviation of 10 nm typically reduces ME by 1 to 2%.

The metrology and servo system play a crucial role with regard to the last two points, since the mirrors must be aligned with a precision of less than a microradian and their distance kept constant to within a few nanometers during an exposure. As we shall see later, monitoring the distance between the two mirrors as well as their alignment many thousand times per second, and a fast correction of any deviations, are required to ensure a constant, high modulation efficiency.

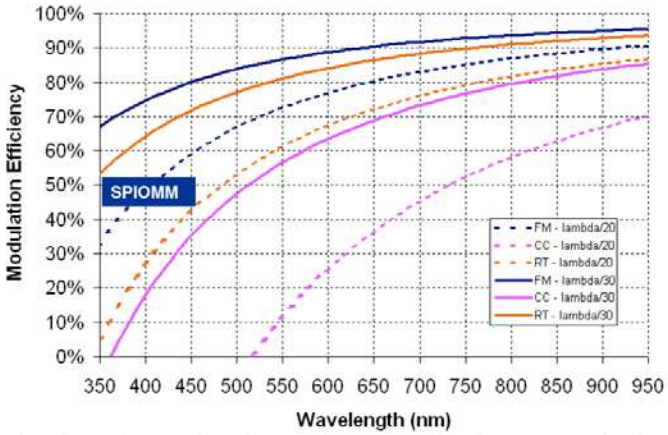


Fig. 5. Wavelength-dependence of an FTS's modulation efficiency for different configurations and mirror quality. Peak-to-peak asperities of $\lambda/20$ and $\lambda/30$ are considered, as well as three mirror configurations: flat mirrors (FM; one reflection on each arm), roof-top mirrors (RT; two reflections), and cube corners (CC; three reflections)

3.4 Four ports design

In a traditional Michelson interferometer, half the light goes back to the source after it interferes with the other beam. But in astronomy, there is so little flux coming from distant nebulae and galaxies that every photon counts; moreover, returning back half of the photons coming from a galaxy after it had crossed millions of light-years seems absurd. A third factor also needs to be considered in astronomical observations: the sky transparency. Any change in the sky transparency during the observations, caused by thin clouds or a change of airmass, will affect the interferogram, introducing noise or even artefacts in the recovered spectrum. But since the sum of both outputs from the interferometer is, in principle, constant during he whole data acquisition, any fluctuation in the total signal can only be interpreted as a sky transparency variation and be corrected for.

Therefore, the Michelson interferometer used in an astronomical IFTS must be designed in a way that allows the two half-beams to be recovered after having interfered with each other. Such configurations, with two output ports, are usually achieved using cube corner retroreflectors or cat's eye mirrors (Maillard & Simmons 1992, Boulanger et al. 2008). If one wants to produce an interferometer that strongly modulates the visible light down to the near UV, the optical surfaces located between the two beamsplitter passes must exhibit excellent surface figures ($< \lambda/20$) to allow a good interference of the two recombining wavefronts, as seen in the previous section. The two mirror options mentioned above complicate this task because of the additive errors of all three reflections encountered (see Fig. 5) in addition to optical shape consideration in the case of cat's eyes or orthogonality issues in the case of cube corners. The flat mirror interferometer, besides providing a better throughput than the others and improved wavefront accuracy, can also provide a two output-ports configuration when used off axis as shown on Fig. 6. There are however two disadvantages of using flat mirrors. First, the alignment of the two mirrors must be very carefully controled by a servo control system (cube corners are insensitive to slight misalignements); this puts serious constraints on the metrology (which measures the

distance between the two mirrors and their alignement) and piezo actuators which must respond very rapidly and accurately to any misalignement. Second, since the center of the interferometric pattern is not aligned with the detector's center, the ultimate spectral resolution attainable by the FTS is lower than that obtained with cube corners. As we shall see, neither of those constraints is a show stopper for practical purposes.

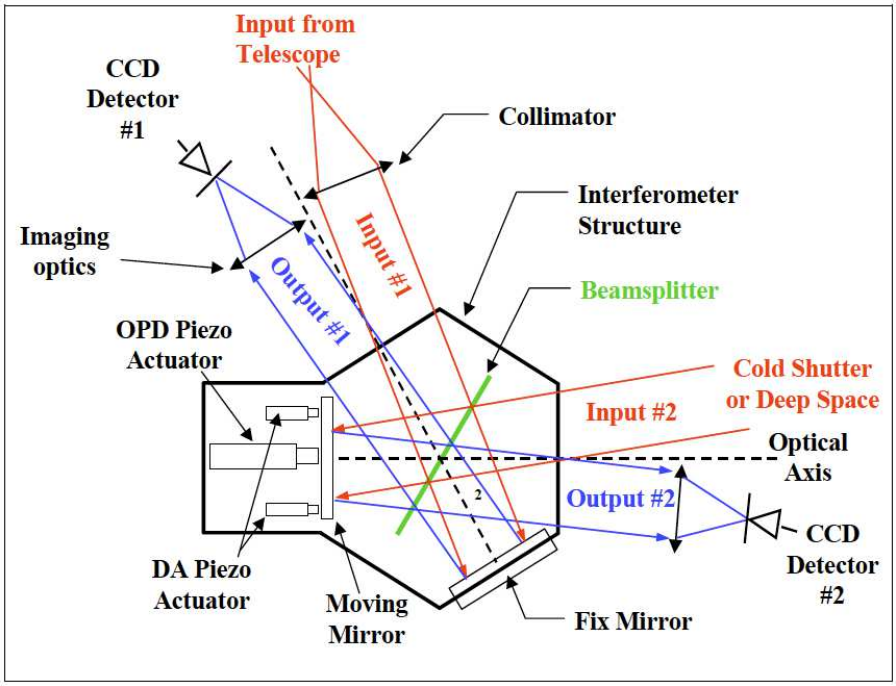


Fig. 6. Interferometer design of an of-axis imaging FTS

3.5 An imaging FTS in the visible band

By adding input and output optics (a group of collimating lenses before the interferometer, and one group of imaging lenses for each output port) to image a wide field of view and a multi-pixel detector such as a CCD at each output port, the standard Fourier transform spectrometer becomes an imaging FTS. Since most Integral Field Spectrographs used in astronomy today cover a small field of view (typically a few arcseconds on a side), an enormous advantage of the imaging FTS is its ability to cover a much wider (arcminutes) field. As we shall see in the next section, a large number of scientific programs benefit from an instrument capable of simultaneously obtaining spatially resolved, high quality spectra of extended (of the order of 10 arcminutes) sources.

The basic design of the optical system surrounding the interferometer is very similar to that of conventional astronomical cameras and focal reducers, but with an additional constraint on the optics. Astronomical imagery is performed with filters which select specific bandpasses aimed at extracting information on the color of objects (broadband filters such as the SDSS u'g'r'i'z' system, with a typical bandwidth $\Delta\lambda \sim 150$ nm) or on specific emission

lines (narrowband filters centered, for example, on the Hα line, with a typical bandwidth Δ λ ~ 2 nm) to highlight the ionized gas in galaxies. These cameras usually observe one band at a time allowing for refocus between each of them, thereby relaxing the constraints on the optical design. The challenge for the optical design of an imaging FTS, which can in principle accept a very wide wavelength range at every single step is to satisfy the image quality requirement over the whole waveband throughout the field of view.
Two more constraints on detector technology are imposed by the imaging FTS concept and need to be discussed because they were partly responsible for the absence of wide FOV imaging FTS up to recently and still drive the choice of the detector today. In order to build an interferogram, an exposure of the scene must be obtained at every mirror step and read out from the detector. In the visible range, the most efficient detector is without contest the charge-coupled device, or CCD. An electronic readout noise is added to the shot noise from the scene every time the CCD is read out. Moreover, it takes time to read an entire CCD and register its content on a hard disk, from a few seconds up to one minute, depending on the readout rate and the number of pixels. In order to minimize the « dead time » due to this transfer of information, the CCD readout rate ought to be very high. But the readout noise is proportional to the readout rate, so the two problems are linked. In the 1990's, the largest CCDs had 1024 x 1024 pixels ; their readout time was around one minute and the readout noise was about 10 electrons. Nowadays, 2048 x 2048 pixel CCDs are common, and many of them are equipped with multiple amplifiers allowing the simultaneous reading of its four quadrants at speeds up to 1 MHz : the readout time of the entire detector is reduced to a couple of seconds and the readout noise is down 2 to 5 electrons. In the near future, zero-noise EMCCDs (Daigle et al. 2009) will replace conventional CCDs for this type of application and will enable speed-scanning of the OPD instead of the step-scan approach; this will at the same time increase the signal-to-noise ratio at very low flux level and lower the dead time due to the CCD readout.

4. SpIOMM, a wide-field imaging FTS for the Mont Mégantic Observatory

In 1999, the *NGST science and technology exhibition* meeting was held in Hyannis, Massachusetts, to present and discuss, among other things, the best possible suite of instruments to be attached to the giant space telescope now known as the James Webb Space Telescope (JWST: www.jwst.nasa.gov). Three teams then proposed concepts of IFTS that would allow the James Webb to acquire hyperspectral images over large field of view (Graham 2000, Morris et al. 2000, Posselt et al. 2000). The scientific rationale and a preliminary technical performance study of a space-based IFTS had been published a year earlier (Graham et al. 1998). Although the enormous qualities of such an instrument were recognized by the panel of scientists and engineers who had to select the JWST instrument suite, the IFTS solution was not selected. One of the reasons invoqued was that an IFTS is competitive with a conventional spectrograph only if the number of sources in the field of view is large. To quote the report of the NGST ad hoc working group on the near-infrared spectrograph study (Huchra et al. 1999) about IFTS:

> *In addition to preserving the diffraction limit of the telescope, this technique allows unbiased spectroscopy since there is no preselection of objects. FTS instruments trade their spatial multiplex gain against the spectral multiplex advantage of grating instruments. Thus the richness of the observation scene largely dictates the advantage of an FTS against dispersive designs. The FTS is most efficient for low spectral resolution of extremely crowded fields.*

We would add to the last sentence "or at moderate to high resolution, over selected spectral wavebands, of extended sources". Indeed, a dispersive spectrograph is more efficient than an FTS for the observation of single sources. Multi-object dispersive spectrographs are also more efficient than IFTS if only a limited number of sources (say, a hundred) are spread over a large area of the sky. This was indeed one of the arguments mentioned to reject IFTS for the JWST: since one of the most important goals of this telescope is to observe the faint, distant Universe for cosmological studies, the number density of sources (galaxies) bright enough for spectroscopic studies was not expected to be large enough to justify the use of IFTS. However, an IFTS becomes the instrument of choice if the target spans a fair fraction of the entire field of view, which is the case for Galactic nebulae and nearby galaxies. This led us to our desire to design and build SpIOMM.

4.1 Science case and science-based requirements

In order to demonstrate the capabilities of a wide-field imaging FTS working in the visible band, our group has designed and built SpIOMM (**Sp**ectromètre **I**mageur de l'**O**bservatoire du **M**ont **M**égantic) at Université Laval in close collaboration with ABB-Bomem, a Québec City-based company specialized in the development and manufacturing of commercial FTS. The 1.6-m telescope of the Mont Mégantic Observatory (OMM), jointly operated by Université Laval and Université de Montréal since 1978, is perfectly suited to train highly qualified personnel such as M. Sc. and Ph. D. students, postdoctoral researchers and engineers, but is also a perfect test bed to design, build and use innovative astronomical instruments such as polarimeters (Manset & Bastien 2002), infrared cameras and spectrographs (Artigau et al. 2009), low-noise detectors for F-P systems (Daigle et al. 2009) as well as imaging FTS such as SpIOMM.

The primary objective of any astronomical instrument development being to address a science case, the development of the IFTS at OMM was fed by the desire to acquire complete data sets to feed existing science projects already ongoing at Université Laval on the interstellar medium, late stages of stellar evolution, star formation and galaxy evolution and which could not be obtained with the current instrument suite of the observatory. The general theme of "cosmic recycling" was our motivation since the beginning: how do chemical elements such as oxygen, carbon or nitrogen, which are produced by nucleosynthesis in stellar cores, get transferred via stellar winds and supernova explosions to the interstellar medium where they "pollute" giant molecular clouds which eventually contract under their own gravity and form new generations of stars, surrounded by planets where life can eventually appear and evolve? How does this process occur in our own galaxy, the Milky way, and in other galaxies? Can we trace the past evolution of galaxies by measuring the spatial distribution of their heavy elements? Obtaining spatially resolved spectra of extended objects such as ionized nebulae, supernova remnants, planetary nebulae and nearby galaxies is a prerequisite to answer these questions and an instrument such as a wide-field imaging Fourier transform spectrometer could provide some answers.

The work started by first identifying the measurement aspect that could benefit the most from the IFTS concept. In this respect the instrument development deviated early on from the JWST instrument concepts and other past IFTS development (such as BEAR) due to the waveband of interest for our particular research: the visible range (from 350 nm to 850 nm) is very rich in diagnostic emission lines allowing us to determine the physical characteristics of ionized nebulae, such as their chemical composition, density or temperature. Theoretical models abound which relate emission line ratios in this wavelength range to the properties

of nebulae and their ionizing stars (Kewley & Dopita 2002). Moreover, spectra of small sections of extended objects are available for a direct comparison with those obtained with an IFTS. But choosing this wavelength range has its drawbacks from an instrumental point of view, as we have seen in section 3.3. It can easily be understood that controlling interference of light mechanically is better achieved at longer wavelengths at which a given fraction of a wave corresponds to a larger displacement hence easier to detect and correct. As such, FTS users have traditionally had more success when operating in the IR and far-IR although usage down to far-UV are documented in the literature in other fields of science. Considering the advances in nanometer level precision actuators, high stability metrology lasers and electronics speed, new tools are now available to tackle challenge of the past. Most importantly, the expertise with laboratory demonstration instruments at ABB Bomem and the NGST breadboard (Wurtz et al. 2002a) on which members of our team had closely worked gave us a sufficient level of confidence that we could succeed in the visible range.

From our point of view, the niche of an astronomical IFTS clearly sits in the wide field coverage, moderate spectral resolution (R ~ 2000 -10 000) and the large waveband, although it is interesting to note that the FTS was classically known for its ability to obtain very high spectral resolution. Using a FTS in an imaging configuration typically does not affect its spectral potential, but the faintness of the target diffuse nebulae make the high resolution measurement difficult to reach in a reasonable time. However, one can still exploit the high spectral resolution capability of an IFTS by reducing its waveband in order to proportionally reduce the number of exposures required to obtain a given spectral resolution.

A set of science-based requirements were established as a starting point to design the instrument:

- Field of view - Wide field, of the order 10 arcminutes or more; this corresponds to the size of some of the largest Galactic planetary nebulae and to a fair number of nearby galaxies of interest, up to the Virgo cluster.
- Wavelength range - Two important diagnostic emission lines in the spectra of ionized nebula define the useful spectral range: [OII] 372.7 nm in the near ultraviolet and [SII] 673.1 nm in the red. Allowing for some redshift due to the expansion of the Universe, which affect galaxies or group of galaxies within reach of the OMM, a lower limit to the red was set to 750 nm.
- Minimal spectral resolution – Should be high enough to separate Hα (656.3 nm) from the [NII] (654.8, 658.4 nm) doublet, as well as both members of the [SII] 671.7, 673.1 nm doublet: $R = \lambda/\Delta\lambda > 500$. Kinematical study of expanding nebulae and galaxies require a higher value of R ~ 2000.

A trade study was then performed to define the best design that would meet these requirements (Grandmont 2007), which led to the current design, described in the next section, and construction of SpIOMM in 2002 – 2004. Details about the design and tests of SpIOMM can be found in a series of SPIE papers (Grandmont et al. 2003, Bernier et al. 2006, Bernier et al. 2008).

4.2 Description of the instrument

SpIOMM is capable of obtaining spectra in selected bands of the visible spectrum (from 350 to 850 nm) of every light source in a 12 arcminute (circular) field of view of the OMM telescope; a 12′ x 12′ FOV is actually recorded on the detector but optical abberations are important in the corners of the field. The spectral resolution is variable, depending on the

need of the observer, from R = 1 (broad-band image) to R = 25 000. The spatial resolution is limited by the seeing disk (atmosphere blur), which is typically 1 – 1.5 arcsecond. The dual output design of SpIOMM (see below and Fig. 6) ensures, in principle, that virtually every photon collected by the telescope reaches the detectors and is analyzed. We have however worked until now, for budgetary reasons, with only one detector, a Princeton Instrument CCD camera with an array of 1340 x 1300 20 μm-pixels (corresponding to 0.55 arcsecond on the sky). A data cube thus results in 1.7 million spectra. But we recently acquired a second CCD, an Apogee Alta 2048 x 2048, 15 μm-pixels that will collect the flux from the second output port. SpIOMM is unique in the world, offering the largest field of view of any integral field spectrograph. After five years of research and development, we have demonstrated that the concept behind SpIOMM is sound and viable, and that such an instrument is capable of producing high quality hyperspectral data cubes over a very extended field of view.

The design (shown on Fig. 6) is based on the Michelson interferometer with a 30 degree incidence angle on the beamsplitter-compensator assembly, which minimizes their circular size. The use of two plane mirrors reduces the number of optical surfaces encountered by the science beam from the telescope output to the camera and therefore grants better throughput. Also, the instrument configuration places the incoming beam 8 degrees off–axis perpendicular to the optical interferometer plane, which allows access to the two output ports. The interferometer uses a dynamic alignment control (metrology laser and piezoelectric actuators) and step-scan operation. The robustness of an interferometer operating in the visible range is of paramount importance. FTS are known to serve as good microphones, seismographs or even thermometer, since they are extremely sensitive to vibrations and temperature changes. The stabilization of a moving mirror at a small fraction of the shortest wavelength accepted translates in this case to nanometer level precision. Any change in temperature, gravity orientation or vibration is at this scale bound to have a noticeable effect on a macroscopic scan mechanism. Hence much care must be taken in either passively reject some of these perturbations through a stiff and athermal design or actively through a well tuned servo control system. SpIOMM's moving mirror is mounted on a parallelogram porchswing-type mechanism which offers frictionless displacement. It is actuated by a piezo-based stepper motor mounted in series with a small range high-frequency response piezo (the OPD piezo). The stepper actuator can travel a Maximum Path Difference of 1 cm, corresponding to an average spectral resolution of R ~ 25 000. In practice however, we have never pushed SpIOMM to this resolution limit. The ensemble allows a high stiffness (10 N/μm) in order to ensure good passive stability of the Optical Path Difference (OPD). A dynamic alignment system using piezoelectric (DA piezo) combined with the mirror displacement mechanism is required for precise positioning and to evenly sample the interferogram. Also, this maintains the mirror position and alignment subject to gravity, vibration and thermal perturbation during data acquisition. The interferogram is sampled at a step size determined by the Nyquist criterion. In order to measure the wavelength modulation at 350nm, the scan step is 175 nm. When it is set to a value larger than 175nm, it has to be used with an appropriate filter for the desired band.

The moving mirror position and alignment are detected by a metrology system and acquired by a servo-computer at a rate of 8 MHz. This feedback is given by an infrared 1550 nm laser beam expanded before passing through the interferometer. We take advantage of the fact that the telescope's primary mirror has a hole in its center : the image of the science

beam in the interferometer therefore shows a hole where the metrology beam is sent. The metrology and science beams are therefore completely separated, both physically and in the frequency domain (since the CCD is completely insensitive to the laser's wavelength). The metrology beam passes through the same optical path as the science beam and strikes the center of the fixed mirror, where a thin circular layer of glass has been deposited. This layer acts as a wave-retarder which causes a retardation of $\lambda/8$ at the metrology source wavelength (1550nm) so that after the reflection on the mirror, the metrology beam is retarded by $\lambda/4$ since it passes twice in the layer. The diameter of the wave-retarder layer is smaller than that of the metrology beam, so that a small misalignement of the two mirrors can be measured (Fig. 7). The metrology beam is then recorded orthogonally by two 2x16 pixel detector arrays. Once the correction values are calculated, commands are sent back to the dynamic alignment piezos which correct the misalignement.

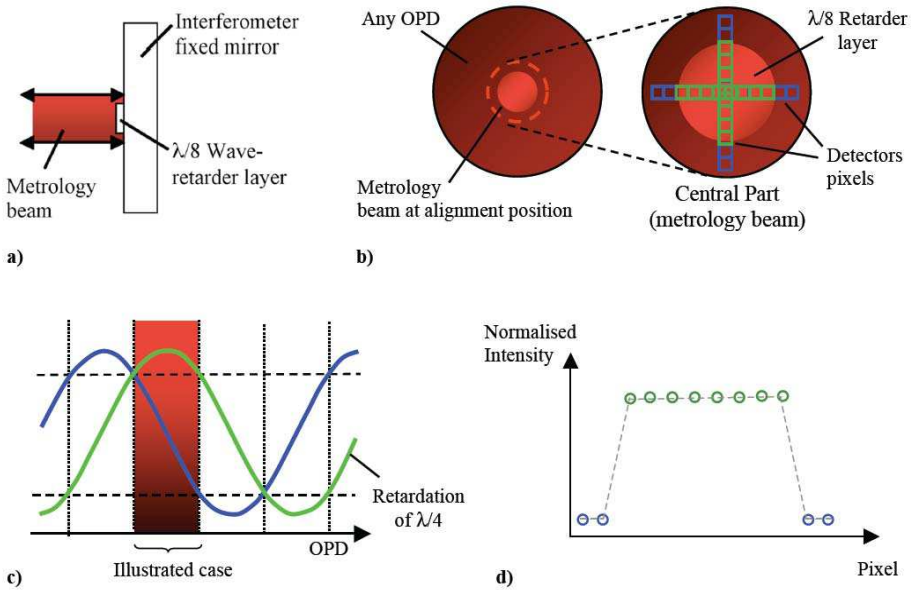


Fig. 7. Design of SpIOMM's metrology system consisting in the detection of a flat fringe pattern (of the IR source) comprising a smaller circular region of $\lambda/4$ retardation created by a wave-retarder layer placed at the center of the fixed mirror (a). The flat fringe pattern is recorded by two detectors orthogonally placed at interferometer output (b). The signal of a pixel located in the inner region of retardation of the metrology beam will be delayed by $\pi/2$ compared to the signal of a pixel outside (c). For a specific OPD, (d) shows the shape of the aligned intensity pattern read by one of the detectors. Any variation from the "perfect" shape shown in (d) causes the piezoelectric actuators to realign the mirror

4.3 Tests and problems encountered

SpIOMM was assembled in the laboratory in early 2004 and the first tests were extremely encouraging: high resolution spectra of laser sources were obtained on a regular basis, and the instrument line shape was exactly like it should (Fig. 8). The laboratory environment is ideal to

test an instrument such as SpIOMM: the temperature is constant to within a degree, there is no wind nor outside vibrations and the gravity vector is constant during data acquisition.

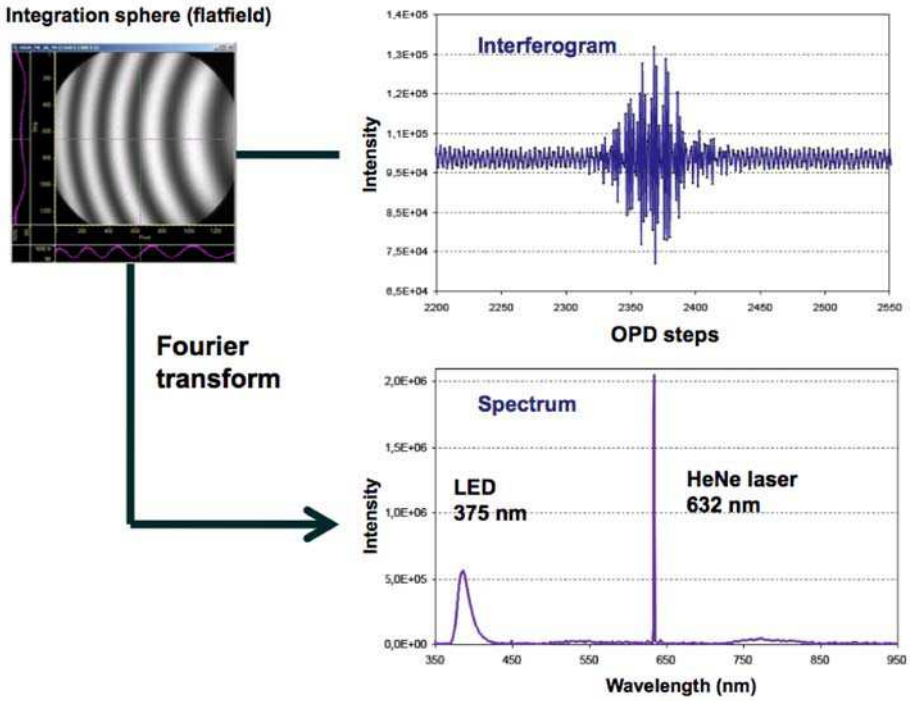


Fig. 8. Example of data obtained in the laboratory. Two light sources (a HeNe laser at 632 nm and a LED at 375 nm) are sent to an integration sphere which is observed with SpIOMM using exactly the same technique as for any astronomical source. The upper left image shows the fringe pattern recorded at a given OPD by the CCD. Notice that the interference pattern is not centered on the CCD; this is caused by the off-axis approach we have chosen and which is shown in Fig. 6. The central part of the interferogram of one pixel, near the ZPD, is shown on the upper right image (the entire interferogram was sampled with 4750 points). A Fourier transform of this interferogram then recovers the spectrum of the two sources

Happy with the lab results, we installed SpIOMM at the telescope in early 2004. But a telescope environment is very harsh for an interferometer. Yearly temperatures vary between +25°C and -35°C, but more importantly excursions of 5 degrees are common during a single data cube observation. Windy nights are frequent and gusts can move the telescope abruptly. The Earth rotates, so does the telescope in the opposite direction to follow the apparent movement of the targets; the orientation of the gravity vector therefore varies constantly with respect to all components of the instrument, including the interferometer's mirrors. The telescope's motors and the air pump that supports the telescope's primary mirror also create vibrations that are transmitted to the instrument. All of these hostile environmental factors impose very severe and stringent constraints on the servo system to maintain the mirror stability. The original servo loop ran at 2.5 kHz, which was sufficient in

the lab to correct any misalignment. It became however obvious that it was not enough to maintain a proper stability at the telescope. The modulation efficiency was not optimal (70 - 75% at 632 nm) and highly variable (excursions of 10% were frequent) during data acquisition. This introduced noise and spurious artifacts in the spectra. The original metrology and servo system (described by Bernier et al. 2006, 2008) was then replaced by the one described above with great success. Science data cubes have been obtained on a regular basis since 2007, with a very stable modulation efficiency very close to the theoretical limit of 85% at 632 nm. Fig. 9 schematically describes the data acquisition of a typical science interferogram and its transformation to a useable data cube. Data processing is an important part of the entire process and is described in more details in the next section.

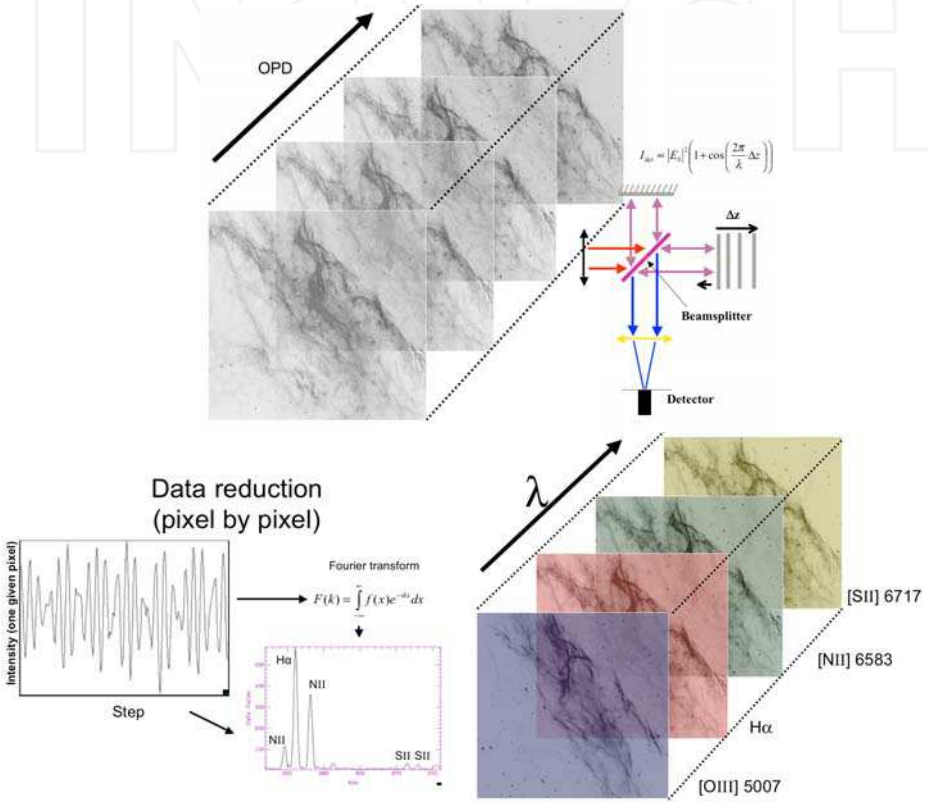


Fig. 9. Data acquisition with SpIOMM (a) By scanning the Optical Path Difference (OPD) of the interferometer and taking images at every step, one gets a datacube composed of one interferogram for every pixel. (b) For a given pixel, the recorded intensity varies as a function of the OPD with a pattern that depends on the spectral content of the source; for example, a monochromatic laser beam would produce a sinusoidal pattern. A Fourier transform of the signal produces a spectrum for every pixel in the image. (c) After Fourier transforming every interferogram, one gets a spectral datacube from which monochromatic images corresponding to the emission lines of interest are extracted. The data shown are extracted from a data cube of the supernova remnant NGC 6992

4.4 Data processing

As with every imaging system using CCD detectors, data processing starts with a series of basic corrections such as bias removal (an electronic signal inherent to the CCD) and flatfield corrections (to correct for the non-flatness of the optical transmission or pixel-to-pixel sensitivity variations). A datacube acquisistion usually takes between one and four hours depending of the spectral parameters and the object's surface brightness. Sky transmission sometimes fluctuates over that period and affects the signal of the interferogram; this is due to the passage of thin clouds or simply the change of airmass. It can add low and high frequency contributions to the spectra. To avoid such artifacts, we want to normalize the baseline signal of the interferogram. To do so, we measure the photometric variations of the at least a dozen stars in the FOV. The interferogram signal of a polychromatic source like a star (almost a perfect blackbody!) modulates only at the ZPD. On either side of the ZPD, the signal intensity is supposed to be flat unless the sky transmission fluctuates (clouds, airmass, etc). Therefore we determine the mean sky transmission signal from the interferograms of a list of stars. Using only one CCD as we have done so far is not an ideal situation because sky transparency fluctuations cannot be completele corrected for, especially around the ZPD. The implemetation of the second output port in 2011 should solve this problem and significantly improve the signal-to-noise ratio of the spectra. Because of flexure within the instrument or between the telescope's guiding system and the instrument during the entire acquisition (the source can be at the zenith at the beginning and 45 degrees from the horizon at the end), the images are not perfectly aligned with each other. Typical shifts of 1 to 3 pixels in both directions are common between the first and last image of an interferogram. The centroid of a dozen stars scattered across the field is then measured in all images, which are then realigned accordingly with a precision of a tenth of a pixel. Finaly, CCDs are very sensitive to cosmic ray hits, which usually affect a few pixels in each image; they are corrected for by comparing signals from adjacent images.

Before transforming the interferogram cube into a hyperspectral cube (x, y, λ) we must apply some operations on the interferogram signal. First, the pre-processed interferogram should have a null mean in order to have its extremities values approaching zero. Therefore, we remove the mean value to each interferogram. We then multiply the interferogram with an apodization function in order to minimize the side lobes in the spectral lines created by the truncation of the signal (finite number of points; see section 3.2). We have tested some apodization windows with our data to find the best function to minimize the side lobes in the spectrum without degrading the spectral resolution too much. The gaussian function provides the best results. Finally, a zero-filling operation is performed on the interferogram in order to increase the number of points up to the next higher power of 2. This is useful for further spectral interpolation and for a faster processing discrete Fourier transforms. We then apply the discrete Fourier transform to each processed interferogram of the datacube (up to 1.7 million) in order to obtain a datacube of spectra. In parallel, we compute the spectral axis scale from the datacube sampling parameters and we interpolate the total number of points on a graduation in nanometers. This rescaling and interpolating operation takes also in account a calibration dataset that corrects for off-axis contributions on the whole FOV: an interferogram of a He-Ne laser obtained at high resolution during the same observing run.

5. Science results with SpIOMM

While the number of scientific applications of an imaging FTS is potentially very large, our group has focused on the interstellar medium of our own galaxy, the Milky way, and nearby spiral galaxies. A meaningful link between local heavy element enrichment and the global chemical evolution of galaxies can only be established by detailed studies of individual wind-blown bubbles in our own galaxy. Winds of evolved stars, and their surrounding bubbles (planetary nebulae, Luminous Blue Variables and Wolf-Rayet ejecta, supernova remnants) are known to be globally enriched with products of nucleosynthesis. While planetary nebulae are ejected by low-mass stars, with slow (20-30 km/s) winds, LBVs and WRs are the late evolutionary stages of the most massive stars with wind velocities of hundreds to thousands of km/s. A complete survey of abundance, density, temperature and kinematic measurements in nebulae surrounding individual evolved stars and ionizing clusters, looking for inhomogeneities in the distribution of processed material (primarily nitrogen and oxygen), which has never been undertaken because it requires wide-field spectroscopic mapping, will provide firm grounds for the interpretation of global galactic abundance studies.

In particular, we have obtained data cubes of planetary nebulae (the envelope of low-mass stars ejected at the end of their life), Wolf-Rayet bubbles (cavities created by the interaction of very massive stars with their surrounding interstellar medium) and supernova remnants, which are the result of the explosion of the most massive stars. We have also targeted nearby spiral galaxies. While the observations could be done without any filter, thereby covering the entire visible range (from 350 nm to 850 nm), we take advantage of the fact that our prefered targets emit most of their flux in a series of emission lines clustered around 500 nm and 660 nm (rest wavelength). The observations are therefore performed in two steps, with a blue (450 – 520 nm) and a red (650 – 680 nm) filter to cover the most intense and diagnostic-rich emission lines. The use of filters also significantly increases the contrast between the targets (nebulae) and the underlying continuum sources (stars) and dramatically reduces the background sky intensity, especially when the Moon is up in the sky. The most important criterium to define the spectral resolution is the capacity to unambiguously separate the lines from the [SII] doublet at 671.7 nm and 673.1 nm (whose ratio is a good indicator of the gas electron density). Such a resolution also ensures a clear separation of the [NII] doublet (654.8, 658.4 nm) from the strong Hα line at 656.3 nm. Such a resolution is obtained in the red filter with 325 steps. The lines full width at half maximum corresponds to a velocity of ~ 130 km/s, but the centroid of each line can be determined with a precision of about 10 – 20 km/s, depending on the line strength, therefore allowing kinematical analysis of nebulae and rotation curves of galaxies. Exposure times vary from 15 seconds per step for the observations of bright nebulae in the red filter to 90 seconds per step for the observations of galaxies in the blue filter. Indeed, a few factors contribute to the difficulty of acquiring blue data : emission lines are generally more intense in the red ; absorption by interstellar dust and the Earth's atmosphere is much more efficient in the blue ; the global throughput of the instrument but especially its modulation efficiency dramatically decreases from the red to the blue end. To the exposure time, one must add the « dead time » caused by the CCD readout and the mirror displacement and stabilization time (a total of 7 seconds per step). During observing runs in 2007 – 2010, we have obtained more than 60 datacubes of galactic HII regions, planetary nebulae, Wolf-Rayet ring nebulae, supernova remnants, nearby spiral galaxies and groups of interacting galaxies.

5.1 Nebulae around evolved stars

Planetary nebulae are the remains of the outer envelope of low-mass stars like the Sun ejected during the late stages of their evolution. They are chemically enriched with the products of stellar nucleosynthesis and expand at relatively low velocities (~ 50 – 100 km/s). The planetary nebula NGC 6853 (M27) was one of the first astronomical targets of SpIOMM, because of its large size, high surface brightness brightness, and because it has been the object of numerous studies, thereby allowing a direct comparison with previous observations with classical techniques. Figure 10 highlights some of the data extracted from the blue and red cubes of M27. While images of this objects had been obtained before at different wavelengths, and long-slit spectra of a tiny fraction of the nebula been analyzed, SpIOMM's data are the first to show spatialy resolved spectra of the entire nebula. In particular, the very high contrast possible by a careful extraction of the images corresponding to particular wavelengths in the data cube allow us to detect very faint structures in the outskirts of the nebula, particularly in the [NII] emission line. This line is particularly difficult to isolate with interference filter imagery because it is very close to the bright Hα line. Moreover, since one spectrum was obtained for every pixel on the scene, we were able to obtain a very detailed diagnostic diagram based on two line ratios. A detailed physical analysis of these data is under way.

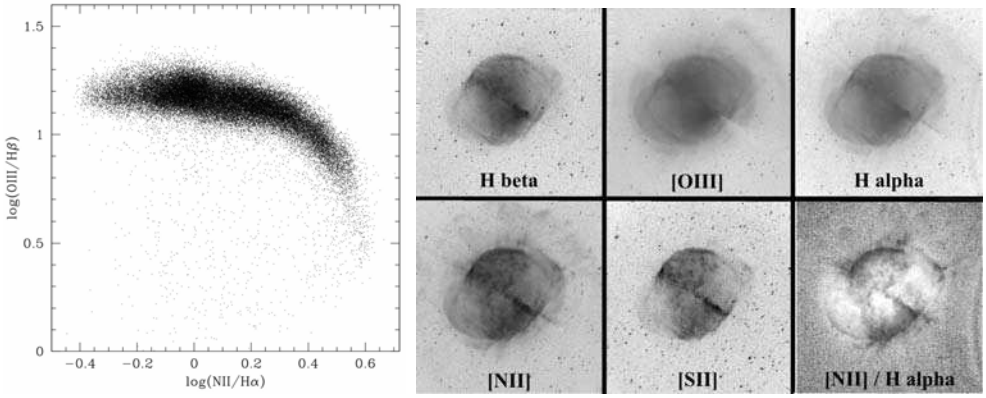


Fig. 10. (left) Diagnostic diagram of the planetary nebula M27; each point represents line ratios for a single 1.1 arcsecond pixel from the data cube. (right) Images of M27 in different emission lines, from the same data cube, showing very different morphologies. The [NII] 658.4 / Hα 656.3 line ratio map (lower right panel) displays unusually large values, characteristic of shocks, at the outskirts of the inner bubble, as well as the periphery. The identification of the individual points on the diagnostic diagram with precise location on the image provides important constraints for the modelisation of the nebula

5.2 Supernova remnants

Extended galactic supernova remnants were also prime targets for SpIOMM. The most massive stars in the universe end their life as catastrophic events. When the nuclear fuel has been entirely consumed by thermonuclear reactions, the core of the star is composed almost exclusively of iron and nickel. The outer envelope of the star collapses, bounces back on the dense nucleus and is then expelled at velocities of ~ 15 000 km/s ; this is called a supernova.

The rapidly expanding bubble of chemicaly enriched material has a profound impact on its surrounding by injecting huge amounts of kinetic energy as well as heavy elements in the interstellar medium. A classic example of such an interaction is the Cygnus Loop, a 15 000 year-old supernova remnant spanning many degrees in the sky. A tiny fraction of this object has been mapped with the camera WFPC2 on the *Hubble Space Telescope* to characterize the motion, structure and dynamical scale of the blast wave currently encountering the surrounding medium, in the northeastern part of the nebula (Blair et al. 2005). We have begun a complete mapping of the Cygnus Loop with SpIOMM is order to characterize this important object in its entirety. Fig. 11 depicts some characteristics of a single red datacube, showing the unusually strong [SII] lines. This cube illustrates that SpIOMM can be used at

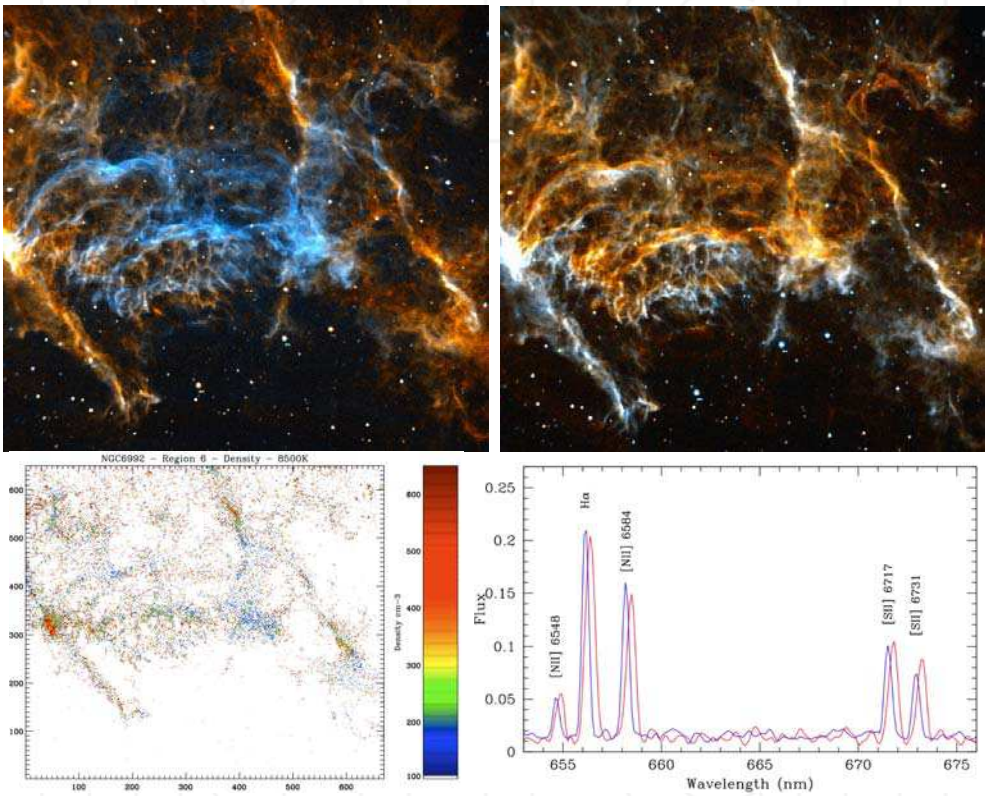


Fig. 11. Example of science results obtained with a single data cube of a section of NGC 6992, an old supernova remnant in the Milky Way, obtained with SpIOMM. The frames are 12 arcminutes on a side and display 435 000 pixels. (Upper left) - Doppler map from the [NII] 654.8 nm and 658.4 nm, Hα 656.3 nm, [SII] 671.7 nm and 673.1 nm emission lines; velocities vary between -20 km/s and + 30 km/s, the blue filaments approaching us and the red ones receiding from us. (Upper right) [NII]/Hα line ratio; Hα is orange and [NII] blue. (Lower left) - Electron density of the gas for the same region, based on the ratio of the [SII] 671.6 and [SII] 673.1 nm lines, assuming an average temperature of 8 500K. (Lower right) Spectra of two filaments (5 x 5 pixel area each) with different Doppler shifts

the same time as an imager with a set of "perfect" narrow-band filters (note in the lower right panel of Fig. 11 how well separated all the emission lines are from their neighbors) as well as a medium-resolution, very-wide field spectrograph.

The case of another, much younger supernova remnant, M1 (also known as the Crab nebula), is particularly interesting as it illustrates the full power of SpIOMM. M1 is a very young objects in Galactic terms, since it is the result of an explosion visible from Earth in 1054 AD. The initial explosion propelled the star's outer envelope at more than 10 000 k/s, but this movement was slowed down by the material surrounding the star. Nevertheless, the gas is still globally expanding today at velocities of up to ~ 1400 km/s, causing Doppler shifts of up to 3 nm. Moreover, because of the presence of shocks, the forbidden lines of [NII] 654.8 and 658.4, as well as the [SII] 671.7, 3.1 doublet are almost as strong as the (usually strongest) Hα. Therefore, up to 10 emission lines can be seen in regions where an approaching and a receding filament are superimposed on the line of sight. Charlebois et al. (2010) presented a detailed analysis of the M1 data cubes, but we reproduce in Fig. 12 some of the results. M1 has been partially mapped before with a classical, long-slit spectrograph; only a small fraction of the nebula could be mapped this way, but we have used these data to demonstrate that the spectra obtained with SpIOMM showed exacty the same features as those obtained with classical spectroscopy, but with a 100% filling factor over the entire nebula. Our data allowed us to obtain a comple tridimensional view of this intriguing object, as well as to identify a notable asymetric evolution of the oposite lobes of the nebula.

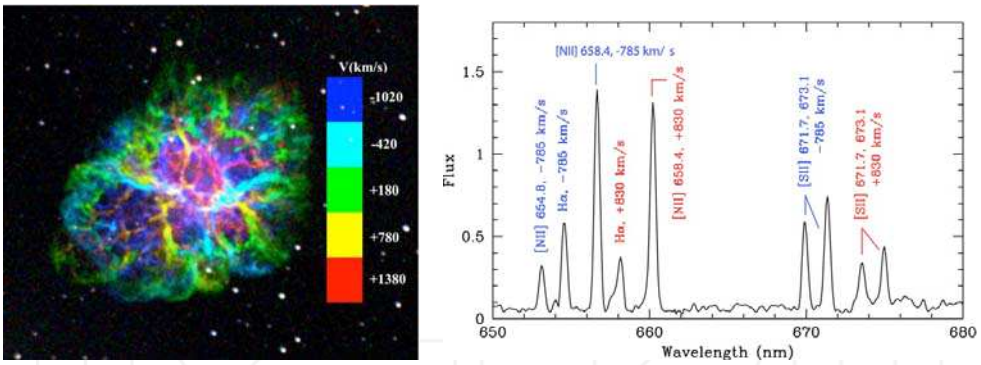


Fig. 12. A Doppler image of the Crab nebula obtained with SpIOMM as determined from the [OIII] 500.7 nm emission line, showing the rapid expansion of the filaments. On the right, the spectrum of a pixel where two filaments are superimposed. We see here two sets of lines, shifted by the Doppler effect

5.3 Galaxies

Powerful constraints on models of galactic chemical evolution, on the star formation histories of galaxies and on the dynamical processes that transform them can be derived from accurate and homogeneous determinations of chemical abundances in individual gaseous nebulae, the distribution of their stellar populations in terms of age and metallicity, and the gaseous and stellar kinematics. So far, systematic studies between chemical properties (central abundance or radial abundance gradient), and other parameters have been conducted for a small sample of spiral galaxies. The effects of the morphological type,

the presence of bars, and environment have all been studied to some extent. Moreover, most studies so far, performed with slit spectrographs, have concentrated on global properties of stellar ejecta or abundance gradients in galaxies, thereby neglecting possible small-scale variations caused by multi-phase stellar wind (individual stars) or localized enrichment by starburst clusters in peculiar evolutionary stages. SpIOMM is an ideal instrument to conduct a systematic study of abundances in nearby galaxies and thus easily detect evidence for small-scale enrichments and establish conditions under which they take place. The possibility to study the multiple emission line ratios and kinematics for hundreds of HII regions (nebulae ionized by massive stars) simultaneously in each individual galaxy is an excellent project for an IFTS. We have targetted a dozen galaxies so far, and show in Fig. 13

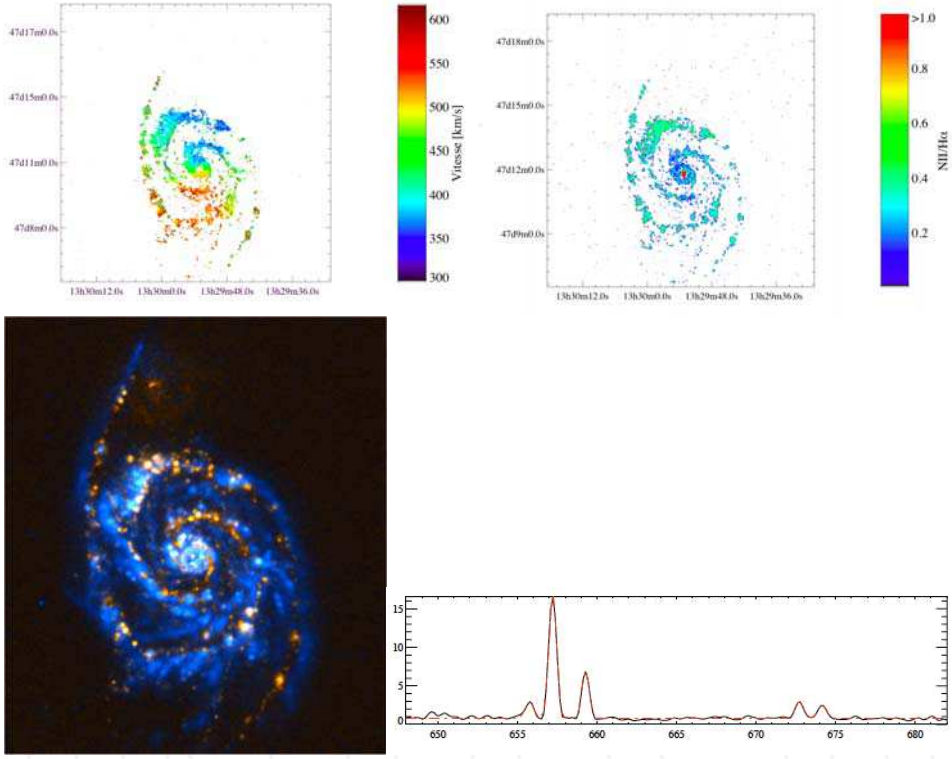


Fig. 13. (Upper left) - Velocity diagram of the spiral galaxy M51, based on the centroid of the Hα 656.3 nm emission line from an SpIOMM data cube. (Upper right) - Image of the [NII] 658.4 / Hα 656.3 line ratio from the same cube. Notice the large ratio in the core of the galaxy, characteristic of shocks driven by the central Active Galactic Nucleus. (Lower left) A comparison between the distributions of the Hα flux (characteristic of ~ 10 Myr old stellar populations, obtained from the SpIOMM data cube, in orange), and the ultraviolet flux (characteristic of ~ 100 Myr old stars, obtained with the Galex space telescope, in blue). The lag between the two stellar populations is indicative of the rotation velocity of the spiral wave pattern in the galaxy. (Lower right) Spectrum of one HII region from the SpIOMM data cube

some data on the nearby interacting spiral galaxy M51. As can be seen in the upper left panel of this figure, the galaxy is not only moving away from us (average recession velocity of 500 km/s) due to the expansion of the Universe, but it is also rotating. M51 is seen almost face-on, but a large inclination increases the velocity difference from one side to the other relative to us. Obtaining line intensities and especially line ratios with narrowband filters would almost require a set of filters per galaxy, and it would be impossible to obtain precise line ratios for edge-on galaxies. Most spectra of star-forming regions in galaxies have been obtained with single slit or multi-object spectrographs, which can only sample a small fraction of each galaxy. An IFTS is thus the instrument of choice for this kind of work.

6. Conclusions and the future

Despite lots of technical hurdles, we have clearly demonstrated that an astronomical imaging Fourier transform spectrograph working in the visible band is not only viable, but also shows an immense potential to spectrally map extended objects. We have so far discussed and presented examples of emission line spectra. Is an imaging FTS capable of obtaining spectra of continuum sources with absorption lines, such as stars? The answer is yes; indeed, the vast majority of commercial applications of IFTS are to detect absorption lines superimposed on a continuum source, and exquisite infrared spectra of stars and planets have been obtained with FTS (Ridgway & Brault 1984, Mosser et al. 1993). We have however observed only emission-line objects with SpIOMM, mostly because we used it, until recently, with only one detector and used the photometry of stars to correct for sky transparency variations. The use of the second output port will allow us to work on absorption line objects. But there is also another reason for us to favor emission-line objects, and this reason is intrinsic to the FTS concept itself. Since at every step an image of the scene in the entire waveband is obtained, contrary to Fabry-Perot systems, one of the great advantages of an IFTS is that a by-product of an interferogram is a deep panchromatic image of the scene. However, this property has also its downside: the photon noise from each wavelength is distributed among every spectral resolution element. The spectrum of an ionized nebula is completely dominated by emission lines and shows very little continuum. Photon noise therefore comes from the emission lines themselves. Moreover, the intensities of the lines we are aiming for are all within a factor of ten from each other. The FTS distributed noise disadvantage is therefore completely negligeable in this case. However, some lines of interest, such as [OIII] 436.3 nm or [NII] 575.5 nm, are extremely weak and the inclusion of much stronger lines, such as [OIII] 500.7 or Hα, in the same bandbass would lower the signal-to-noise of the faint lines and make them very difficult to detect. Once the second detector becomes available for SpIOMM, we will aim for both absorption line targets (star clusters, elliptical galaxies) and very faint lines in extended nebulae.

We have recently obtained funding to design and build an improved version of SpIOMM for the Canada-France-Hawaii telescope. SITELLE, another collaboration between Université Laval and ABB-Bomem, will have the same field of view, 12' x 12', with 0.35'' pixels, and will see its first light, if all goes according to plans, in early 2013. Because the CFHT has a larger primary mirror than the OMM telescope (3.6m vs. 1.6m) and SITELLE will from the start be equipped with two detectors, because of a better sky transparency (especially in the near UV; the CFHT is located at an altitude of 4200 m) and a better optics quality, we expect that SITELLE will offer a factor of 15 improvement over SpIOMM in terms of global efficiency. As such, it will be an ideal instrument to target very faint objects

or very faint diagnostic lines in extended objects. We expect that a fair fraction of its observing time will be devoted to the study of the star formation at cosmological distances.
As we have seen, technology has evolved enough during the past ten years on different fronts to allow us to design and build very efficient, wide-field imaging FTS in the visible range: servo systems and piezoelectric actuators for an accurate and fast correction of the mirror alignment to enhance the modulation efficiency, high quantum efficiency (~ 90%) large detectors with fast readout rate and low noise to increase the field of view and the spectral resolution (which is proportional to the maximum OPD and thus the number of CCD readouts), and of course computer power and random access memory to allow the data reduction and analysis of huge data cubes.

7. Ackowledgements

We would like to acknowledge financial contributions from the Canadian Foundation for Innovation, the Canadian Space Agency, the Natural Sciences and Engineering Council of Canada, the Fonds Québécois de la Recherche sur la Nature et les Technologies, and Université Laval. We also thank Ghislain Turcotte, Bernard Malenfant and Pierre-Luc Lévesque for their help at the telescope.

8. References

Allington-Smith, J., et al. 2002, PASP, 114, 892
Artigau, É., Bouchard, S., Doyon, R., & Lafrenière, D. 2009, ApJ, 701, 1534
Bacon, R., et al. 2001, MNRAS, 326, 23
Bacon, R., et al. 2010, *Proc SPIE*, 7335, 7B
Bennett, C. L. 2000, in Imaging the Universe in Three Dimensions, *ASP Conf. Ser*. Vol 195, p. 58
Bernath, P. F., et al. 2005, *Geoph. Res. Let*. 32, L15S01
Bernier, A.-P., Grandmont, F., Rochon, J.-F.. Charlebois, M., & Drissen, L. 2006, *Proc. SPIE*, 6269, 135
Bernier, A. -P., Charlebois, M., Drissen, L., & Grandmont, F. 2008, *Proc. SPIE*, 7014, 70147J
Blair, W. P., Sankrit, R., & Raymond, J. C. 2005, AJ, 129, 2268
Boulanger, F., et al. 2008, Experimental Astronomy, 23, 277
Chalabaev, A. A. & Maillard, J.-P. 1985, ApJ, 294, 640
Charlebois, M., Drissen, L., Bernier, A.-P., Grandmont, F., & Binette, L. 2010, AJ, 139, 2083
Colless, M. M., et al. 2001, MNRAS, 328, 1039
Daigle, O., Carignan, C., Gach, J.-L., Guillaume, C., Lessard, S., Fortin, C.-A., Blais-Ouellette, S. 2009, PASP, 121, 866
Flasar, F. M., et al. 2004, SSRv 115, 169
Friesen, R. K. et al. 2005, *MNRAS*, 361, 460
Graham, J. R. 2000, in Next Generation Space Telescope Science and Technology, *ASP Conf. Ser*., vol. 207, p. 240
Graham, J. R., et al. 1998, PASP, 110, 1205
Grandmont, F. 2007, Ph. D. Thesis, Université Laval
Grandmont, F., Drissen, L., & Joncas, G. 2003, *Proc. SPIE*, 4842, 392
Hernandez, O., et al. 2008, *PASP*, 120, 665
Huchra, J., et al. 1999, NGST NIR Spectrograph Study

Kewley, L. J., & Dopita, M. A. 2002, *ApJS*, 142, 35
Lagrois, D., & Joncas, G. 2010, *ApJ,* 700, 1847
Maillard, J. P., Crovisier, J., Encrenaz, T., Combes, M. 1987, A&A, 187, 398
Maillard, J.-P. & Simons, D. 1992, in Progress in Telescope and Instrumentation Technologies, *ESO Conference and Workshop Proceedings, ESO Conference on Progress in Telescope and Instrumentation Technologies*, ESO, p.733
Maillard, J.-P., Drissen, L., Grandmont, F., & Thibault, S. 2011, submitted to *Experimental Astronomy*.
Manset, N., & Bastien, P. 2002, AJ, 124, 1089
Monnet, G. 2009, *Experimental Astronomy*, Volume 25, p. 91
Morris, S. L., et al. 2000, in *NGST Science and Technology*, ASP Conf. Ser., vol. 207, p. 276
Mosser, B., Mekarnia, D., Maillard, J. P., Gay, J., Gautier, D., & Delache, P. 1993, A&A, 267, 604
Naylor, D. A., Gom, B. G., Tahic, M. K. & Davis, G. R. 2004, SPIE, 5498, 685
Naylor, D., et al. 2010, SPIE, 7731, 29
Paumard, T., Maillard, J.-P. & Morris, M. 2004, A&A, 426, 81
Posselt, W., et al. 2000, in NGST Science and Technology, *ASP Conf. Ser.*, vol. 207, p. 303
Ridgway, S. T. & Brault, J. W. 1984, *Ann. Rev. Astron. Astrop.* 22, 291
Ridgway, S. T., Carbon, D. F., Hall, D. N. B., & Jewell, J. 1984, *ApJS*, 54, 177
Sanchez, S. F., Becker, T., & Kelz, A. 2004, *Astron. Nachr.*, 325, 171
Scoville, N., Hall, D. N. B., Kleinmann, S. G., & Ridgway, S. F. 1979, *ApJ*, 232, L121
Stoughton, C. et al. 2002, *AJ,* 123, 485
White, G. J., et al. 2010, A&A, 518, 114
Wurtz, R., Wishnow, E., Blais-Ouellette, S., Cook, K., Carr, D., Lewis, I., Grandmont, F., & Stubbs, C. 2002a, in Next Generation Wide-Filed Multi-Object Spectroscopy, *ASP. Conf. Ser.*, Vol. 280, p. 139
Wurtz, R., Wishnow, E., Blais-Ouellette, S., Cook, K., Carr, D., Grandmont, F., Lewis, I., & Stubbs, C. 2002b, in Galaxies: The Third Dimension, *ASP. Conf. Ser.*, Vol. 282, p. 445