LAPPEENRANTA UNIVERSITY OF TECHNOLOGY
Faculty of Technology
Technomathematics and Technical Physics

LINEAR PREDICTION IN OPTICAL SPECTROSCOPY

Kolesnichenko Pavel

Examiners: Associate Professor Erik Vartiainen
Professor Tuure Tuuva

Supervisor: Associate Professor Erik Vartiainen

ABSTRACT

Lappeenranta University of Technology
Faculty of Technology
Technomathematics and Technical Physics

Author: Kolesnichenko Pavel

Title: **Linear prediction in optical spectroscopy**

Master's thesis

Year: 2014

57 pages, 50 figures

Examiners: Associate Professor Erik Vartiainen
Professor Tuure Tuuva

Keywords: cars spectroscopy, maximum entropy, mem, line narrowing, linear prediction, q-factor, lomep.

A linear prediction procedure is one of the approved numerical methods of signal processing. In the field of optical spectroscopy it is used mainly for extrapolation known parts of an optical signal in order to obtain a longer one or deduce missing signal samples. The first is needed particularly when narrowing spectral lines for the purpose of spectral information extraction. In the present paper the coherent anti-Stokes Raman scattering (CARS) spectra were under investigation. The spectra were significantly distorted by the presence of nonlinear nonresonant background. In addition, line shapes were far from Gaussian/Lorentz profiles. To overcome these disadvantages the maximum entropy method (MEM) for phase spectrum retrieval was used. The obtained broad MEM spectra were further underwent the linear prediction analysis in order to be narrowed.

ACKNOWLEDGMENTS

I am grateful to Associate Professor Erik Vartiainen for sharing his experience, constructive remarks accompanying this work, and providing me with the all necessary materials. I also wish to express my gratitude to Professor Erkki Lähderanta for supporting the Master's Degree Program.

I wish to thank my parents and my girlfriend for their support and inspiration.

TABLE OF CONTENTS

INTRODUCTION

Coherent anti-Stokes Raman scattering (CARS) technique is a powerful spectroscopy method exploiting nonlinear optical properties of different media (gases, flames, organic compounds etc.). This is a non-degenerate four-wave mixing (FWM) process giving the signal at anti-Stokes Raman frequency which proportional to third-order nonlinear susceptibility of a medium. However, CARS spectra are far from those obtained, for instance, by stimulated Raman scattering technique. There are several kinds of distortions affecting spectral background (which in general is nonlinear) as well as convenient Gaussian/Lorentz profiles of spectral lines. The nonlinear background results from the presence of nonresonant response of the medium. As a consequence, the analysis of CARS spectra becomes quite unobvious, especially when line broadening mechanisms lead to spectral lines being overlapped. Notwithstanding this situation, there exist several approaches to resolving the mentioned problems. One of such techniques is the maximum entropy method (MEM) which allows scientists to derive Raman line shapes (MEM spectra) directly from the CARS spectra [10, 14]. There are several methods of estimation nonlinear background as well, amongst which is, for example, wavelet analysis [26, 27]. Line narrowing methods are quite useful for obtaining the information on peak frequencies and relative peak's heights thus enhancing spectral resolution of obtained spectra [28]. The essence of line narrowing methods is the linear prediction (LP) algorithm. One of the main delimitations of such techniques is that they cannot properly process spectra with non-constant spectral background because the last is considered by the algorithms as the set of additional resonant frequencies that do not characterize the medium resonant quantum transitions.

The present paper is aimed to develop the LOMEP line narrowing algorithm [28] using MatLab environment and apply it to the known MEM spectra. The paper is organized as follows. The Section 2 provides the description of the CARS process including theoretical and experimental considerations. The Section 3 gives the key steps of derivation MEM spectra and introduces the spectra to be analyzed. The Section 4 explains the ideas of line narrowing techniques and describes the LOMEP procedure. The Section 5 presents the main results obtained by this method. Finally, in the Section 6 analysis of the results is made. In the Appendix section one can find the sources of the MatLab programs created for the current paper.

1. GENERAL PRINCIPLES OF CARS SPECTROSCOPY

1.1. Historical remarks

The discipline of spectroscopy originates from 1802 when the English chemist and physicist Wollaston discovered the presence of dark lines in the spectrum of sunlight [1]. Later, mainly due to Fraunhofer, Bunson, Talbot, Kirchhoff and others the discipline was further developed and elaborated. However, at that time spectroscopy was essentially an empirical science and confined itself to cataloging and classification of newly obtained spectra. New period started after Bohr has proposed his atomic theory in 1913. This theory provided the first explanation of existing spectra by introducing the concept of discrete atomic levels. From this moment spectroscopic experiments started to collaborate with atomic theory.

In 1923 the Austrian theoretical physicist Smekal first theoretically predicted the phenomenon of inelastic scattering of photons. In 1928 the Indian physicist Raman discovered the effect experimentally for which two years later he was awarded the Nobel Prize in physics. He observed that a portion of light scattered from material had a different frequency than the incident light. This effect has led to the whole new branch of spectroscopy named after the scientist.

Raman spectroscopy has been a powerful tool for studying molecular vibrations and rotations for many years. However, this technique had an essential shortcoming. Before laser has been invented Raman signal had been difficult to observe because of its weakness and low intensity [2]. As a matter of fact, differential cross-section of spontaneous Raman scattering is about 10^{-30} cm^2/sr resulting in low instrumental sensitivity and spectral resolution [3]. The invention of the laser in 1960 has revolutionized this type of spectroscopy which has become mostly nonlinear. In 1962 two physicists Woodbury and Ng have discovered infrared (IR) component in the spectrum of the light emitted by ruby laser with the nitrobenzene Kerr cell [4]. This observation revealed a new type of Raman scattering called stimulated Raman scattering.

In 1965 Maker and Terhune published the paper [5] on induced third order polarization. In the paper the process of FWM was demonstrated for the first time. However, methodical elaboration of the nonlinear spectroscopy techniques has begun only in 1971 when

frequency-tuned lasers became widely in use [6]. One of the FWM spectroscopy techniques appeared as the method of CARS.

1.2. Theoretical background

In order to describe the process of CARS let us start from the description of medium susceptibility χ. This physical quantity can be referred to the measure of how medium reacts to electromagnetic (EM) radiation. It is in one-to-one correspondence with the internal structure of medium describing both relative positions of atoms inside the unit cell of crystal lattice and different vibrational, rotational modes of molecules including dipole transitions of electrons. Depending on the intensity of light, medium response can be either linear χ_L or nonlinear χ_{NL}. The first occurs when the intensity of external EM field is much lower than the intensity of intra-atomic fields ($\sim 10^8$ V/cm) while the last takes place when we apply strong EM fields with electrical strength $\gtrsim 10^8$ V/cm (e.g., emitted by lasers). Obviously, in order to adequately describe linear and nonlinear susceptibilities one should utilize rigorous principles of quantum mechanics. However, in many cases it is sufficient to exploit simple models that reveal particular properties under consideration. Here we briefly consider the classic model of anharmonic oscillator thus revealing resonant properties of a medium.

Let us consider an isotropic medium consisting of oscillators of one kind and suggest the whole system is under strong EM radiation ($\gtrsim 10^8$ V/cm). In this case the motion of each oscillator can be described by the well-known nonlinear second-order inhomogeneous differential equation in the form

$$\frac{d^2x}{dt^2} + \Gamma\frac{dx}{dt} + \omega_0^2 x + \sum_{j=2}^{\infty} \beta_j x^j = -\frac{e}{m}E, \tag{1}$$

where m is the mass of electron, e is the electron charge, Γ is the damping parameter, β_j is the normalized nonlinear spring coefficients, ω_0 is the damped oscillator's natural frequency and $E = E_0 e^{-i\omega t}$ is the complex electric field [7]. Under the assumption of weak nonlinearities, following [7], we can immediately write the solution of (1) in the form of series expansion

$$x = \sum_{l=1}^{\infty} x_l, \tag{2}$$

where

$$\begin{aligned} x_1 &= -\frac{e}{m}\frac{E_0 e^{-i\omega t}}{\omega_0^2 - \omega^2 - i\Gamma\omega}, \\ x_2 &= \frac{e^2}{m^2}\beta_2 \frac{E_0^2 e^{-2i\omega t}}{(\omega_0^2 - 4\omega^2 - 2i\Gamma\omega)(\omega_0^2 - 4\omega^2 - i\Gamma\omega)^2}, \\ x_3 &= -\frac{e^3}{m^3}\left[\frac{2\beta_2^2}{\omega_0^2 - 4\omega^2 - 2i\Gamma\omega} + \beta_3\right]\frac{E_0^3 e^{-3i\omega t}}{(\omega_0^2 - \omega^2 - i\Gamma\omega)(\omega_0^2 - 9\omega^2 - 3i\Gamma\omega)}, \\ &\ldots \ . \end{aligned} \tag{3}$$

Defining polarization of a medium $\boldsymbol{P}$ as the vector sum of the oscillators' dipole moments $N\boldsymbol{p} = -Ne\boldsymbol{x}$ per unit volume V we find the following series expansion for the absolute value P

$$P = -\frac{N}{V}ex = -ne\sum_{l=1}^{\infty} x_l = \chi^{(1)}E + \chi^{(2)}E^2 + \chi^{(3)}E^3 + \cdots, \tag{4}$$

where $\chi^{(1)}$ is the linear susceptibility of a medium, $\chi^{(i)}$, $i = 2,3,\ldots$, are the nonlinear susceptibilities of i^{th} order and n is the concentration of oscillators in the medium. From (4) one can directly conclude that the second term corresponds to a second-order nonlinear process which results in generating new frequencies being linear combinations of two incident frequencies (three-wave mixing processes). At the same time the third term describes the third-order nonlinear processes that generate new frequencies as linear combinations of three initial frequencies (four-wave mixing processes). At this point it is worth noting that the second term in (4) is vanished for noncentro-symmetrical systems (that are dominated over centrosymmetrical ones in nature) leading to $\chi^{(3)}$ being the first nonlinear term. Expressions (3)–(4) exhibit resonant behavior of susceptibilities and of a medium in general.

In order to find exact expressions for the susceptibilities one should apply quantum mechanical formalism of density matrices [4]. Furthermore, in the general case of anisotropic medium and EM radiation consisting of more monochromatic components

$\omega_1, \omega_2, \omega_3, \ldots$ it appears that susceptibilities are of the tensor form ($\chi_{ij}^{(1)}(\omega_1, \omega_2, \omega_3, \ldots)$, $\chi_{ijk}^{(2)}(\omega_1, \omega_2, \omega_3, \ldots)$, $\chi_{ijkl}^{(3)}(\omega_1, \omega_2, \omega_3, \ldots)$ etc.). In addition, quantum mechanical approach reveals nonresonant terms in the susceptibilities as well as resonant ones. Thus, nonlinear susceptibilities can be written as the sum of resonant and nonresonant terms, for instance,

$$\chi_{ijkl}^{(3)}(\omega) = \left(\chi_{NR}^{(3)}\right)_{ijkl} + \left(\chi_{R}^{(3)}\right)_{ijkl}(\omega). \tag{5}$$

There are generally 48 terms in the expression for $\chi_{ijkl}^{(3)}$ (for each of 3^4=81 elements) [4]. For isotropic medium the tensor $\chi_{ijkl}^{(3)}$ has 21 (out of 81) nonzero elements and only 3 of them are linearly independent [3]. The element $\chi_{1111}^{(3)}$ can be considered as one of them. It corresponds to the linearly polarized vibrations of a medium and therefore is often observed by means of linearly polarized laser beams.

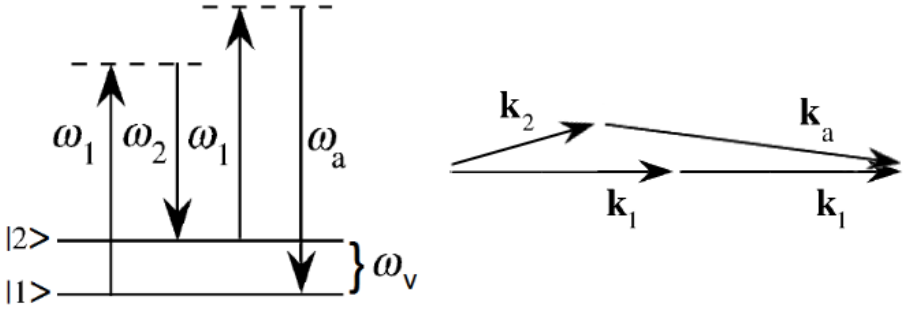

Fig.1. CARS process: energy (left) and momentum (right) conservation laws [8].

Now we can directly proceed to the process of CARS. This process is generally a non-degenerate FWM process. However, the degenerate CARS process shown in the Fig.1 is implemented more frequently. The process assumes two incident pump fields at frequencies ω_1 and $\omega_2 < \omega_1$ forming together the signal at the beat frequency $\omega_1 - \omega_2 = \omega_v$ of a molecular vibrational mode. The latter signal coherently excites medium increasing its energy from level $|1\rangle$ to level $|2\rangle$ by the value $\hbar(\omega_1 - \omega_2)$. Before relaxation of a medium has occurred, a photon from the probe beam with the energy $\hbar\omega_1$ excites the medium to a virtual energy level[*]. According to the quantum mechanical uncertainty principle the last process is not forbidden even though it is instant meaning that the relaxation of the medium to lower states occurs instantly. There is a nonzero probability of the medium to lose its energy down to the level $|1\rangle$. In this case an anti-Stokes photon with

[*] Generally speaking, this level must not necessarily be virtual one.

the higher energy $\hbar\omega_{\mathrm{a}}$ appears. One could consider this photon being a spontaneously emitted in a random direction one. However, this is not the case. Speaking in the terms of classical optics, the EM field at the frequency ω_1 is scattered not by the random thermal fluctuations, but by the regular diffraction grating formed inside crystal. This grating is induced by the biharmonic wave at the frequency $\omega_1 - \omega_2$. As a consequence, for a given set of input parameters (angles of incidence, frequencies ω_1 and ω_2) there exists only one direction of constructive interference that satisfies Bragg condition [6]. Thus, we can write for the CARS process both conservation energy law and conservation momentum law (phase matching condition) as the following

$$\omega_{\mathrm{a}} = 2\omega_1 - \omega_2, \tag{6}$$

$$\boldsymbol{k}_{\mathrm{a}} = 2\boldsymbol{k}_1 - \boldsymbol{k}_2. \tag{7}$$

As was mentioned above, the process of CARS is the FWM process and therefore described by the third-order nonlinear susceptibility $\chi_{\mathrm{a}}^{(3)} = \chi^{(3)}(\omega_{\mathrm{a}} = 2\omega_1 - \omega_2)$. The intensity of anti-Stokes component observed in the experiment is proportional to $\left|\chi^{(3)}(\omega_{\mathrm{a}} = 2\omega_1 - \omega_2)\right|^2$:

$$I_{\mathrm{a}}(\omega_{\mathrm{a}}) \sim \left|\chi^{(3)}(\omega_{\mathrm{a}} = 2\omega_1 - \omega_2)\right|^2 I_1^2 I_2 \left| \frac{\sin\left(\frac{\Delta k L}{2}\right)}{\frac{\Delta k}{2}} \right|^2, \tag{8}$$

where L is the length of the nonlinear layer and $\Delta k = k_{\mathrm{a}} - 2k_1 + k_2$ [6]. According to (7) and (8) the intensity I_{a} has its maximum value whenever the phase matching condition $\Delta k = 0$ takes place. The typical CARS line shape is shown in the Fig.2. This kind of line shape is obtained, for example, by scanning the nonlinear medium with the field at the tunable frequency ω_1. Similar result was obtained by Levenson et al. and published in 1972 [9]. It is notable that the obtained line shape was far from Lorentzian line shape. This is because of presence of nonresonant contribution from the electronic response of molecules in the third-order susceptibility $\chi_{\mathrm{a}}^{(3)}$ (see (5)) which leads to the additional nontrivial background:

$$\left|\chi_{\mathrm{a}}^{(3)}\right|^2 \sim \left|\left(\chi_{\mathrm{a}}^{(3)}\right)_R + \left(\chi_{\mathrm{a}}^{(3)}\right)_{NR}\right|^2 = \left|\left(\chi_{\mathrm{a}}^{(3)}\right)_R\right|^2 + 2\left(\chi_{\mathrm{a}}^{(3)}\right)_{NR} Re\left\{\left(\chi_{\mathrm{a}}^{(3)}\right)_R\right\} + \left(\chi_{\mathrm{a}}^{(3)}\right)_{NR}^2, \quad (9)$$

where

$$\left(\chi_{\mathrm{a}}^{(3)}\right)_R \sim \frac{A_{\mathrm{v}}}{\omega_{\mathrm{v}} - (\omega_1 - \omega_2) - i\Gamma_{\mathrm{v}}}. \quad (10)$$

Here A_{v} is the amplitude of a vibrational Raman mode and Γ_{v} is its line-width [10].

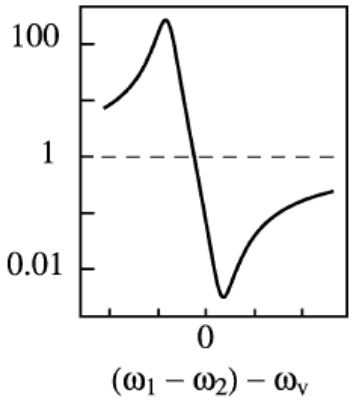

Fig.2. Typical CARS single line shape [8].

1.3. Experimental setup

With the development of the techniques for generating ultra-short pulses the CARS spectroscopy methods have become more efficient. It is the well-known fact that the shorter the pulse, the more continuously distributed frequencies it consists of. Thus, in the context of CARS spectroscopy by means of ultra-short pulses one can cover several vibrational molecular modes at once without scanning procedure (multiplex CARS). As a matter of fact, there exist several different modifications of the experimental setups for CARS spectroscopy. Each concrete modification can be designed for specific purposes. The example of an experimental setup is given by Levenson et al. in [9]. They have used the second harmonic of a Q-switched Nd:glass laser and continuously tunable narrow-band dye laser in order to obtain the line shape as that in Fig.2. Rotational modes can be studied with the other modification of experimental setup presented in [11]. Another example of an experimental setup applicable for studying vibrational molecular modes is shown in Fig.3. The most common laser system for CARS spectroscopy was the pair of a dye laser and a

Nd:YAG laser with pulse repetition rate of approximately 10 Hz and pulse duration ~0.1–10 ns. The Nd:YAG laser operated in broadband mode and was coupled with a shutter (e.g., acousto-optic Q-switch, electro-optical Kerr cell switch etc.) in order to generate short pulses at the pump/probe frequency ω_1. Today Ti:sapphire lasers can generate ultra-short femtosecond (fs-) pulses that have much broader bandwidth than nanosecond (ns-) and picosecond (ps-) pulses. Instead of relatively cumbersome dye lasers optical parametric oscillators (OPO) and optical parametric amplifiers (OPA) are used in order to obtain different frequencies. Detection system often consists of a high-resolution spectrometer and a CCD-camera. Special optics is needed for the high-energy pulses passing through the optical elements. Filters are necessary in order to block input beams and probable fluorescence appearing in a sample under observation.

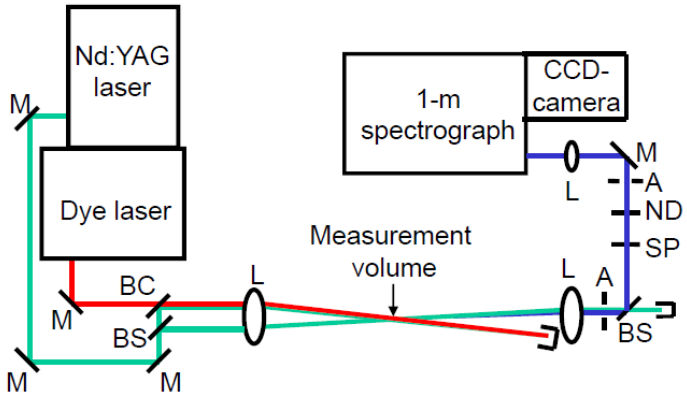

Fig.3. Experimental setup of vibrational CARS spectroscopy [12]. ‘M’ denotes mirror, ‘BC’ – beam combiner, ‘BS’ – beam splitter, ‘A’ – aperture, ‘ND’ – neutral density filter, ‘CCD’ – charge-coupled device, ‘L’ – lens, ‘SP’ – short-pass filter.

In [9], [11] and [12] noncollinear geometries of CARS experimental setups are presented. However, the most common geometry for CARS spectroscopy is the collinear one (Fig.4).

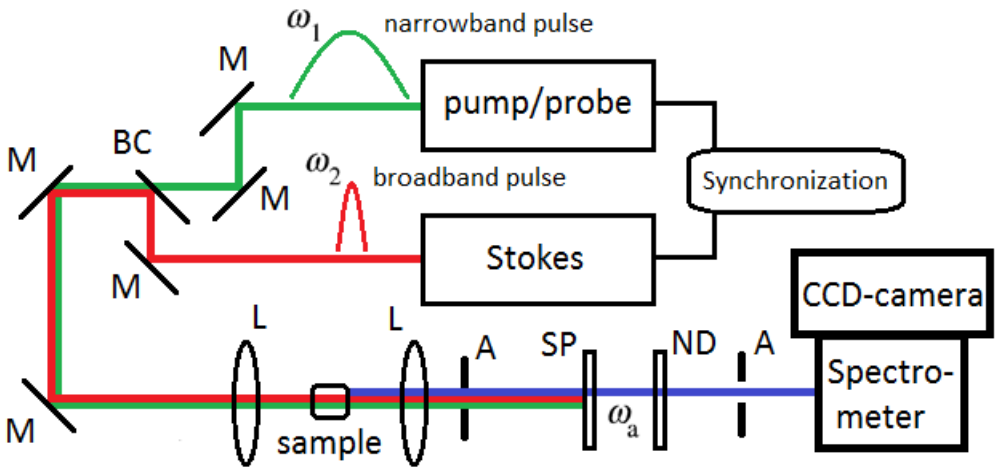

Fig.4. Collinear geometry of a vibrational CARS experimental setup.

Let us discuss this issue more carefully. Mathematically there are several ways to satisfy phase-matching condition (7). All of them can be separated in a few groups [12]. These are collinear phase-matching, planar boxcars phase-matching and folded boxcars phase-matching (Fig.5). When wave vectors participating in the process are all parallel to each other then the collinear phase matching is implemented. Planar boxcars phase matching takes place whenever all wave vectors are coplanar. Finally, folded boxcars phase matching implies wave vector $\boldsymbol{k}_1$ belonging to one plane while both $\boldsymbol{k}_2$ and $\boldsymbol{k}_a$ are belonging to another plane. It is shown in the paper [16] that CARS interaction volume depends not only on the focusing parameters but in addition on the type of phase matching used. In the context of CARS spectroscopy each of these implementations can be revealed in a specific type of measurement. Thus, in the paper [13] the folded boxcars case has been employed in order to obtain pure rotational CARS spectra to within a few wavenumbers of the incident pump frequency. Collinear phase matching is well implemented in gases where dispersion within a small interval of frequencies is negligible [2-3]. In liquids dispersion is much higher and the phase matching condition is well implemented whenever beams are crossing each other at a certain angle (that satisfies the phase matching condition) [2].

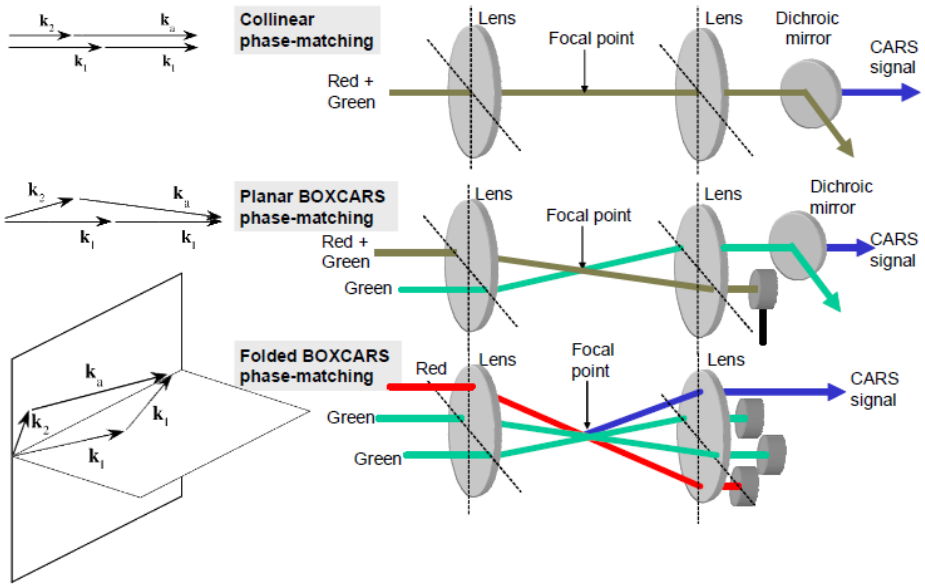

Fig.5. Different implementations of phase-matching condition [10].

The data (see Section 3.2) for the current work were obtained by means of collinear experimental CARS scheme similar to that as in Fig.4. The pump/probe laser operating at 710 nm emitted 10-ps pulses each having the bandwidth of 1.5 cm^{-1}. This laser determined the spectral resolution. The second (Stokes) laser was frequency-tunable within the range of 700 – 1000 nm thus allowing experimenters to cover the vibrational range of 950 – 3150 cm^{-1}. This laser emitted 80-fs pulses with the bandwidth of 184 cm^{-1}.

1.4. Advantages and disadvantages

Speaking of the advantages of CARS spectroscopy, authors (for instance, [2-3]) commonly highlight the following ones:

1. The main advantage with respect to spontaneous Raman spectroscopy is the high intensity of anti-Stokes signal. For comparison, its magnitude can be 10^4–10^5 times bigger than that of spontaneously scattered Raman signal.
2. CARS beam is significantly narrow and propagates in the solid angle of order ~10^{-6} sr. This allows experimenters to discriminate anti-Stokes signal from fluorescent background or thermal radiation (especially when studying plasma).
3. Scattering volume leading to appearance an anti-Stokes component is essentially around a focal point. Such a volume is of order ~10^{-6} cm^3 and consists of ~10^7 molecules. Therefore it is enough to investigate small amounts of a sample.
4. It is relatively easy to filter anti-Stokes signal owing to its bigger frequency.

5. By means of ultra-short pulses it is possible to obtain high spectral resolution (0.3–0.001 cm^{-1}) and also observe processes having fast relaxation times.

Owing to these advantages CARS spectroscopy methods are extremely effective when studying flames, gas discharges, burning processes as well as biological and organic samples. With the aid of fs-pulses it is possible to investigate fast processes and vibrational dynamics of molecular systems in gaseous, liquid and solid media.

The main drawbacks of CARS spectroscopy are the relatively high prices of experimental equipment and strong fluctuations of the CARS signal caused by instabilities both in beams adjustment and pulses' intensities [2]. With the aid of pure CARS signal it is impossible to directly obtain Raman spectra because of the presence of nonlinear nonresonant contribution in the third-order susceptibility which results in both high distortions of line shapes and presence of nonresonant background in output spectra. In the CARS spectroscopy the absolute value $\left|\chi^{(3)}\right|^2$ is measured, therefore it is impossible to directly obtain phase spectra as well. The last is an issue of high importance because, as was shown in a series of papers [10, 14–15], in the context of CARS spectroscopy phase spectra are directly proportional to corresponding Raman spectra containing information on resonant components of molecules. In the next section the maximum entropy principle is applied in order to extract phase function from obtained CARS spectra. It is possible due to the interference between resonant part of FWM CARS process and nonresonant background which results in the implicit record of phase information within CARS spectra.

2. MAXIMUM ENTROPY METHOD

2.1. Introductory remarks

In 1948 Shannon published the paper "A Mathematical Theory of Communication" [17] where he introduced the quantitative measure of information in the form of entropy. Generally speaking, entropy is the measure of either chaos or uncertainty. The more chaos (randomness or uncertainty) is in a system under consideration, the greater value of entropy the system has, and vice versa. Shannon used the concept to measure information uncertainty particularly in cryptography. However, entropy is widely used in thermodynamics, statistical mechanics, cosmology, astrophysics, probability theory, philosophy, as well as in different applications such as natural language processing [18-19] and additional spectroscopic information extraction [7], [10, 14, 20-23]. According to the second law of thermodynamics (law of nature) entropy of a system left to its own resources can only be increased. This is called the maximum entropy (ME) principle. In this section the ME principle is used in order to extract phase information from the CARS signal. The fundamental idea is to find the ME model for the finite set of spectral data, i.e. to fit known samples with the new ones giving the most random corresponding time samples (discrete Fourier transform, DFT), whose autocorrelation values are in agreement with the given data set [23]. Mathematically it can be done by exploiting the methods of variation calculus where the entropy functional (which might include additional constraints) is maximized. The idea has been proposed by Burg [24].

It is worth noting that MEM is not the only theoretical method for phase retrieval. Another method is based on the well-known Kramers-Kronig (KK) relations. However, these relations cannot be applied with sufficient confidence whenever one deals with finite and discrete set of experimental data. Moreover, in nonlinear optics MEM covers a wider subset of problems than the KK relations [7, 10, 23].

2.2. Theoretical background

Following the papers [7], [10], [14], [21] and [23], let us represent here the most important steps of the phase function ME model derivation process. Let x_m, $0 \leq m \leq M$, be a measured experimental equidistant (with a distance Δt) time series and $C(m) = \epsilon[x_n, x_{n+m}]$ be a corresponding autocorrelation function (ϵ is the mathematical expectation

operator). Then $C(m)$ and the power spectrum $S(\nu)$ are connected through the following Fourier transforms:

$$S(\nu) = \sum_{m=-\infty}^{\infty} C(m) e^{-i2\pi m\nu}, \tag{11}$$

$$C(m) = \int_0^{\frac{1}{\Delta t}} S(\nu) e^{i2\pi m\nu}\, d\nu. \tag{12}$$

One can obtain a power spectrum estimate by applying a Fourier transform to $w(m)C(m)$ under the assumption that $C(m) = 0, |m| > M$; $w(m)$ is a window function. Next, one should apply the Burg's MEM [24] mentioned above. The output is the maximum entropy model for $S(\nu)$. As a starting point for phase retrieval procedure, one has to maximize the entropy rate

$$h \sim \int_0^1 \log S(\nu)\, d\nu \tag{13}$$

under the constraint

$$C(m) = \int_0^1 S(\nu) e^{i2\pi m\nu}\, d\nu\,, |m| \le M, \tag{14}$$

i.e. to solve

$$\frac{\partial h}{\partial C(m)} = 0, |m| > M. \tag{15}$$

The solution of (15) can be written as the power spectrum estimate, namely

$$S(\nu) = \frac{1}{\sum_{k=-M}^{M} a_k e^{-i2\pi k\nu}}, \tag{16}$$

where

$$a_k = \int_0^1 \frac{1}{S(\nu)} e^{i2\pi k\nu}\, d\nu\,, |k| \le M. \tag{17}$$

Now, one needs to substitute (16) into (14), but for $0 \le m \le M$. This operation corresponds to obtaining such coefficients a_k for which (16) would be in accordance with

the known autocorrelation values $C(m)$. The methods of complex analysis lead to the set of linear equations which can be represented using Toeplitz matrix as

$$\begin{pmatrix} C(0) & C(-1) & \cdots & C(-M) \\ C(1) & C(0) & \cdots & C(1-M) \\ \vdots & \vdots & \ddots & \vdots \\ C(M) & C(M-1) & \cdots & C(0) \end{pmatrix} \begin{pmatrix} 1 \\ c_1 \\ \vdots \\ c_M \end{pmatrix} = \begin{pmatrix} b_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{18}$$

where b_0 is real and c_k, $1 \le k \le M$, are complex coefficients. It appears that ME model (16) can be written in the terms of these coefficients as

$$S(\nu) = \frac{b_0}{|1 + \sum_{k=1}^{M} c_k e^{-i2\pi k\nu}|^2}, \tag{19}$$

where c_k and b_0 can now be calculated from (18).

It can be shown [25] that there is one-to-one correspondence between the ME model (19) and the autoregressive (AR) model of a stationary stochastic signal (time response function) $x(t)$

$$x_m = -\sum_{k=l}^{M} c_k x_{m-k} + e_m, \tag{20}$$

where e_m is a random error. By taking a DFT of (20) we obtain the following expression for the complex spectral response

$$f(\nu) = \frac{E(\nu)}{1 + \sum_{k=l}^{M} c_k e^{-i2\pi k\nu}}, \tag{21}$$

where $f(\nu) = \mathcal{F}\{x(t)\}$, $E(\nu) = \mathcal{F}\{e(t)\} = |\beta| e^{-i\phi(\nu)}$ is the error spectrum, $\phi(\nu)$ is the phase of the error spectrum, which cannot be defined from measured $|f(\nu)|$. As follows from (21), the phase retrieval process can be reduced to the estimation of the error phase function $\phi(\nu)$. The denominator of (21) can be presented as $1 + \sum_{k=l}^{M} c_k e^{-i2\pi k\nu} = |A_M(\nu)| e^{-i\psi(\nu)}$, where $\psi(\nu)$ is called MEM phase. True phase of a CARS signal can be shown to be the sum of the MEM phase function and the error phase function. It appears that for the most cases of CARS spectroscopy the error phase function $\phi(\nu)$ is almost zero and the MEM phase $\psi(\nu)$ coincides with the true phase spectrum.

According to (9), nonlinear cubic susceptibility registered in a CARS experiment consists in general of three terms including quadratic combinations of resonant and nonresonant components which make obtained CARS spectra difficult to analyze. The resonant part of the susceptibility in complex compounds can be represented as the sum of (10). Such a sum can be compared with the corresponding spontaneous Raman signal

$$I_{SRS}(\omega) \sim \sum_j \frac{A_j \Gamma_j}{\omega_j - \omega + \Gamma_j^2}, \tag{22}$$

where A_j, ω_j and Γ_j are the amplitude, the frequency and the line width of j^{th} Raman mode respectively. From this comparison, one can derive the Raman spectra to be

$$\mathrm{Im}\chi_R^{(3)}(\omega_a) \approx \left|\chi_{1111}^{(3)}(\omega_a)\right| \sin(\theta'), \tag{23}$$

where θ' is the estimate of the $\chi_{1111}^{(3)}$ phase. It is seen from (23) that there is still nonresonant background in the Raman spectra. Taking into account the presence of nonresonant susceptibility leads to the following expression for the phase away from resonances:

$$\theta(\omega) = \tan^{-1}\left[\frac{\mathrm{Im}\chi_R^{(3)}(\omega)}{\chi_{NR}^{(3)} + \mathrm{Re}\chi_R^{(3)}(\omega)}\right]. \tag{24}$$

The diagram showing the phase of nonlinear susceptibility on the complex plane is presented in the Fig.6. Whenever experimentalists deal with relatively simple samples (such as gases or flames) there is almost constant nonresonant background in the obtained spectra. As can be seen from the Fig.6, in this case $\chi_R^{(3)}$ is almost equaled to $\mathrm{Im}\chi_{1111}^{(3)}$ which exactly coincides with the corresponding Raman spectrum. However, for the complex organic compounds this is not the case and phases' ME models become extremely useful.

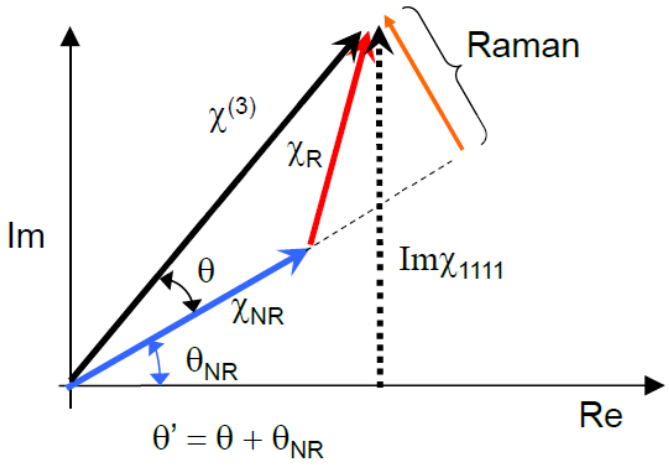

Fig.6. Representation of nonlinear susceptibility on the complex plane [29].

2.3. Application of MEM to experimental and theoretical spectra

In the present paper different CARS spectra were used to be analyzed. The pictures below present the CARS spectra (left) and the corresponding MEM phase spectra (right) of water (Fig.7), water solutions of sucrose (Fig.8 and Fig.9), simulated $\chi^{(3)}$ without background (Fig.10) and with non-constant background (Fig.11), and equimolar (500 mM) mixture of adenosine mono-, di- and triphosphate nucliotides (AMP, ADP and ATP) in water (Fig.12).

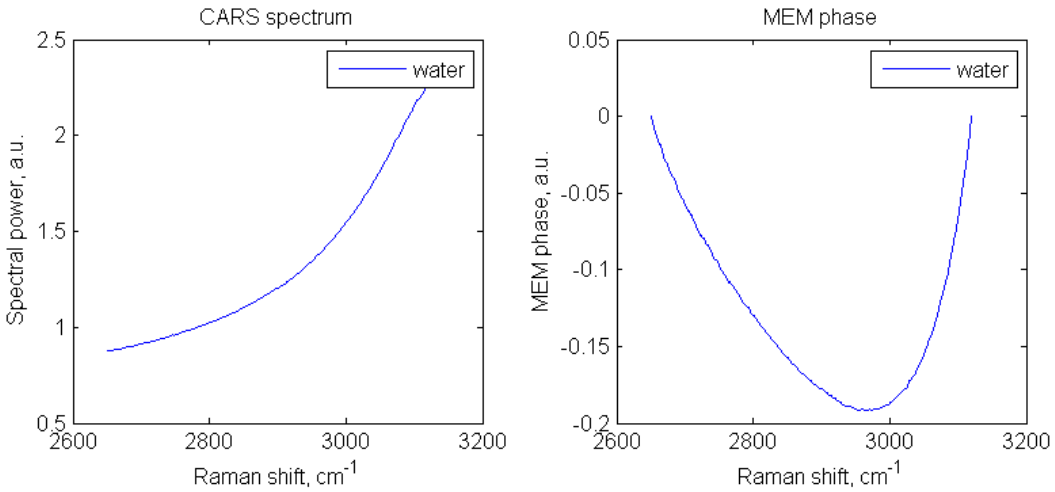

Fig.7. Spectra of water without sucrose.

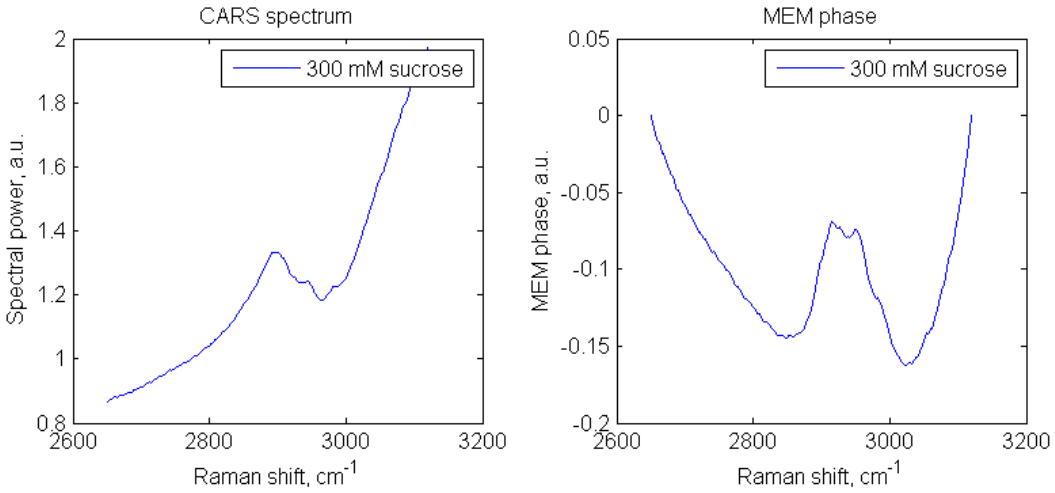

Fig.8. Spectra of water solution of 300 mM sucrose.

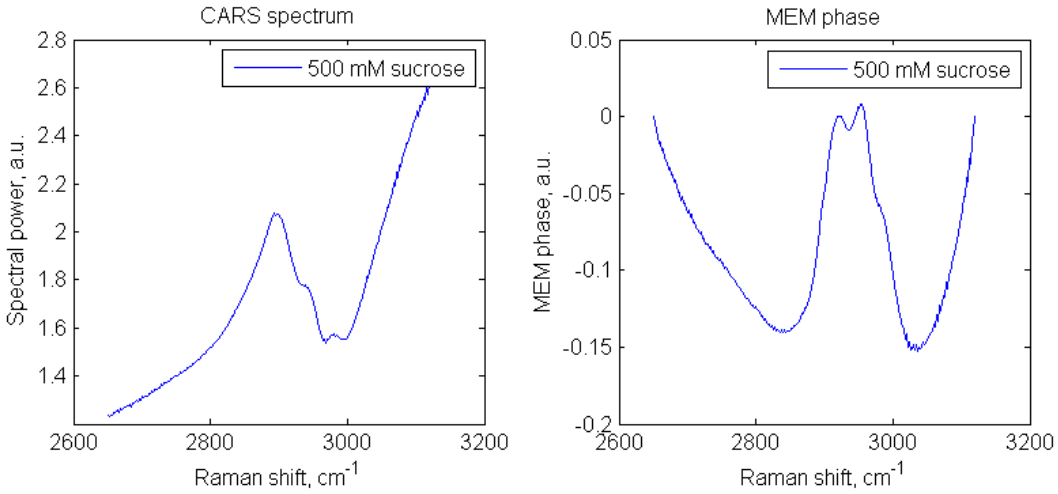

Fig.9. Spectra of water solution of 500 mM sucrose.

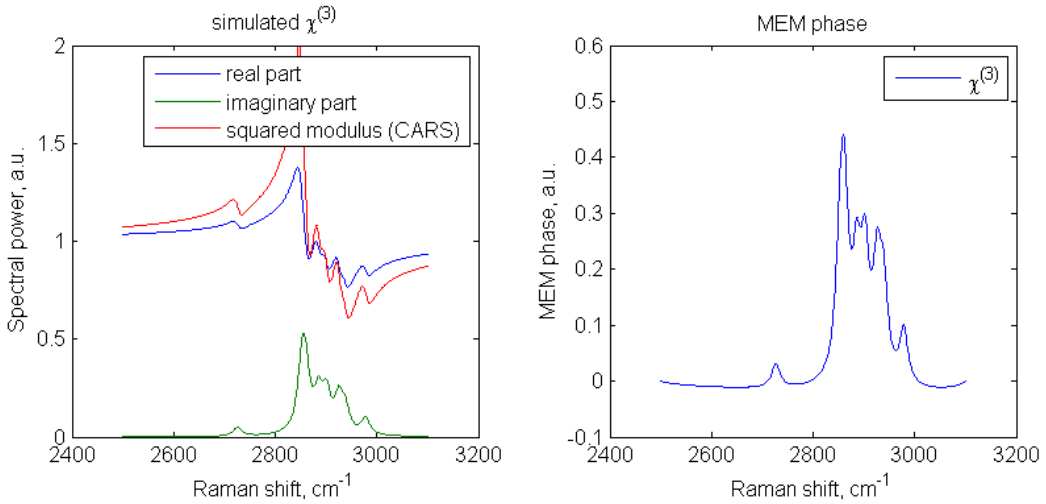

Fig.10. Simulated $\chi^{(3)}$ spectra with constant background.

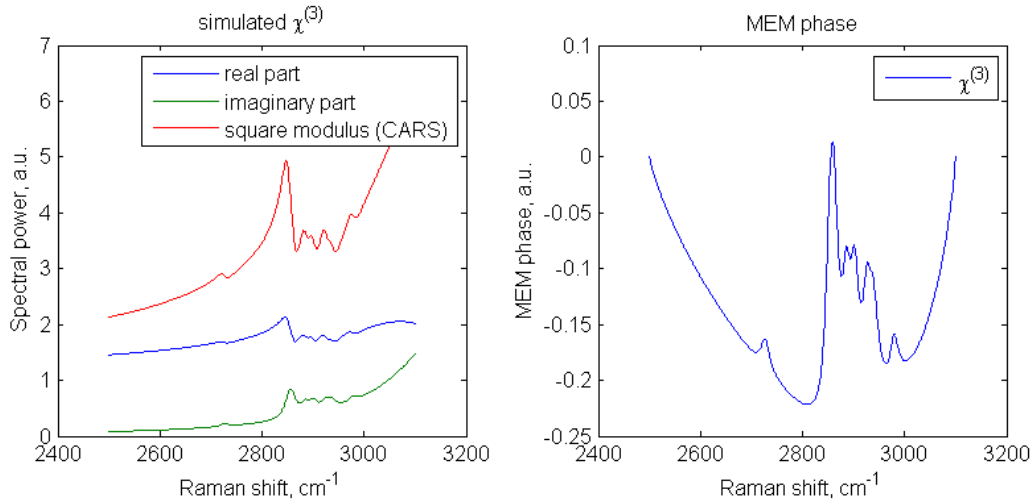

Fig.11. Simulated $\chi^{(3)}$ spectra with non-constant background.

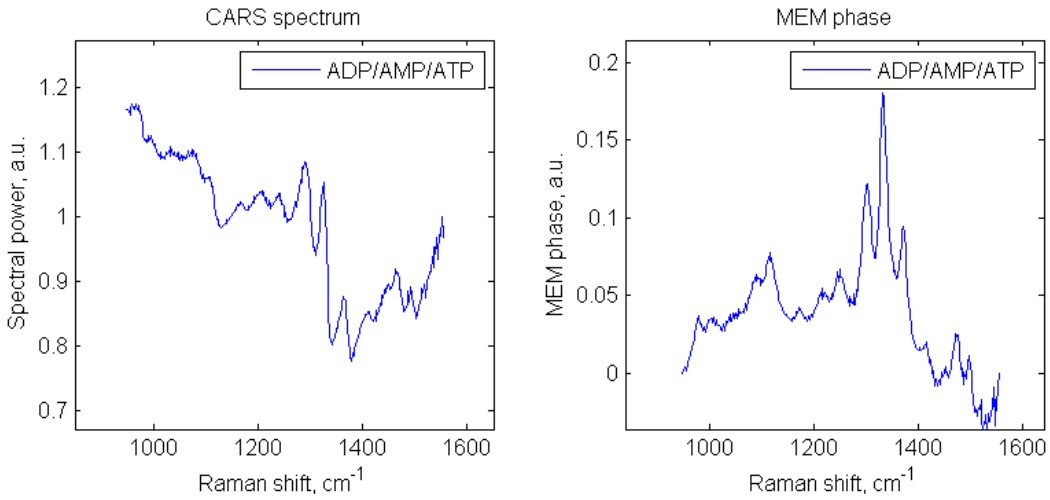

Fig.12. Spectra of ADP/AMP/ATP.

It is seen from the figures above that the MEM phase spectra are similar to those of Raman scattering processes, even though they still need to be analyzed for significant nonlinear background.

3. LINE NARROWING PROCEDURE

3.1. Introductory remarks

Line broadening is one of the problems experimentalists encounter when measuring CARS spectra. This is in addition to the presence of slow varying nonresonant background and line shape distortions. The last problem can be solved by means of MEM algorithm briefly described in the Section 3. Nonresonant background can be estimated, for instance, using the MEM approach coupled with wavelet analysis [26-27]. Another approach to the problem could be the procedure of line narrowing which narrows spectral lines to a less FWHM. After this is accomplished one can easily correct the estimate of slow varying spectral background. In addition, narrowed spectra are more informative meaning knowledge about the number of spectral lines and their relative peak spectral powers. Thus, line narrowing methods become extremely important for the spectral analysis of CARS spectra.

3.2. Fourier self-deconvolution (FSD)

The fundamental idea of any line narrowing method is the fact that the broader the pulse, the narrower corresponding spectral lines it has. Therefore first of all one should convert the spectrum (discrete set of frequencies) under consideration to the corresponding radiation (time series). This is always done by the Fourier transform. In the current paper the inverse DCT was exploited in the following form:

$$x_k = \sqrt{\frac{1}{N}} f_0 + \sqrt{\frac{2}{N}} \sum_{j=1}^{N-1} f_j \cos\left(\frac{\pi}{2N}(j-1)(2k-1)\right), k = 0, \dots, N-1, \tag{25}$$

where N is the number of spectral samples, f_j is the j^{th} spectral sample and x_k is the k^{th} time sample. The corresponding direct DCT is

$$\begin{gathered} f_k = \sqrt{\frac{2}{N}} \sum_{j=0}^{N-1} x_j \cos\left(\frac{\pi}{2N}(2j-1)(k-1)\right), k = 1, \dots, N-1, \\ f_0 = \sqrt{\frac{1}{N}} \sum_{j=0}^{N-1} x_j . \end{gathered} \tag{26}$$

These forms are straight and do not assume N to be equaled to 2^n. After the time series x_k has been calculated by using (25), the next step should be to make the signal broader. This can be done by the FSD procedure (which corresponds to deapodization in the time domain) implying the factorization of the original signal by the inverse Fourier transform of a window function $W(f)$ [28]. Let us consider this issue more carefully. Following [28], let a measured spectrum be of the form

$$H_w(f) = W(f) * H(f) = \int_{-\infty}^{\infty} W(f')H(f - f')df', \tag{27}$$

where $H(f)$ is the real spectrum, $W(f)$ is an instrumental window function. The spectrum $H(f)$ can be obtained by deconvolution if $W(f)$ is defined. However, deconvolution (in the frequency domain) is rather difficult to implement, therefore deapodization† (in the time domain) is used more frequently. Thus, one should take Fourier transform of (27) which leads to the following equation:

$$\mathcal{F}\{H_W(f)\} = \mathcal{F}\{W(f)\}\mathcal{F}\{H(f)\} \tag{28}$$

Finally, the real spectrum can be obtained from (28) by taking inverse Fourier transform as

$$H(f) = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{H_W(f)\}}{\mathcal{F}\{W(f)\}}\right\}. \tag{29}$$

In the line narrowing process self-deconvolution is used. In this case spectrum is deconvolved by the line shape of the spectrum itself [28]. Let $W_R(f)$ be a registered single line shape. As follows from (29),

$$\mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{W_R(f)\}}{\mathcal{F}\{W_R(f)\}}\right\} = \mathcal{F}^{-1}\{1\} = \delta(f). \tag{30}$$

Thus, in the ideal case of the single line shape without noise the FSD procedure leads to the Dirac's delta function which has an infinitely small FWHM. However, the real spectra are always noisy requiring the additional smoothing operation by multiplying the corresponding spectrum by some smoothing window function. Moreover, measured spectra lie within the finite range of frequencies giving sinc-function instead of the delta function in (30). Finally, the spectra are usually consisted of several spectral lines. Fortunately, the FSD procedure is a linear operation meaning that the deapodization of the

† Further 'deapodization' and 'deconvolution' will be used as synonyms.

signal composed of several frequencies implies the narrowing of each spectral line separately. It does not affect the relative heights of the spectral lines as well as their positions and areas under the corresponding line shapes.

The FSD procedure affects the tails of the signal more than its core, in a gradient-like way, resulting in the increase of intensity, particularly of the tails. As a result, the signal may seem to be wider and the corresponding spectral lines narrower. The problem is what kind of window function would be more appropriate in order to obtain better results. In this paper the Lorentz function was used as the most relevant to physical considerations. As a matter of fact, there are two general types of spectral line broadening. These are homogeneous (giving Lorentz line shape) and inhomogeneous (giving Gaussian line shape) broadening processes. Of these, there are different mechanisms inside media leading to one of the two line broadening types. One of the main contributions to inhomogeneous line broadening is the big velocities of molecules in gases. In this paper the nongaseous media are under consideration and therefore homogeneous broadening dominates inhomogeneous one.

The inverse Fourier transform of the Lorentz function is the exponentially decaying function $e^{-x/b}$. In the numerical simulations parameter b was varied in order to obtain the most adequate results. This parameter corresponds to the FWHM of the Lorentz function in the frequency domain. The exponential decay $e^{-x/\tilde{b}}$ with the parameter $\tilde{b}$ was chosen also as the smoothing function in the time domain. This function gives to the narrowed spectral lines Lorentz shapes as well.

3.3. Linear prediction algorithm

The LP algorithm can be the part of a line narrowing method. By means of this algorithm it is possible to extend the time series beyond its existence thus making the signal much broader and therefore the spectral lines much narrower. The signal can be predicted forward and backward. The first forward predicted sample $\hat{x}_j$ which follows the last known one x_{j-1} can be estimated by the linear combination of the form

$$\hat{x}_j = \Delta x \sum_{l=1}^{M} h_l x_{j-l}, \tag{31}$$

whereas the backward predicted sample that precedes the first known one can be found as

$$\hat{x}_j = \Delta x \sum_{l=1}^{M} h_l^* x_{j+l}. \tag{32}$$

By applying (31) and (32) in a cycle one can extrapolate the original "short" signal as far as desired [28]. The coefficients h_l are known as theoretical impulse response coefficients. One of the known approaches to the problem of finding the impulse response coefficients is exploiting of the Burg's formula including the Levinson-Durbin recursion.

If the signal is predictable without errors by means of the known $M = m - 1$ impulse response coefficients $h_1^{(m-1)}, \dots, h_{m-1}^{(m-1)}$ then it is possible to obtain a new set of m impulse response coefficients using the Levinson-Durbin recursion

$$\begin{cases} h_1^{(m)} = h_1^{(m-1)} - \Delta x r h_{m-1}^{(m-1)*}, \\ h_2^{(m)} = h_2^{(m-1)} - \Delta x r h_{m-2}^{(m-1)*} \\ \vdots \\ h_{m-1}^{(m)} = h_{m-1}^{(m-1)} - \Delta x r h_1^{(m-1)*} \\ h_m^{(m)} = r, \end{cases} \tag{33}$$

where r is called the reflection coefficient and can be calculated using the Burg's formula

$$r = h_m^{(m)} = \frac{2}{\Delta x} \frac{\sum_{n=m}^{N-1} e_n^{(m-1)} b_{n-1}^{(m-1)*}}{\sum_{n=m}^{N-1} \left[\left| e_n^{(m-1)} \right|^2 + \left| b_{n-1}^{(m-1)} \right|^2 \right]}, \tag{34}$$

N is the number of samples involved in the calculations and $N > M$ [28]. The values $e_n^{(m-1)}$ and $b_n^{(m-1)}$ are called prediction residuals and can be computed by the recursion

$$\begin{cases} e_n^{(m-1)} = e_n^{(m-2)} - \Delta x h_{m-1}^{(m-1)} b_{n-1}^{(m-2)}, \\ b_n^{(m-1)} = b_{n-1}^{(m-2)} - \Delta x h_{m-1}^{(m-1)*} e_n^{(m-2)} \end{cases} \tag{35}$$

with the initial values $e_n^{(0)} = b_n^{(0)} = x_n, n = 0, \dots, N-1$.

3.4. Quantity factor

In order to define whether the given impulse response coefficients are capable of extrapolating the known time series with the minimum errors, one should use some criterion. The quantity factor (q-factor) is approved to be a good one [28]. The idea is to extrapolate the known part of the signal and then to compare it with the corresponding

original samples. For conjugate symmetric signals it is reasonable to use q-factor of the following form:

$$q = 1 - \sqrt{\frac{\sum_{j=J}^{N_0-1} \left| \hat{x}_{-j} - x_{-j} \right|^2}{\sum_{j=J}^{N_0-1} \left| x_j \right|^2}}, \tag{36}$$

where N_0 is the number of the last tested sample. This form assumes the backward prediction of the signal. To be more specific, the positive half of the signal is extrapolated backward and the extrapolated negative half is further compared with the corresponding original half. The q-factor depends on N and M, i.e. $q = q(N, M)$. Moreover, the better the extrapolation, the higher the q-factor is ($\max(q) = 1$). The negative sign of the q-factor may indicate that extrapolated signal is mostly of opposite phase regarding to the original signal's phase. Therefore, one should analyze the q-factor for the local extremums, where derivatives of it are equaled to zero.

3.5. The LOMEP algorithm

The LOMEP (Line shape Optimized Maximum Entropy linear Prediction) algorithm is described by the following steps [28]:

1. Convert the spectra under consideration to the time series using inverse Fourier transform (25).
2. Deapodize the signal obtained at the step 1.
3. a) Choose the number of samples to be extrapolated and compute the impulse response coefficients using Levinson-Durbin recursion (33) and Burg's formula (34).
   b) Compute the q-factor (36) as the function of N and M.
   This step is implemented repeatedly until the best q-factor is obtained.
4. Accept the impulse response coefficients that correspond to the best q-factor and extrapolate the signal as far as desired.
5. Apodize the result in order to smooth the signal, to obtain desired line shape and to minimize sinc-function effects of the sharply truncated signal.
6. Convert the result back to the frequency domain by using direct Fourier transform (26).

These steps are graphically demonstrated in the Fig.13.

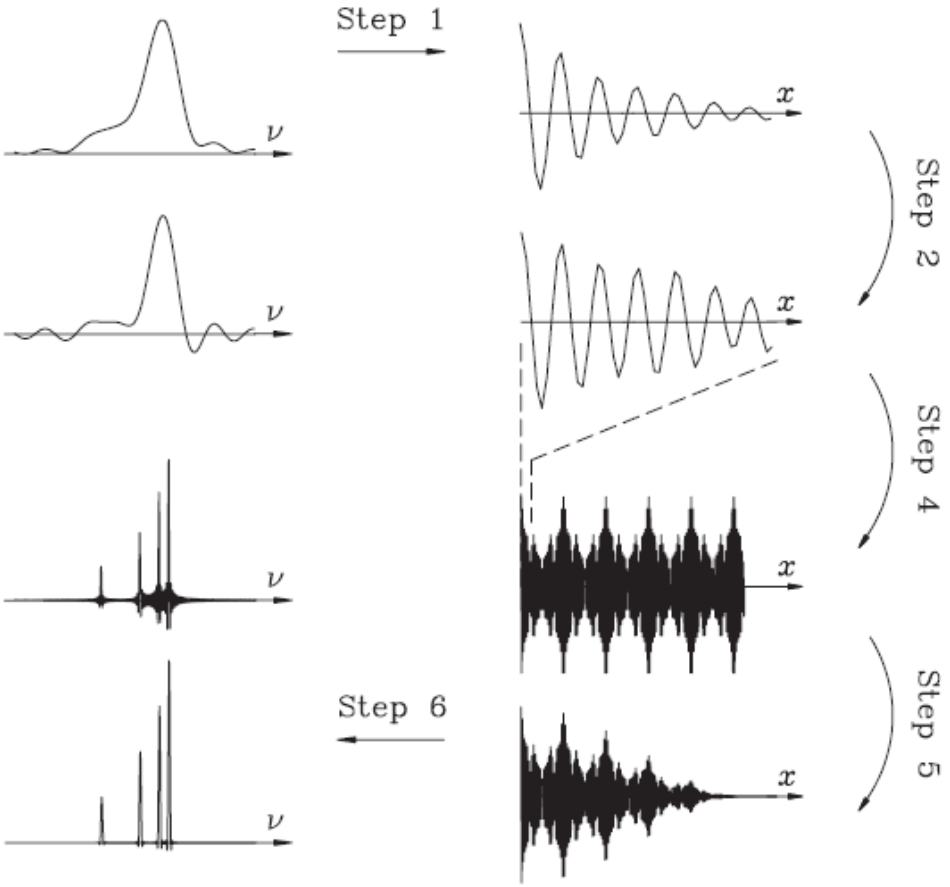

Fig.13. Graphical demonstration of the steps of the LOMEP method [28].

4. SPECTRA ANALYSIS AND RESULTS

5.1. Water solution of 300 mM sucrose

Let us start our analysis from the MEM phase spectrum of water solution of 300 mM sucrose shown in the Fig.8. First, we consider the spectrum without nonlinear background which is obtained by subtracting the corresponding spectrum of water without sucrose (Fig.7). The result is shown in the Fig.14. At the right a smoothed version with suppressed noise is presented. Smoothing was implemented in the process of cubic spline interpolation using Hiroshi Akima's method [30]. The smoothed spectrum will be taken as an example of line narrowing analysis which will be considered in detail. The analysis of the other spectra will be accompanied only with the key results. The words on the spectra including non-constant background are reserved to the Conclusions section.

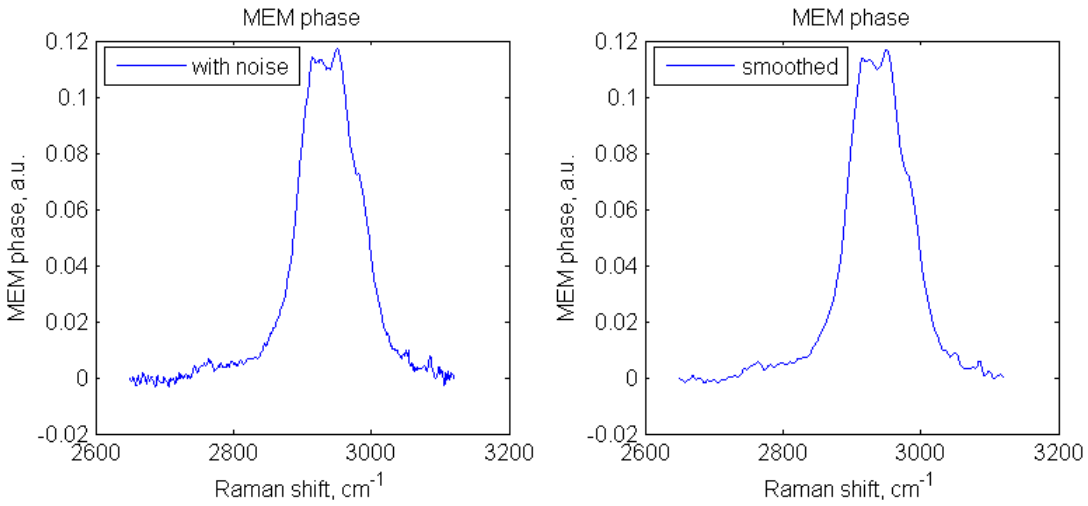

Fig.14. MEM spectra of water solution of 300 mM sucrose without nonlinear background: original (left) and smoothed (right).

The corresponding signal of the smoothed spectrum was obtained by the inverse DCT (25) and shown in the Fig.15.

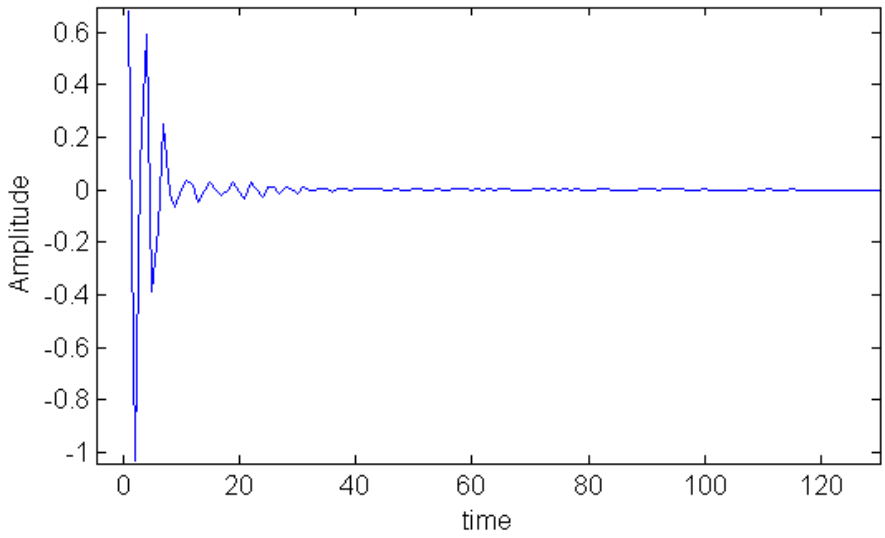

Fig.15. Signal corresponding to the smoothed spectrum of water solution of 300 mM sucrose.

As prescribed by the LOMEP algorithm, the next step is to deapodize the signal. As was mentioned above, for this reason exponential decay function $e^{-x/b}$ was used. The problem now is what value b should be in order to obtain the best results. At this point q-factor criterion was used as well. Thus, q-factor was considered as the function of three parameters, namely b, N and $m \equiv N - M$, where N is the number of the last sample used in the LP procedure and M is the number of impulse response coefficients. In the Fig.16 the q-factor as the function $\tilde{q}(b) = \max_{m,N} q(b, m, N)$ of one variable is presented. X-axis stands for the parameter b while y-axis stands for the value of the q-factor. As seen from the figure, there are several local extremums which are to be checked for the best result. The last is obtained for the set $(X, Y) = (b, q) = (15, 0.8086)$. The value $b = 10$ is too small resulting in significant increase of noise close to the tail. The value $b = 23$ in turn is too large resulting in the small effect of apodization procedure.

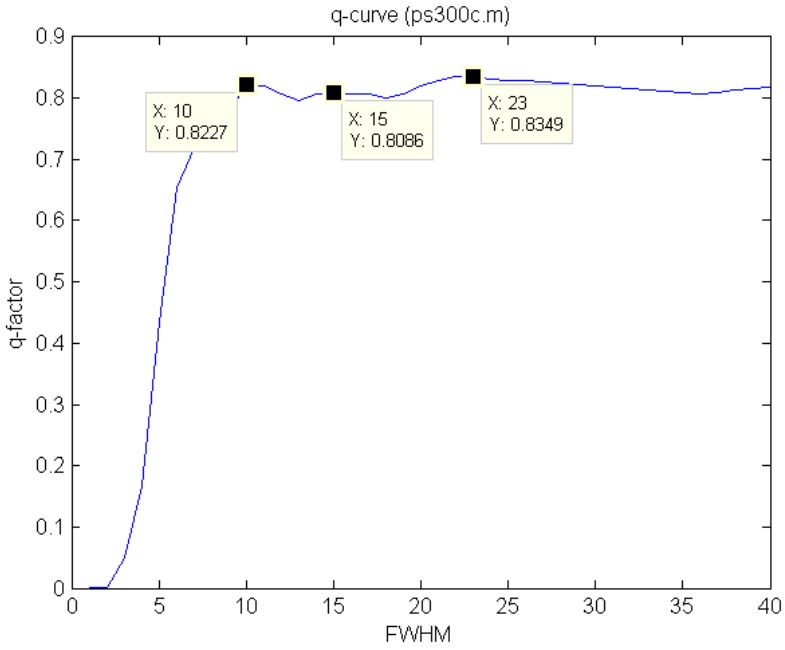

Fig.16. The q-factor as the function of variable b for 300 mM sucrose.

Once parameter b is defined, the next step is to deapodize the signal and then define other two parameters m and N, i.e. decide how many impulse response coefficients are the best to use and where to cut the signal to extrapolate it further. The result of deapodization is shown in the Fig.17.

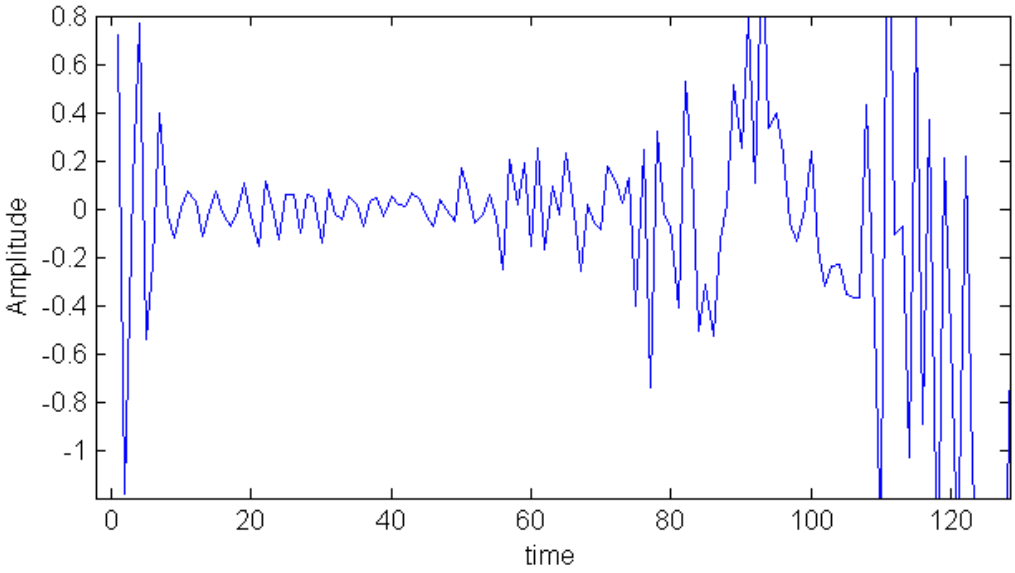

Fig.17. Signal after deapodization.

Now we can immediately proceed to finding M and N by using Levinson-Durbin recursion (33) and Burg's formula (34). For this purpose a contour map was plotted showing different values of the q-factor (Fig.18). From this map one can easily estimate the values of m and N to be 4 and 35 correspondingly. These are the values of the highest point on the map.

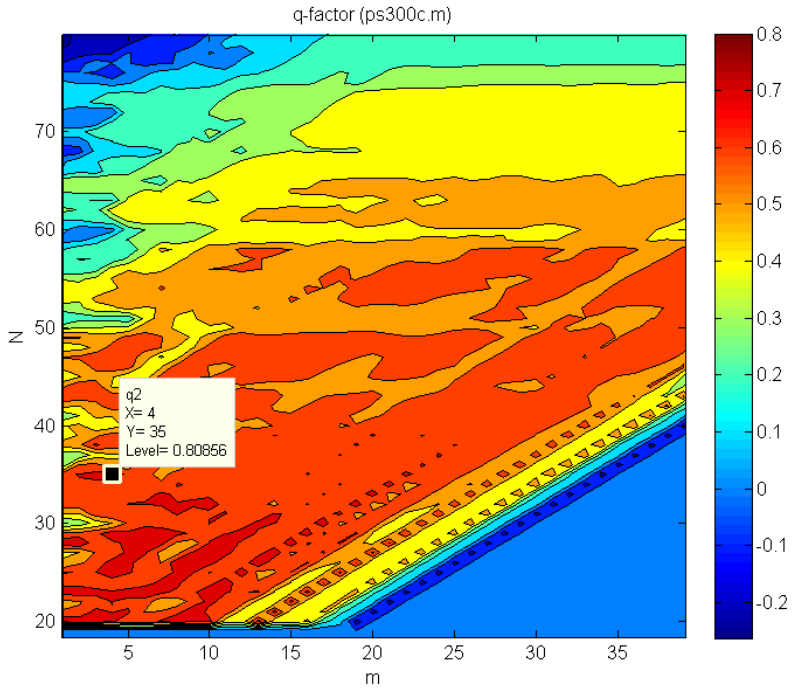

Fig.18. Contour map of the surface $q(15, m, N)$.

Using the obtained set of the parameters that provides the best q-factor $q(15,4,35) = 0.8086$, the signal was linearly predicted forward by 1500 new samples. The result is shown in the Fig.19.

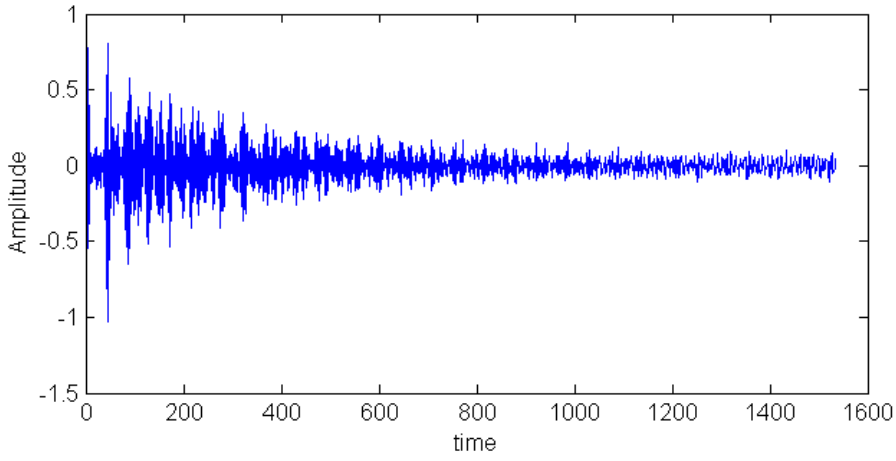

Fig.19. Forward predicted signal.

In order to decrease sinc-function effect of abrupt truncation at the end of the signal, the last was iteratively apodized (smoothed) by exponential decay function with the parameter $\tilde{b} = 500$.

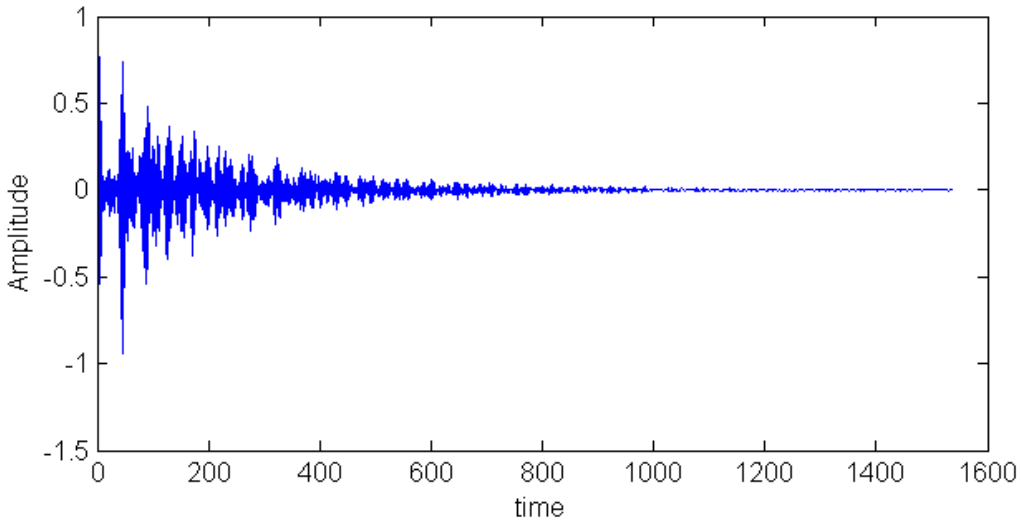

Fig.20. Apodized forward predicted signal.

Taking the DCT of the result by using (26) we finally obtain narrowed spectral lines (Fig.21).

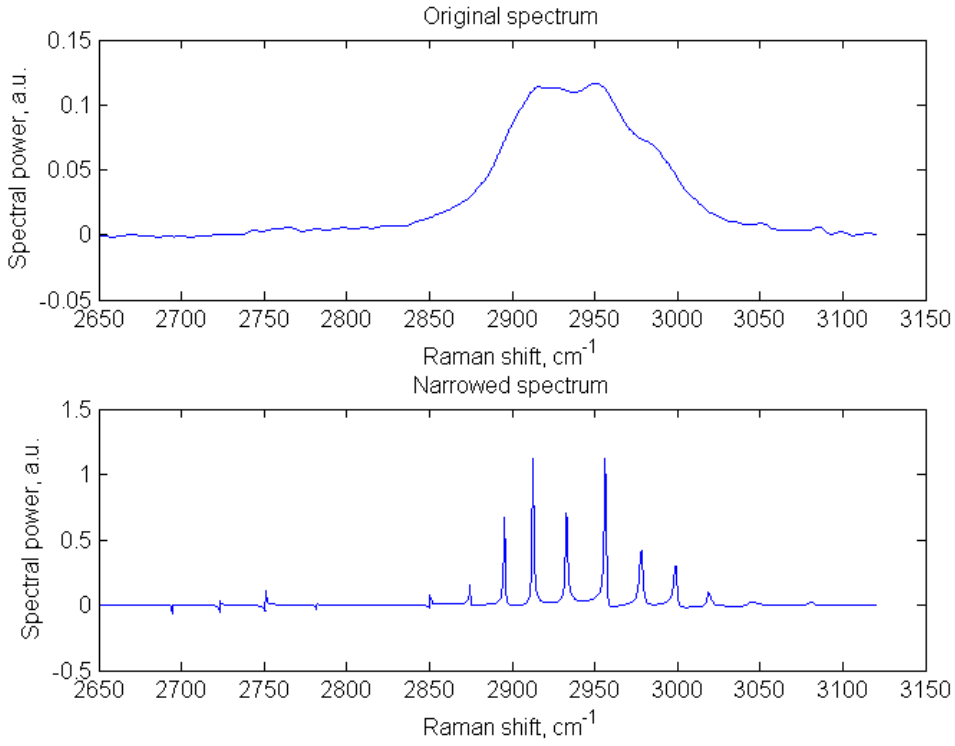

Fig.21. Narrowed spectrum of 300 mM sucrose compared to the original one ($\tilde{b} = 500$).

For the noised spectrum parameters appear to be the same. In this case the q-factor is equaled to 0.8119 and the narrowed spectrum visually looks as shown in the Fig.21.

4.2. Water solution of 500 mM sucrose

Here and further only the key results are presented.

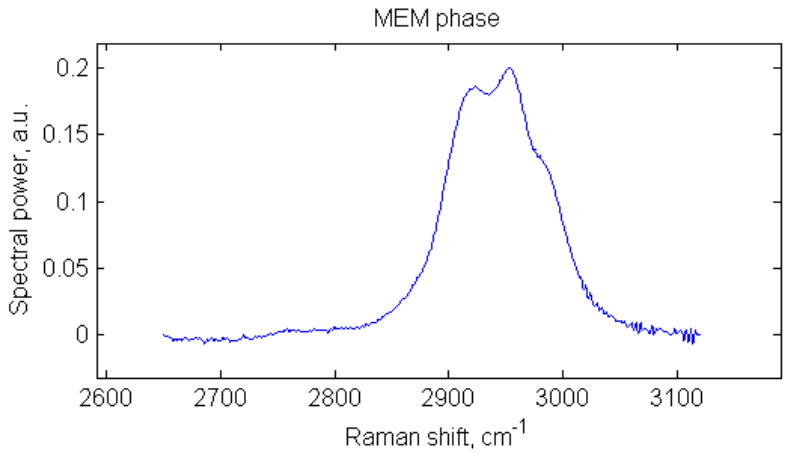

Fig.22. MEM spectrum of water solution of 500 mM sucrose without nonresonant water background.

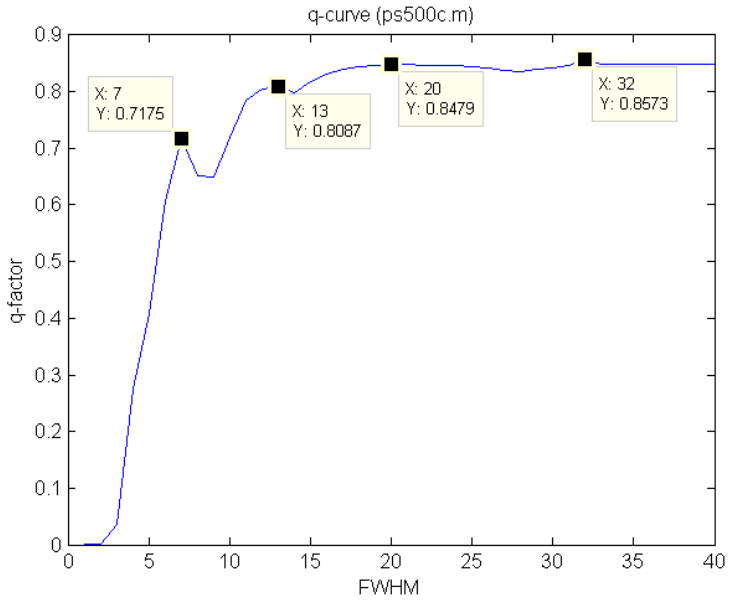

Fig.23. The q-factor as the function $\tilde{q}(b) = \max_{m,N} q(b, m, N)$.

In the Fig.24–26 q-factor contour maps are shown for the different values of parameter b. It is worth noting how contours are changed with the increase of the parameter. For $b = 20$ there is a large island with the big values of q-factor.

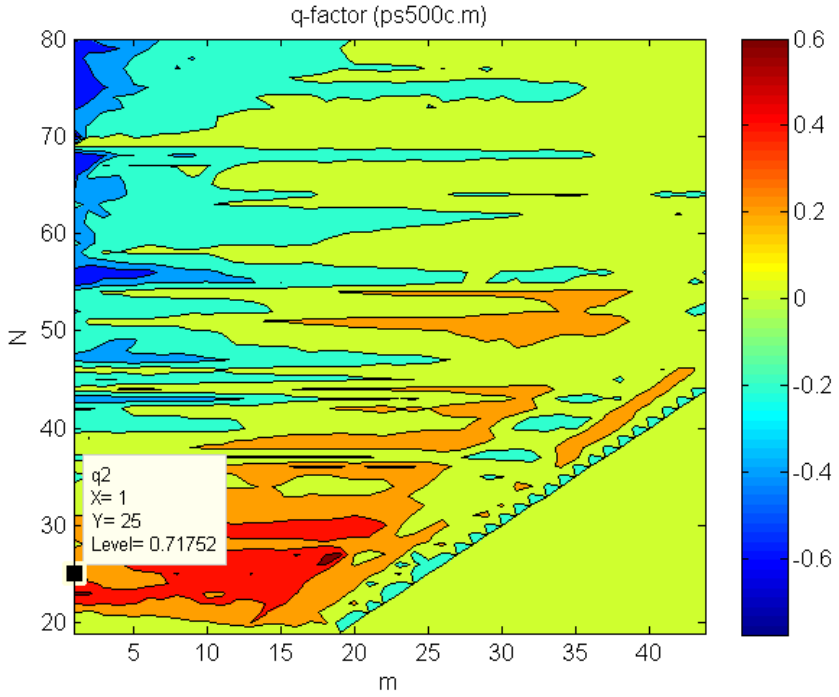

Fig.24. Contour map of the surface $q(7, m, N)$.

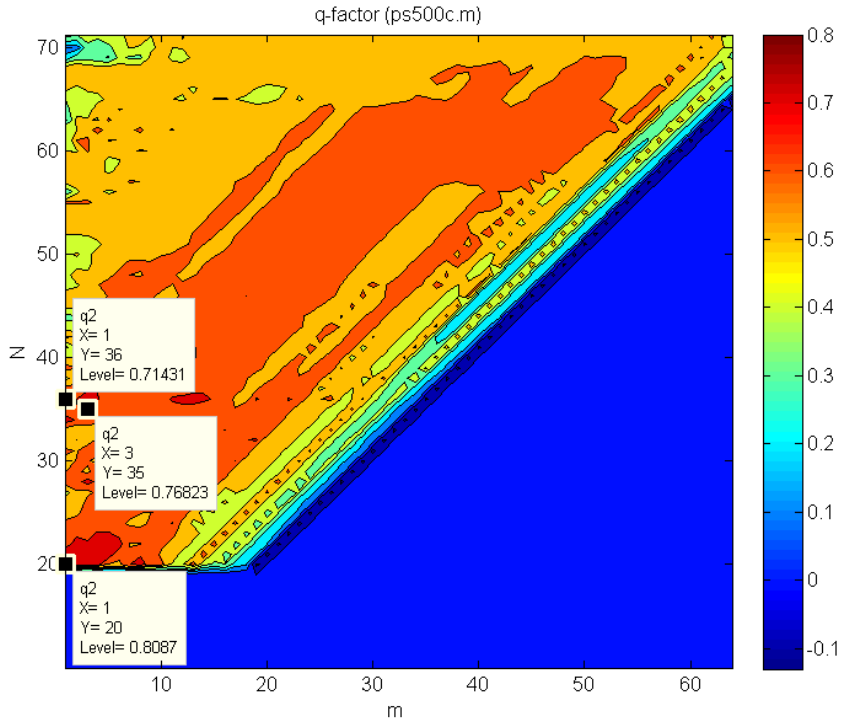

Fig.25. Contour map of the surface $q(13, m, N)$.

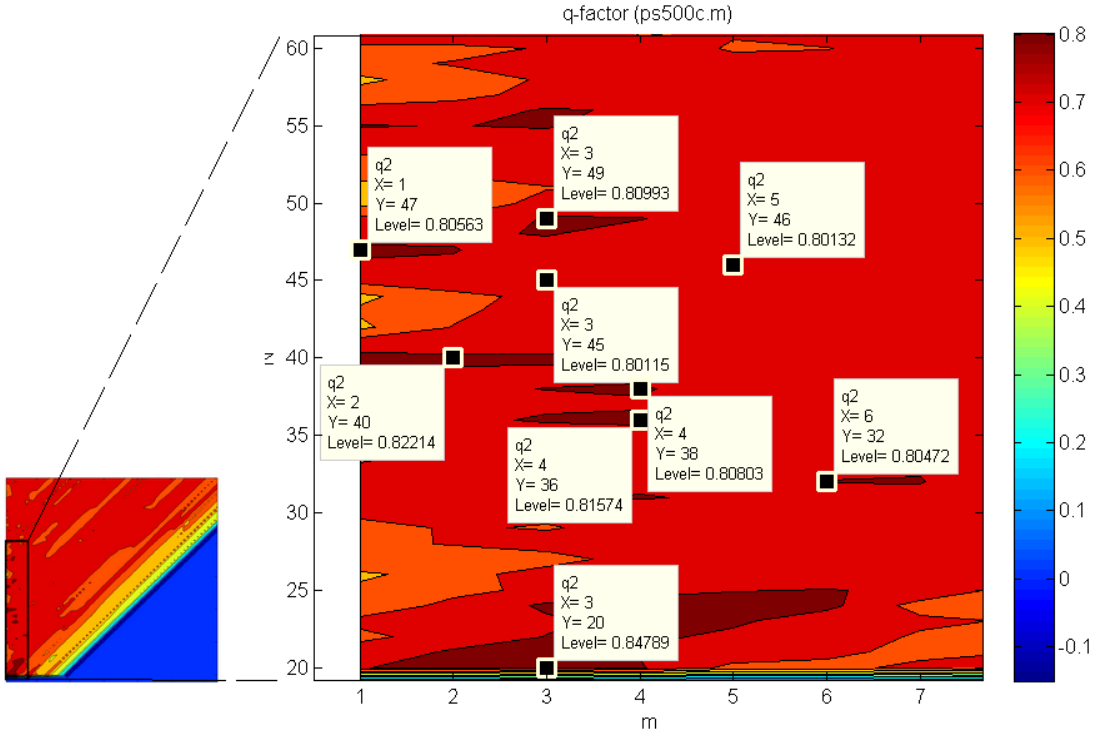

Fig.26. Contour map of the surface $q(20, m, N)$.

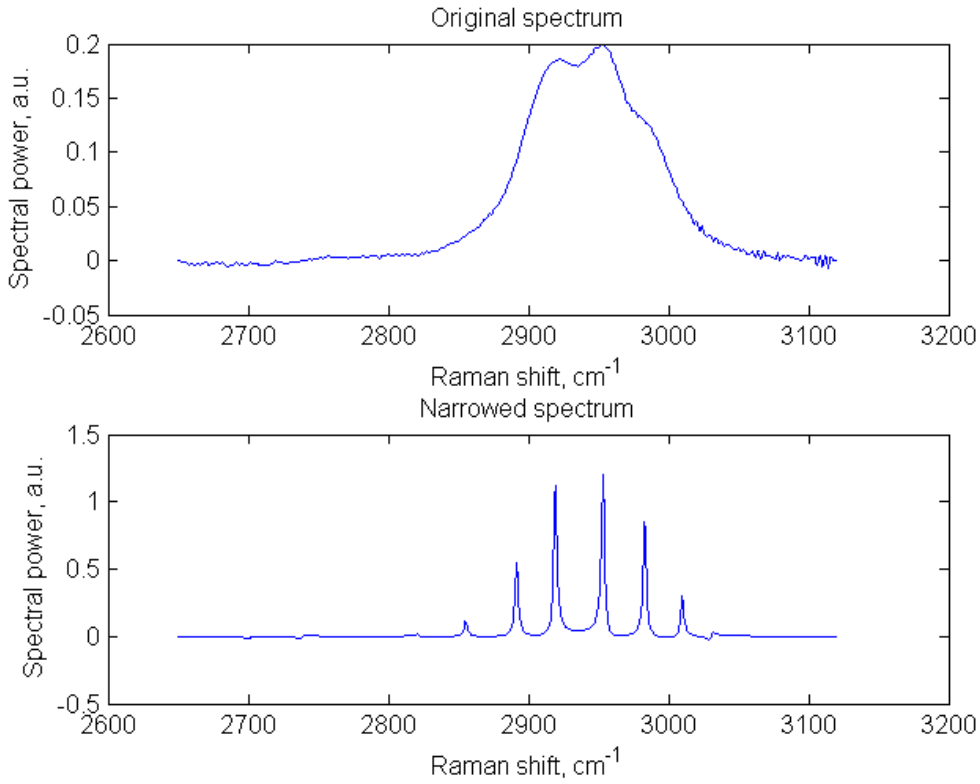

Fig.27. Narrowed spectrum of 500 mM sucrose compared to the original one ($\tilde{b} = 300$).

4.3. Simulated $|\chi^{(3)}|^2$ spectrum

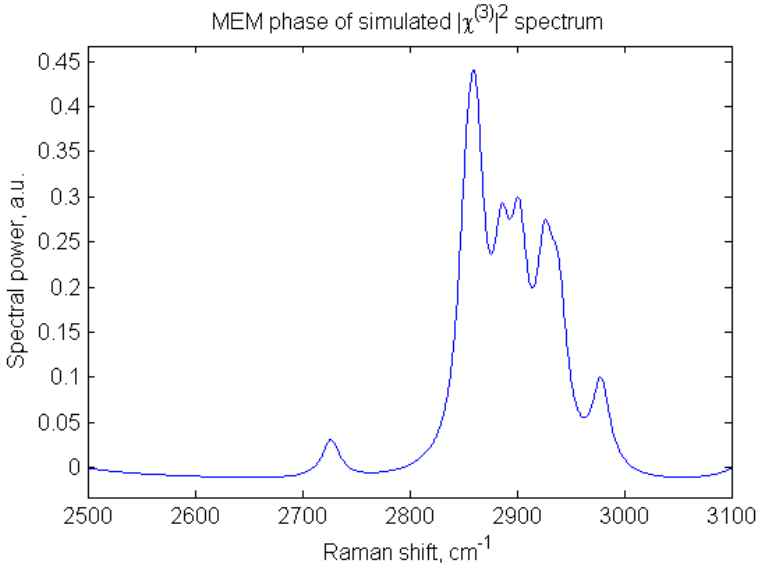

Fig.28. MEM spectrum of simulated $|\chi^{(3)}|^2$.

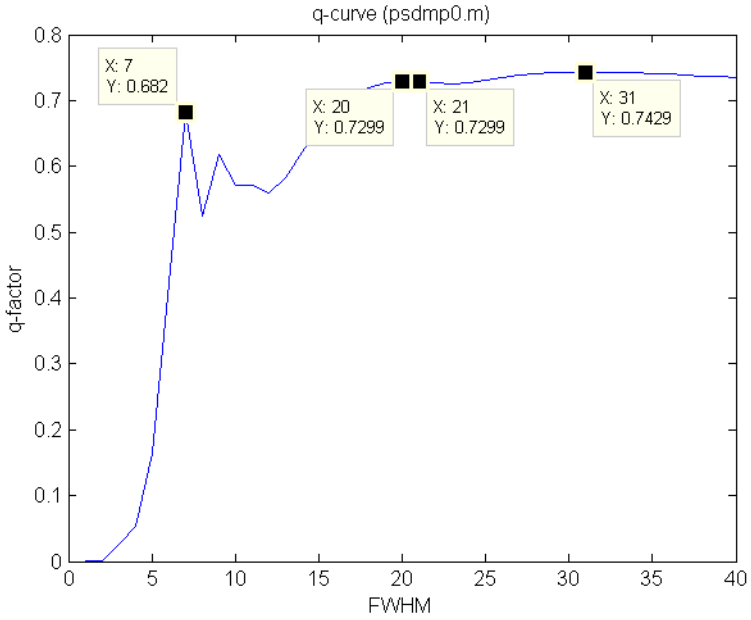

Fig.29. The q-factor as the function $\tilde{q}(b) = \max_{m,N} q(b, m, N)$.

The Fig.30–33 show the results obtained with the parameters $b = 20, N = 100$ and $m = 52$ for different values $\tilde{b}$.

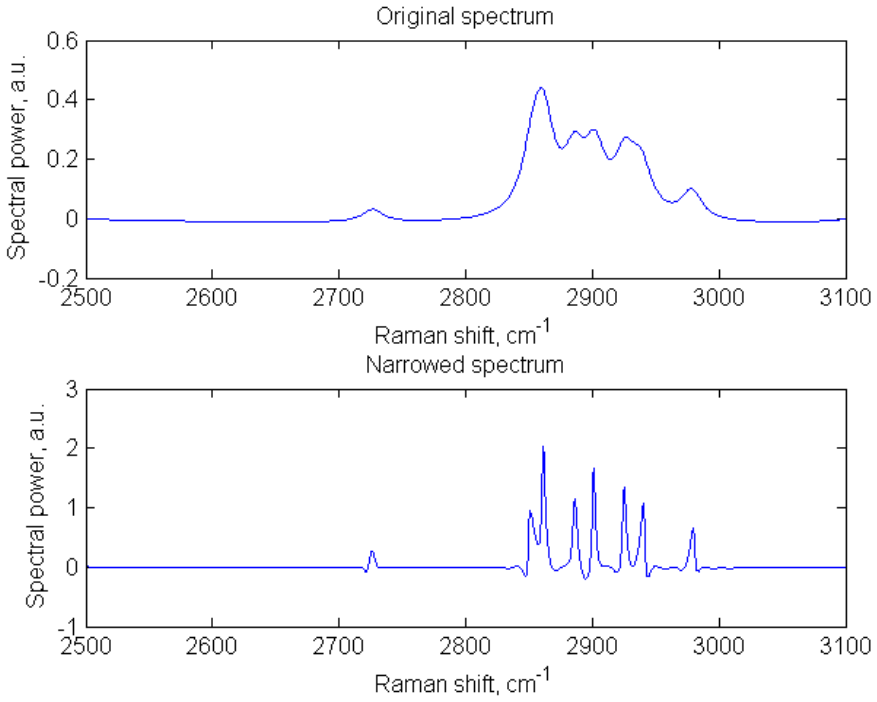

Fig.30. Narrowed spectrum ($\tilde{b} = 200$).

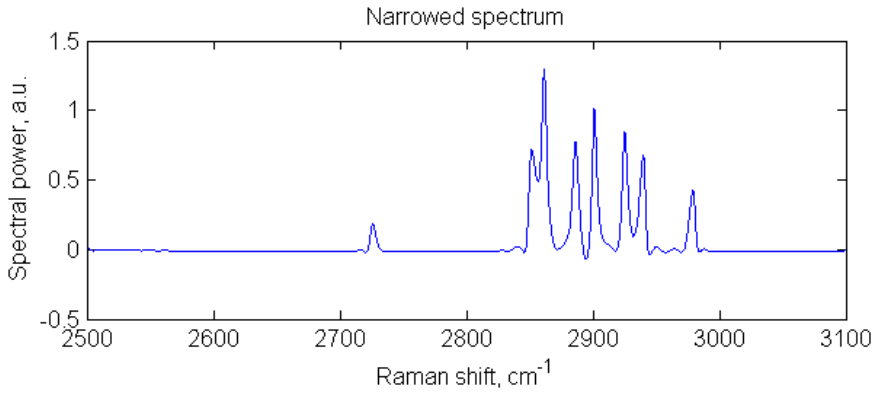

Fig.31. Narrowed spectrum ($\tilde{b} = 100$).

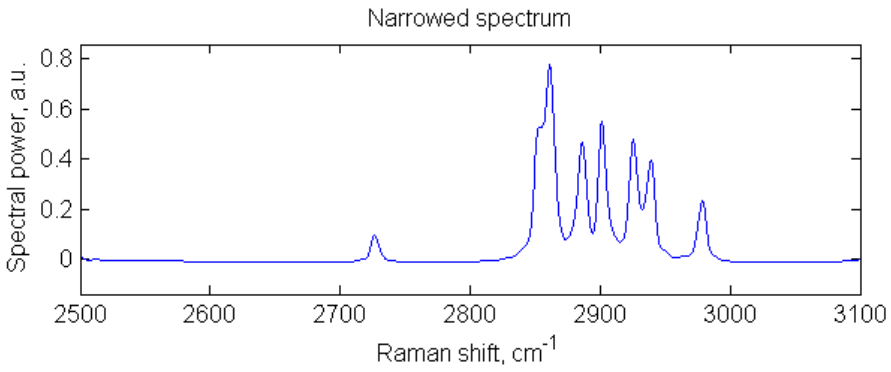

Fig.32. Narrowed spectrum ($\tilde{b} = 50$).

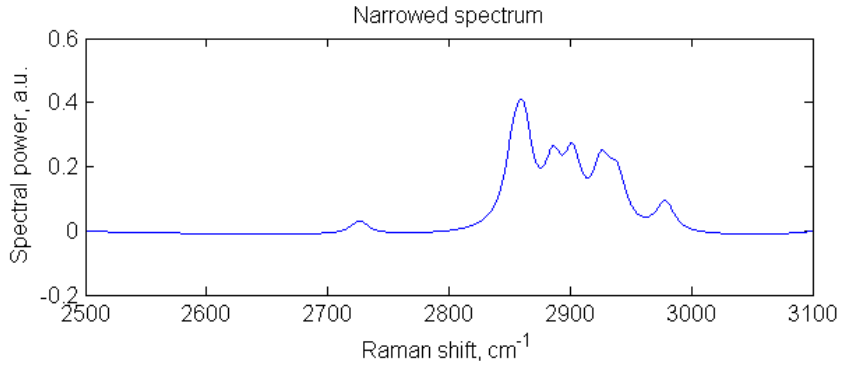

Fig.33. Narrowed spectrum ($\tilde{b} = 20$).

4.4. ADP/AMP/ATP

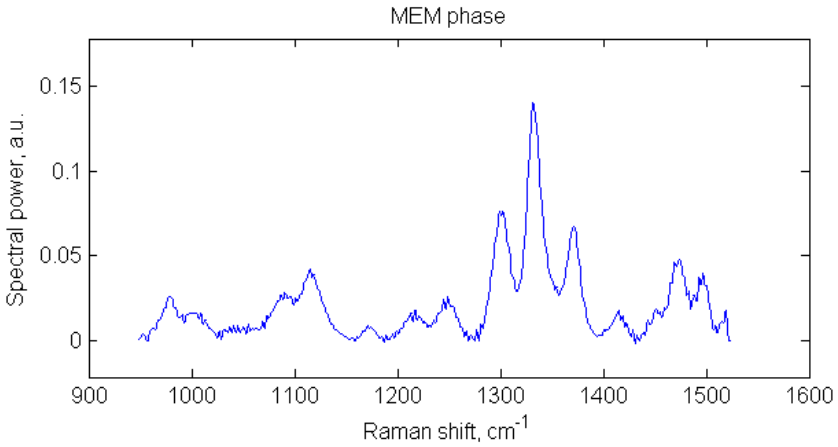

Fig.34. MEM spectrum of ADP/AMP/ATP.

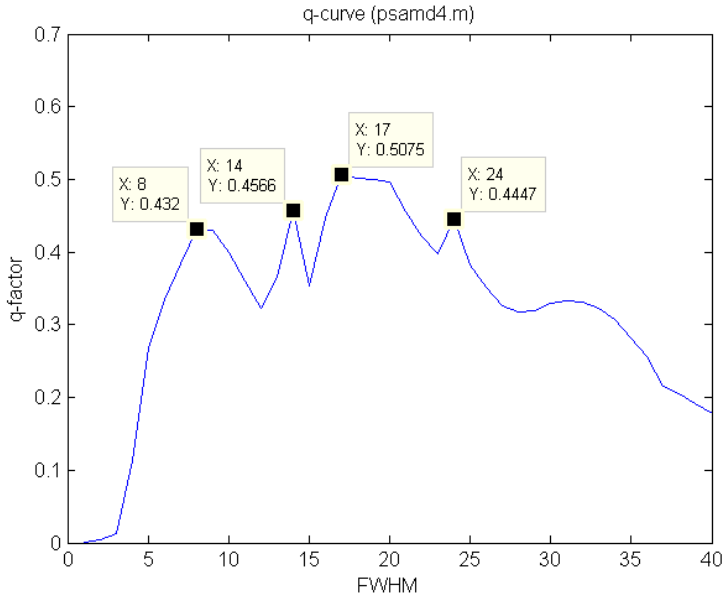

Fig.35. The q-factor as the function $\tilde{q}(b) = \max_{m,N} q(b, m, N)$.

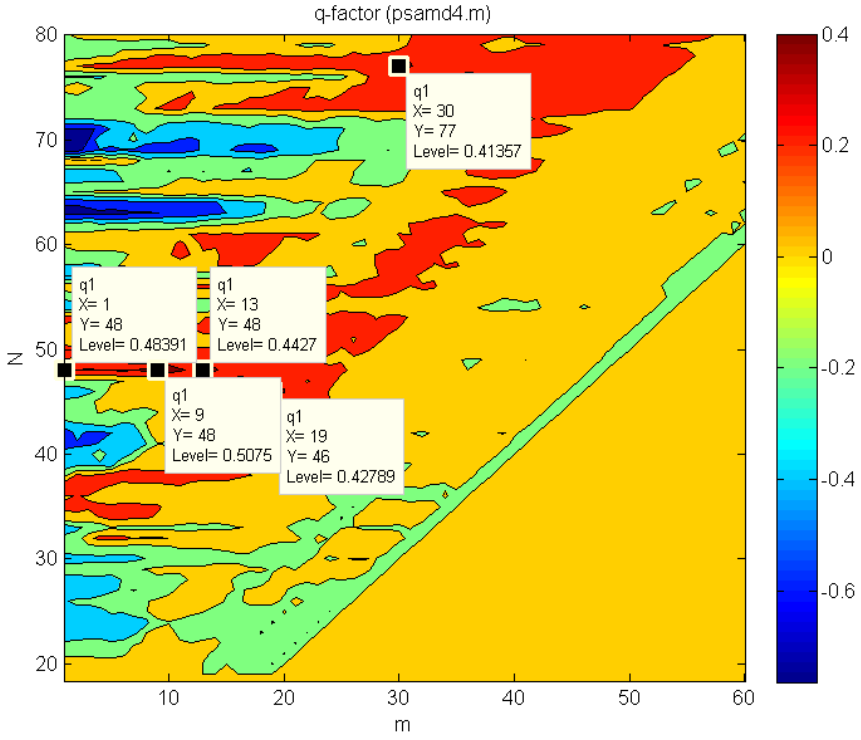

Fig.36. Contour map of the surface $q(17, m, N)$.

As seen from the Fig.36, the best set of parameters is $(b, m, N) = (17,9,48)$ which gives the result shown in the Fig.37.

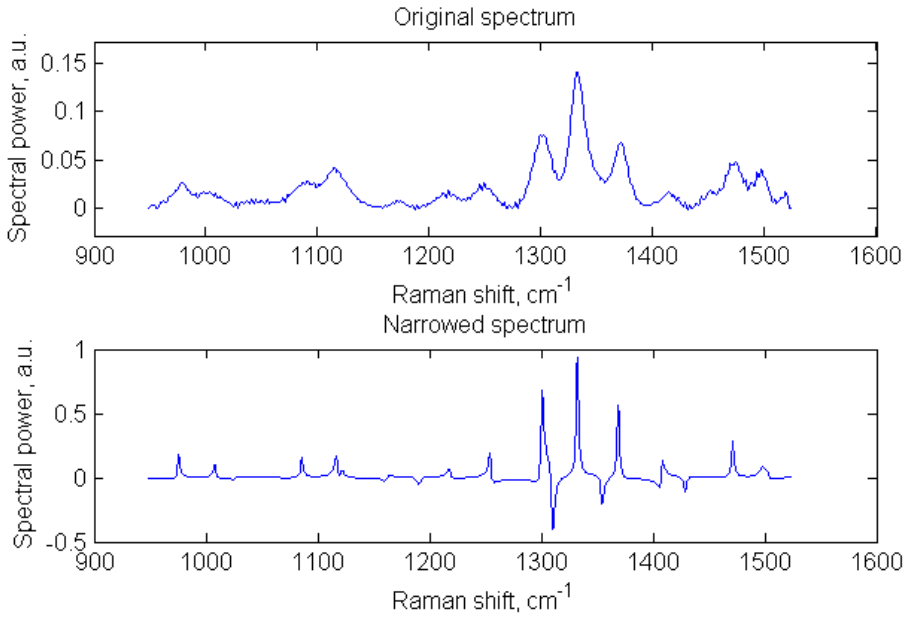

Fig.37. Narrowed spectrum ($\tilde{b} = 150$).

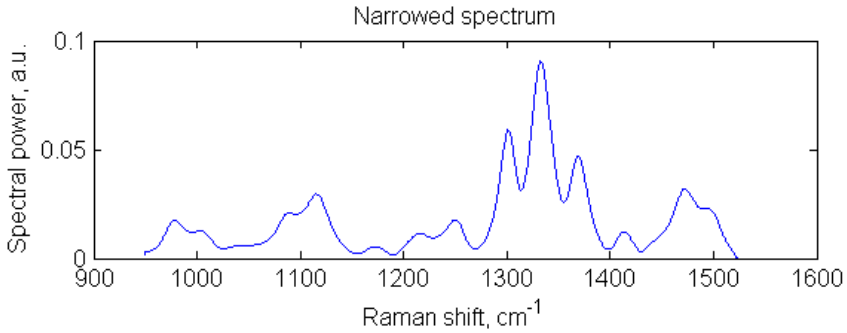

Fig.38. Narrowed spectrum ($\tilde{b} = 14$).

4.5. Simulated Lorentz peaks

In order to understand better how the line narrowing method works, the last was applied to the simulated sets of Lorentz peaks without noise and with random noise (signal-to-noise ratio SNR = 100). The line widths of all spectral lines are FWHM = 20 cm^{-1}. It was also simulated the same spectra with FWHM equaled to 10 cm^{-1}, 5 cm^{-1} and 1 cm^{-1} in order to be compared with the results of the LOMEP procedure.

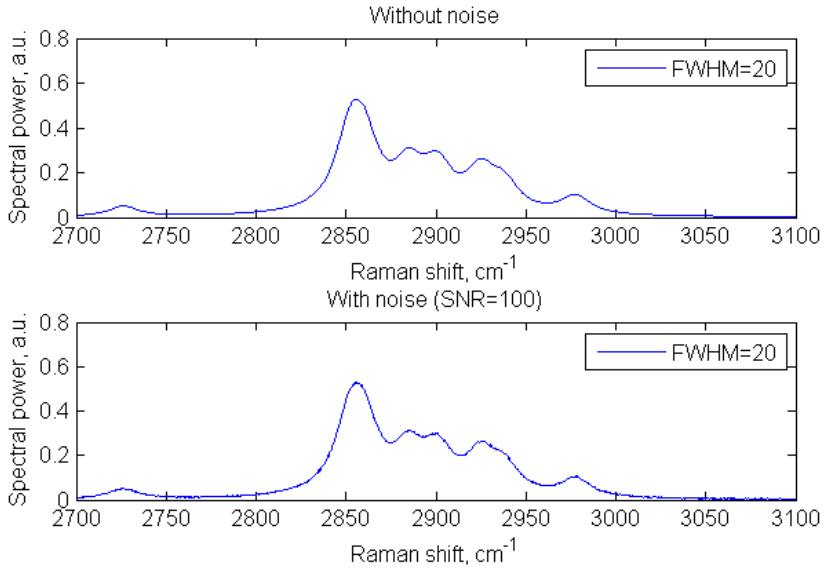

Fig.39. Simulated Lorentz peaks.

In the Fig.40 the q-curve is presented for the noised case. The peak ($q = 0.8012$) of the curve is at the point 11. The same position holds for the spectrum without noise for which $q = 0.7972$. In the Fig.41–44 the narrowed spectra are presented compared to the original ones with noise and different FWHM. The same spectra but without noise are presented in the Fig.45–48. In the case of the noised spectra the set of the parameters $(b, m, N) =$

$(11,2,37)$ was chosen, whereas for the spectra without noise the parameters were $(b, m, N) = (11,24,55)$. These parameters in both cases do not correspond to the maximum values of q-factor though (Fig.49–50).

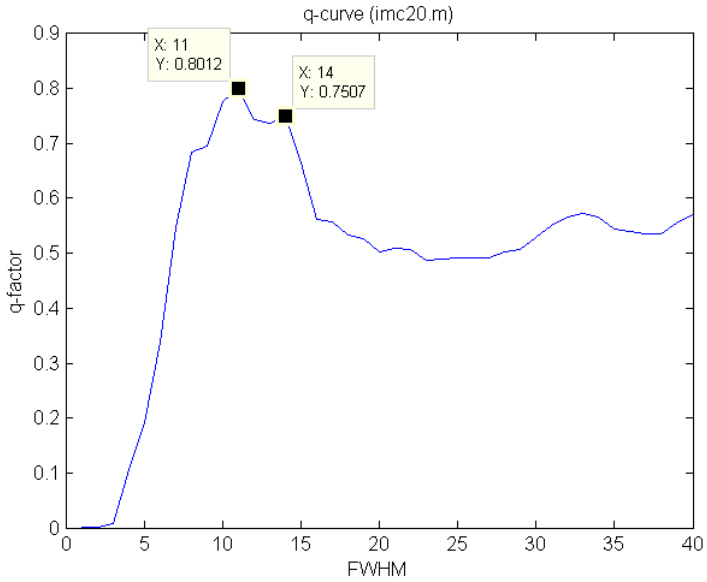

Fig.40. The q-factor for noised spectrum.

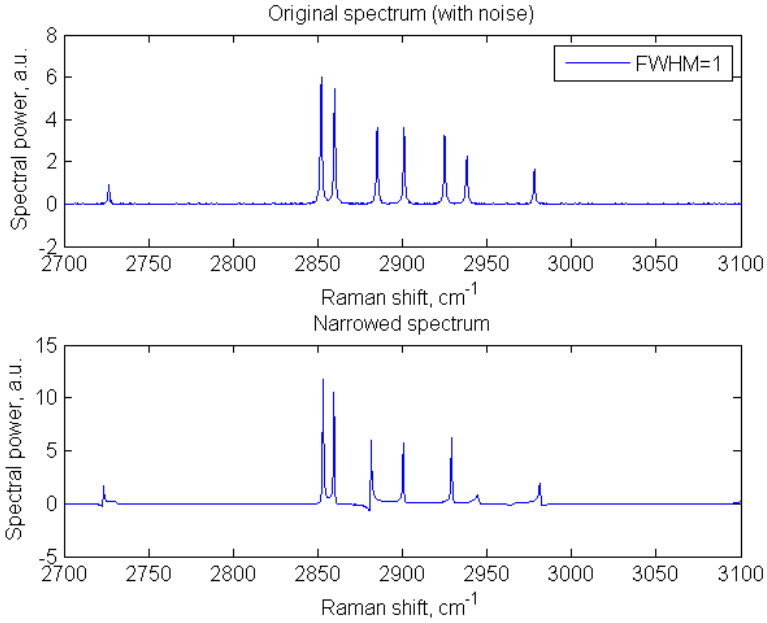

Fig.41. Narrowed spectrum ($\tilde{b} = 400$).

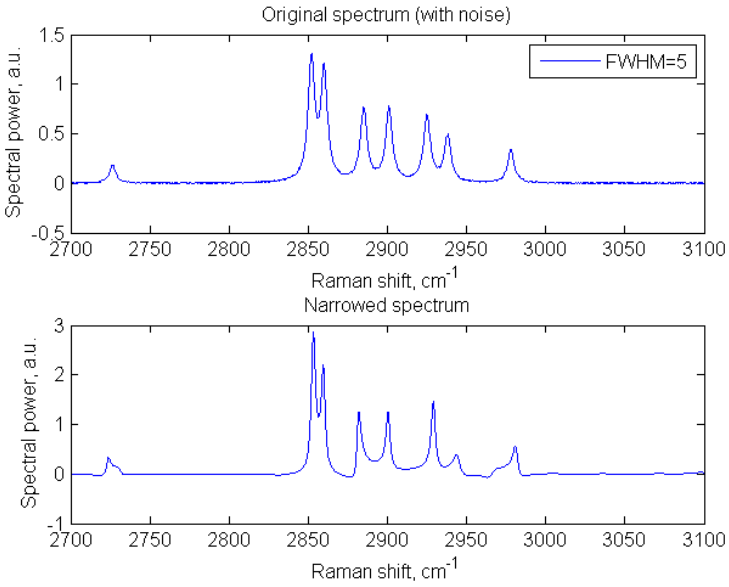

Fig.42. Narrowed spectrum ($\tilde{b} = 80$).

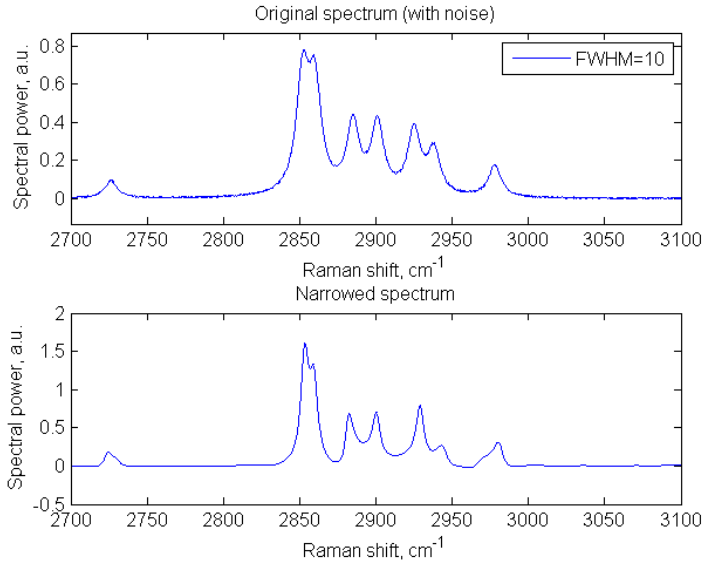

Fig.43. Narrowed spectrum ($\tilde{b} = 40$).

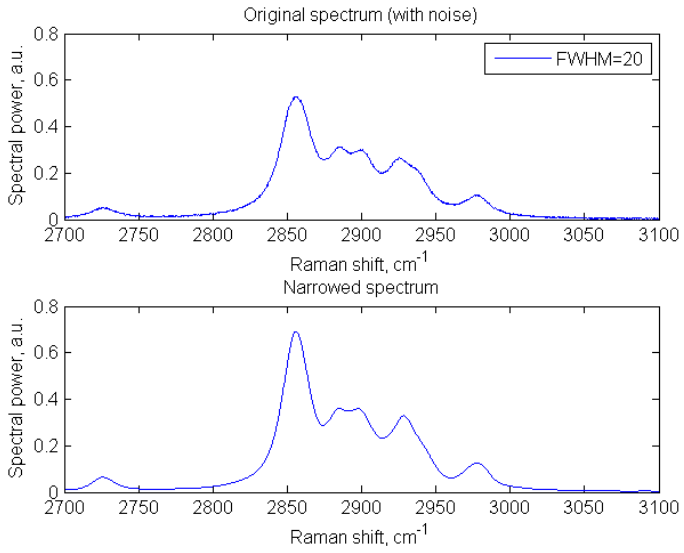

Fig.44. Narrowed spectrum ($\tilde{b} = 12.5$).

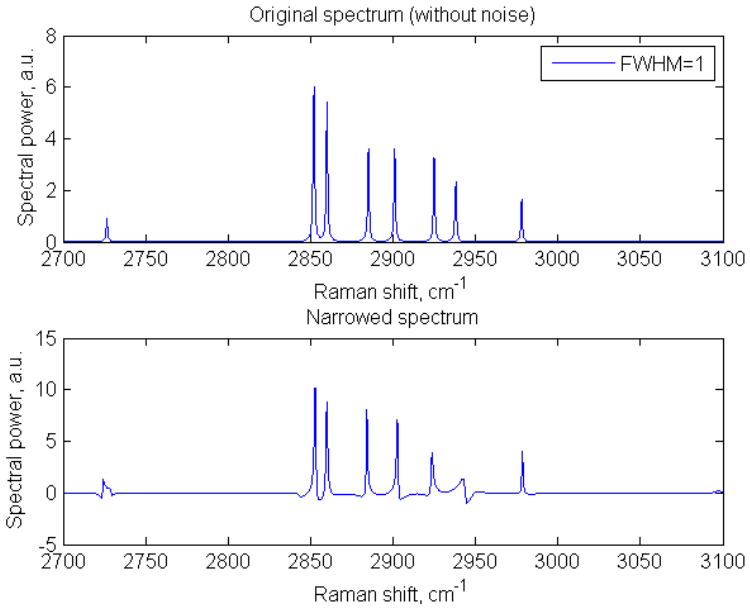

Fig.45. Narrowed spectrum ($\tilde{b} = 400$).

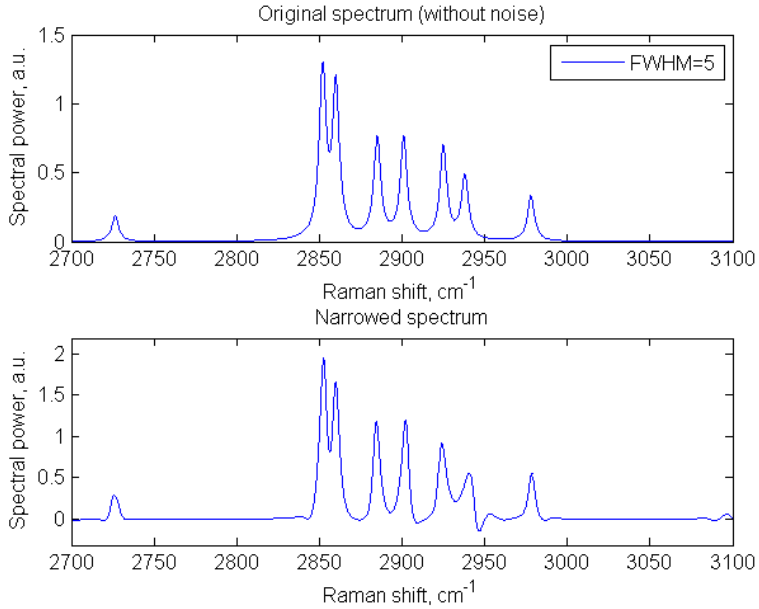

Fig.46. Narrowed spectrum ($\tilde{b} = 50$).

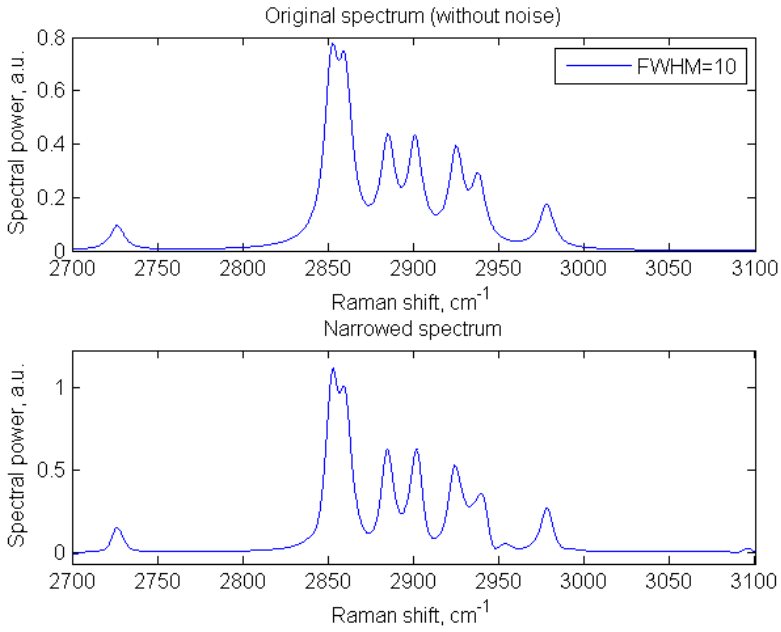

Fig.47. Narrowed spectrum ($\tilde{b} = 26$).

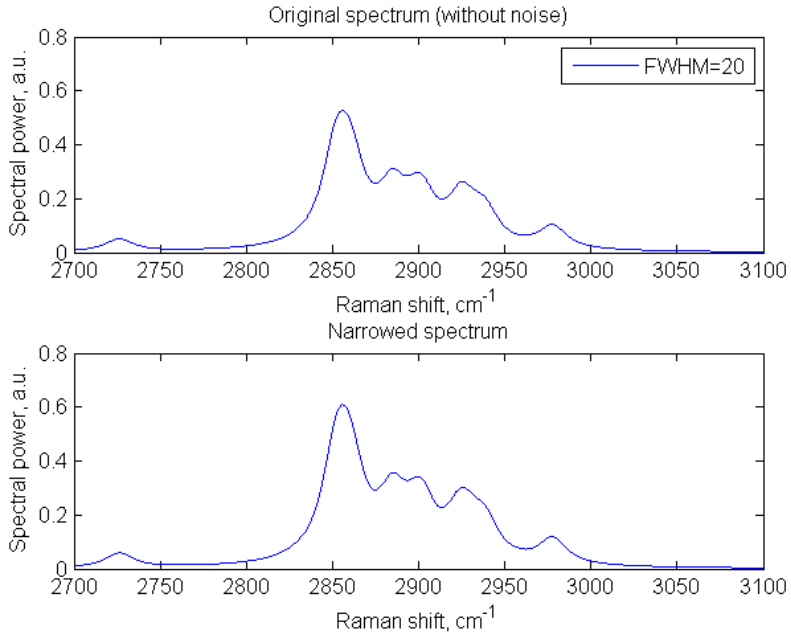

Fig.48. Narrowed spectrum ($\tilde{b} = 11$).

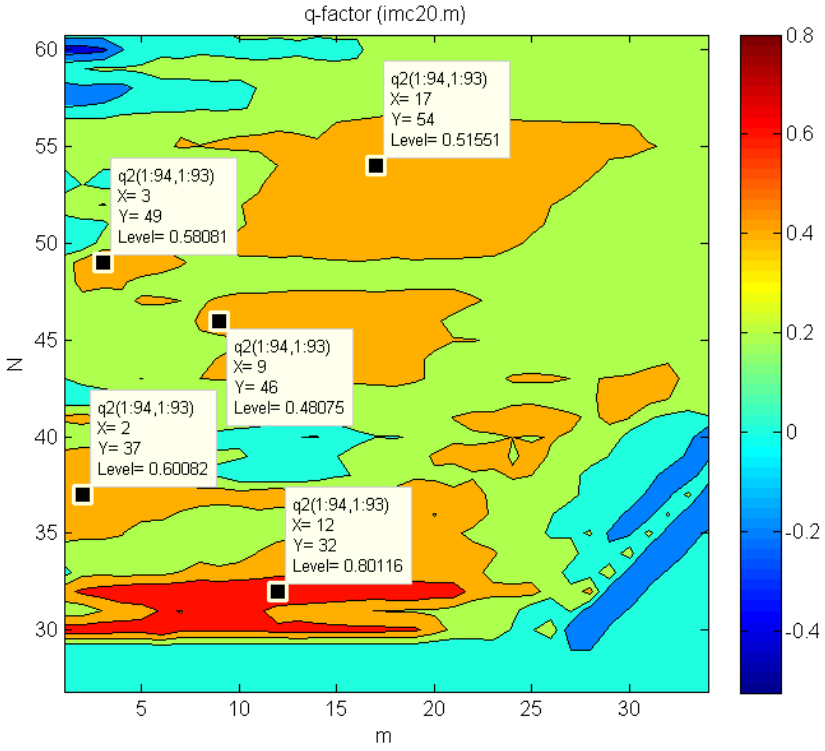

Fig.49. Contour map of the surface $q(11, m, N)$ for the case of the spectrum with noise.

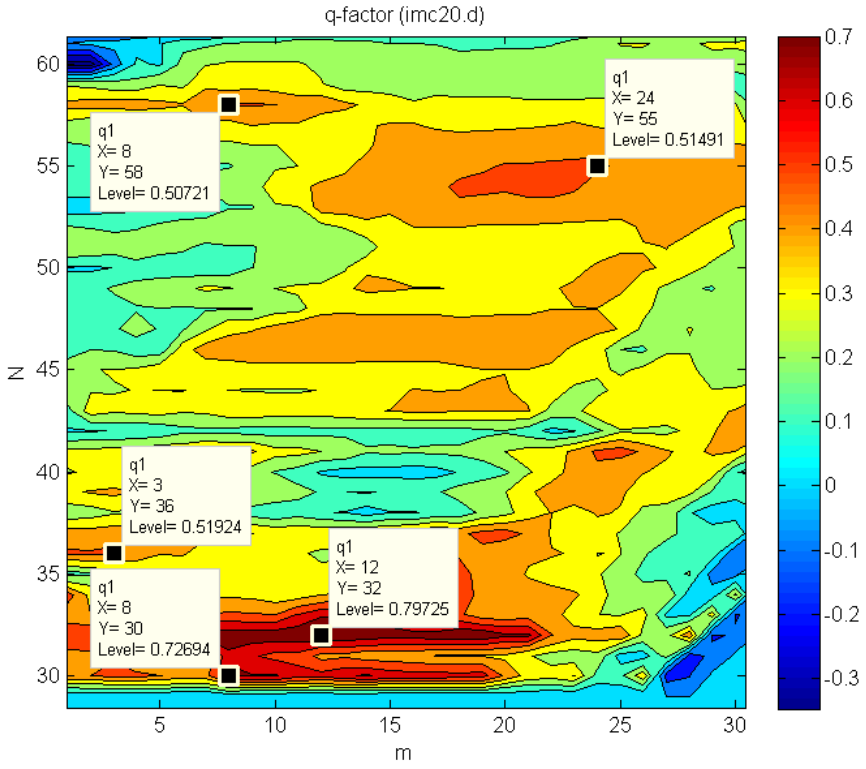

Fig.50. Contour map of the surface $q(11, m, N)$ for the case of the spectrum without noise.

CONCLUSIONS

As seen from the results, the LOMEP algorithm is quite effective theoretical tool for the line narrowing. However, the results of this paper are far from ideal, and algorithm implemented in the MatLab environment requires further elaboration. First of all, the symmetry of the spectral lines is distorted resulting in slight displacement of their positions. Therefore frequency tuning method could be quite appropriate [28]. Second, the maximum value of the q-factor is not necessarily the best as it may correspond to the parameters for which not the all spectral features are taken into account. For instance, relatively small values of N can cut away several frequencies, whereas small values of M (that correspond to large values of m) mean that there are not enough impulse response coefficients to extrapolate all frequencies. The results show that the number of the coefficients should belong to the approximate range 30–50. The greater values of this number could lead to appearance of additional spectral components that were not in original spectra. In general, the minimal number of the impulse response coefficients must be equaled to twice the number of spectral lines [28]. Finally, noised spectra are narrowed with difficulties meaning that the best results might be obtained for the q-factors that are not local extremums at all. Nevertheless, the peaks of the q-factor can be exploited as the starting anchor points for the parameters estimation.

It is worth saying a few words about how the LOMEP method works with the spectra containing non-constant background. Spectral nonresonant background is considered by the algorithm as the set of additional resonant frequencies that do not characterize a medium. As a result, in the output narrowed spectrum there may present a comb of these frequencies. In addition, q-factors in this case are mostly negative. The best positive q-factors are close to zero. Thus, the line narrowing procedure cannot directly narrow such kind of spectra. However, the LOMEP method could be used together with some background estimation procedure as the part of an integrated iterative process of background estimation. Knowing exact frequency values could also help to fit background with a spline within the interpolation process.

In conclusion, in CARS spectrometer setups the whole LOMEP procedure could be as a part of an automatic real-time system with the feedback based on the q-factor anchor points.

REFERENCES

[1] Bertolotti, M., 1999. The history of the laser. Translated from Italian by M. Bertolotti, 2005. Philadelphia: Institute of Physics Publishing.

[2] Demtröder, W., 2003. Laser spectroscopy: basic concepts and instrumentation. 3rd ed. Berlin: Springer.

[3] Smirnov V.V., 1984. Lazernaja spektroskopija kogerentnogo antistoksova rassejanija sveta molekuljarnyh gazov? [Coherent anti-Stokes light scattering laser spectroscopy of molecular gases?]. DPhil. USSR Academy of Sciences.

[4] Shen, Y.R., 1984. The principles of nonlinear optics. New York: Wiley-Interscience.

[5] Maker, P.D., Terhune R.W., 1965. Study of optical effects due to an induced polarization third order in the electric field strength. *Physical Review*, **137** (3A), A801-A818.

[6] Akhmanov, S.A., Koroteev, N.I., 1981. Metody nelinejnoj optiki v spektroskopii rassejanija sveta: aktivnaja spektroskopija rassejanija sveta? [Nonlinear optics methods of light scattering spectroscopy: active light scattering spectroscopy?]. Moscow: Nauka.

[7] Peiponen, K.-E., Vartiainen, E.M., Asakura, A., 1999. Dispersion, complex analysis and optical spectroscopy: classical theory. Berlin: Springer.

[8] Boyd, R.W., 2008. Nonlinear Optics. 3rd ed. Orlando: Academic Press.

[9] Levenson, M.D., Flytzanis, C., Bloembergen, N., 1972. Interference of resonant and nonresonant three-wave mixing in diamond. *Physical Review B*, **6** (10), 3962-3965.

[10] Vartiainen, E.M., Peiponen, K.-E., 2013. Optical and terahertz spectra analysis by the maximum entropy method. *Reports on Progress in Physics*, **76** (6), 066401.

[11] Bohlin, A., Bengtsson, P.-E., 2008. CARS spectroscopy. [online] Lund University. Available at: <http://www.forbrf.lth.se/english/research/measurement_methods/cars_spectroscopy/> [Accessed 12 March 2014].

[12] Alden, M., 2011. Combustion Laser Diagnostics. [pdf] Princeton University. Available at: <http://www.princeton.edu/cefrc/Files/2011 Lecture Notes/Alden/Lecture-10-CARS.pdf> [Accessed 25 March 2014].

[13] Shirley, J.A., Hall, R.J., Eckbreth, A.C., 1980. Folded BOXCARS for rotational Raman studies. *Optics Letters*, **5** (9), 380-382.

[14] Vartiainen, E.M., Rinia, H.A., Müller, M., 2006. Direct extraction of Raman line-shapes from congested CARS spectra. *Optics Express*, **14** (8), 3622-3630.

[15] Day, J.P.R., Domke, K.F., Rago, G., Kano, H., Hamaguchi, H., Vartiainen, E.M., Bonn, M., 2011. Quantitative coherent anti-Stokes Raman scattering (CARS) microscopy. *The Journal of Physical Chemistry B*, **115**, 7713-7725.

[16] Müller, M., Squier, J., De Lange, C.A., Brakenhoff, G.J., 2000. CARS microscopy with folded BoxCARS phasematching. *Journal of Microscopy*, **197** (Pt 2), 150-8.

[17] Shannon, C.E., 1948. A mathematical theory of communication. *Bell System Technical Journal*, **27** (3), 379-423.

[18] Berger, A.L., 1996. A brief maxent tutorial. [online] Carnegie Mellon University. Available at: <http://www.cs.cmu.edu/afs/cs/user/aberger/www/html/tutorial/tutorial.html> [Accessed 25 March 2014].

[19] Berger, A.L., Della Pietra, V.J., Della Pietra, S.A., 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*, **22** (1), 39-71.

[20] Unuma, T., Ino, Y., Kuwata-Gonokami, M., Vartiainen, E.M., Peiponen, K.-E., Hirakawa, K., 2010. Determination of the time origin by the maximum entropy method in time-domain terahertz emission spectroscopy. *Optics Express*, **18** (15), 15851-15858.

[21] Saarinen, J.J., Uozumi, J., Vartiainen, E.M., Peiponen, K.-E., 2012. Phase retrieval of reflectance for nanoparticle optical identification. *Optics Letters*, **37** (12), 2202-2204.

[22] Vartiainen, E.M., Ino, Y., Shimano, R., Kuwaata-Gonokami, M., Svirko, Y.P., Peiponen, K.-E., 2004. Numerical phase correction method for terahertz time-domain reflection spectroscopy. *Journal of Applied Physics*, **96** (8), 4171-4175.

[23] Lucarini, V., Saarinen, J.J., Peiponen, K.-E., Vartiainen, E.M., 2005. Kramers-Kronig relations in optical materials research. Berlin: Springer.

[24] Burg, J.P., 1975. Maximum entropy spectral analysis. Ph.D. Stanford University.

[25] van den Bos, A., 1971. Alternative interpretation of maximum entropy spectral analysis (Corresp.). *Information Theory, IEEE Transactions on*, **17** (4), 493-494.

[26] Hu, Y., Jiang, T., Shen, A., Li, W., Wang, X., Hu, J., 2007. A background elimination method based on wavelet transform for Raman spectra. *Chemometrics and Intelligent Laboratory Systems*, **85** (1), 94-101.

[27] Ramos, P.M., Ruisánchez, I., 2005. Noise and background removal in Raman spectra of ancient pigments using wavelet transform. *Journal of Raman Spectroscopy*, **36**, 848-856.

[28] Kauppinen, J., Partanen, J., 2001. Fourier transforms in spectroscopy. Berlin: Wiley-VCH.

[29] Vartiainen, E.M., 2014. Applied optics, *BM30A0500 Technical physics*. Lappeenranta University of Technology, unpublished. (See lecture 10)

[30] Hiroshi, A., 1970. A new method of interpolation and smooth curve fitting based on local procedures. *Journal of the ACM*, **17** (4), 589-602.

APPENDIX

In this section the sources of Matlab programs are presented. For different spectra separate Matlab files were created. Here only one of them (for the simulated $|\chi^{(3)}|^2$ spectrum, psdmp0.m) is demonstrated. The others can be implemented in a similar way.

1. dft_cos.m

The function computes the spectrum of a symmetrical even signal by using the DCT transform (26).

```
function [ spectrum ] = dft_cos( signal )
%   Discrete cosine transform
%
N = length(signal);
spectrum = zeros(1,N);
for j = 1:N
        spectrum(1) = spectrum(1) + signal(j)/sqrt(N);
end;
for k = 2:N
    for j = 1:N
        spectrum(k) = ...
            spectrum(k) + sqrt(2/N)*signal(j)* ...
                cos(pi/(2*N)*(2*j-1)*(k-1));
    end;
end;
end
```

2. idft_cos.m

The function computes the signal from its spectrum by using the inverse DCT transform (25).

```
function [ signal ] = idft_cos( spectrum )
%   Inverse discrete cosine transform
%
N = length(spectrum);
signal = zeros(1,N);
for k = 1:N
    for j = 2:N
        signal(k) = ...
            signal(k) + sqrt(2/N)*spectrum(j)* ...
                 cos(pi/(2*N)*(j-1)*(2*k-1));
    end;
    signal(k) = signal(k) + sqrt(1/N)*spectrum(1);
end;
end
```

3. burg_imp_resp.m

The function computes the impulse response coefficients by using the expressions (33)–(35).

```
function [ h ] = burg_imp_resp( signal,N_use,M )
%   Burg's impuse response coefficients
%
% initialization
h = zeros(1,M);  % output impulse response coefficients
e = zeros(M,N_use); % residuals (35)
b = zeros(M,N_use); % residuals (35)
for i=1:N_use
    e(1,i) = signal(i);
    b(1,i) = signal(i);
end;

%% Burg's algorithm
% initialization of impulse response coeffitients to be iterated
h1 = zeros(M,M);
for m=1:M
    % initialization of numerator and denumerator of (34)
    S1 = 0; S2 = 0;
    for  i=1:N_use-m
        % numerator of reflection coefficient (34)
        S1 = S1 + e(m,i)*b(m,i+1);
        % denumerator of reflection coefficient (34)
        S2 = S2 + abs(e(m,i))^2 + abs(b(m,i+1))^2;
    end;
    h1(m,m) = 2*S1/S2; % reflection coefficient (34)
    for j=1:m-1
        % equations (29) for response coefficients
        h1(m,j) = h1(m-1,j) - h1(m,m)*h1(m-1,m-j);
    end;
    for k=1:N_use-m
        e(m+1,k) = e(m,k) - h1(m,m)*b(m,k+1); % residuals (35)
        b(m+1,k) = b(m,k+1) - h1(m,m)*e(m,k); % residuals (35)
    end;
end;

h(:) = h1(M,:);

end
```

4. qfactor.m

The function calculates the q-factor (36). The input parameters are the half of an original symmetric even signal and the corresponding predicted version of this part which was predicted knowing another half part of the signal.

```
function [ q ] = qfactor( signal,predicted )
%   Q-factor
%
q1=0; q2=0; % initializing of numerator and denumerator of (36)
N = length(signal);
for j=1:N
    % calculating numerator of (36)
    q1 = q1 + abs(signal(j)-predicted(j))^2;
    % calculating denumerator of (36)
    q2 = q2 + abs(signal(j))^2;
end;
q = 1 - sqrt(q1/q2); % output q-factor

end
```

5. deapodize_and_cut.m

The function returns the signal `s` having been deapodized by some window function `win_func` and cut at the chosen point `N_last`.

```
function [ s ] = deapodize_and_cut( signal,win_func,N_last )
%   Deapodize-and-cut function
%
s = signal./win_func;
for i=N_last+1:length(signal)
    s(i)=0;
end;

end
```

6. predict.m

The function returns the signal `result` having been predicted by `N_fut` samples using the vector `h` of impulse response coefficients. The program allows to choose between two options by setting the parameter `opt` to be `'f'` or `'b'` (forward or backward prediction respectively).

```
function [ result ] = predict( signal,N_fut,h,opt )
%   Linear prediction (either forward or backward)
%
M = length(h);
N_sig = length(signal);
buf_sig = zeros(1,N_fut); % Predicted part of the signal
result = zeros(1,N_fut);

if (opt=='f')
    for i=1:N_sig
        buf_sig(i) = signal(i);
```

```
    end;
elseif (opt=='b')
    for i=1:N_sig
        buf_sig(i) = signal(N_sig+1-i);
    end;
end;

for j=1:N_fut
   for k=1:M
       % forward linear prediction (31)
       result(j) = result(j) + h(k)*buf_sig(N_sig-k+1);
   end;
   for i=1:N_sig-1
       buf_sig(i) = buf_sig(i+1);
   end;
   buf_sig(N_sig) = result(j);
end;

if (opt=='f')
    return;
elseif (opt=='b')
    for i=1:N_fut
        buf_sig(i) = result(N_fut+1-i);
    end;
    result = buf_sig;
end;

end
```

7. predict2sides.m

The function returns the signal `yy` having been predicted by `N_fut` samples both in forward and backward directions using the vector `h` of impulse response coefficients. `N_last` is the last nonzero sample of the cut signal after deapodization and `s` is the input signal.

```
function [ yy ] = predict2sides( s,N_last,N_fut,h )
%    Forward and backward prediction
%
yy = zeros(1,N_last+2*N_fut);
yy(N_fut+1:N_fut+N_last) = s(1:N_last);
% backward prediction
yy(1:N_fut) = predict(s(1:N_last),N_fut,h,'b');
% forward prediction
yy(N_fut+N_last+1:N_last+2*N_fut) = ...
predict(s(1:N_last),N_fut,h,'f');

end
```

8. psdmp0.m

The example of the MatLab procedure that analyzes the given spectrum $|\chi^{(3)}|^2$ by the LOMEP method in order to be narrowed.

```
clear all
close all
clc

% read multi-column spectral data
spectrumDif = dlmread('psdmp0.d');
% forget about x-axis
spectrum = spectrumDif(:,2);
Nl = length(spectrum(:));

% derivation of the signal from the spectrum by the inverse DCT
% (25)
signal = idft_cos(spectrum(:));

N_last1 = 100; % the last nonzero signal sample
N_fut = 1500;  % the number of samples that are to be predicted
N_total = N_fut+N_last1;
q = zeros(N_last1:N_last1);

for k=20:20
    apodization = exp(-(1:1:Nl)/k);
    for N_last = N_last1:N_last1
        for m=52:52  % the max(m) must be equaled to N_last-1
            % iterations' number and impulse response length
            M = N_last-m;
            % number of samples used in Burg's algorithm
            N = N_last;
            s = deapodize_and_cut(signal,apodization,N_last);

            %% find Burg's impulse response coefficients
            % max M allowed is N-1
            h = burg_imp_resp(s,N,M);

            %% linear prediction (both backward and forward)
            % expanding the signal to have N_fut zero values at
            % both sides
            yy = predict2sides( s,N_last,N_fut,h );

            % forming two signals to be compared in the terms of
            % q-factor
            for i=1:N_last
                y1(i) = s(N_last+1-i);
                y2(i) = yy(N_fut-N_last+i);
            end;
            q(N_last,m) = qfactor(y1,y2);
            % 'clear' is needed for iterations in order to compute
            % each variable independently of its value at the
            % previous iteration
```

```
            % clear y1 y2 y3 y4 M N s1 s2 h yy yy2
        end;
    end;
    qq(k) = max(max(q));
end;

%% secondary apodization by exponential decay function
% width parameter b
b = 200;
apodization2 = exp(-((1:1:N_total))/b);

%% plot narrowed spectrum
% normalized x-axis (Raman shifts)
freqs = spectrumDif(1,1):...
    (spectrumDif(Nl,1)-spectrumDif(1,1))/(N_total-1):...
    spectrumDif(Nl,1);
% deapodized and predicted forward signal
yyy(1:N_total) = yy(N_fut+1:N_last+2*N_fut);

figure
subplot(2,1,1); plot(spectrumDif(:,1),spectrumDif(:,2))
title('Original spectrum');
ylabel('Spectral power, a.u.');
xlabel('Raman shift, cm^{-1}');
subplot(2,1,2);
plot(freqs...
    ,dft_cos(deapodize_and_cut(yyy,...
    exp(-(1:1:N_total)/360000),N_total-...
    1).*apodization2(1:N_total)));
title('Narrowed spectrum');
ylabel('Spectral power, a.u.');
xlabel('Raman shift, cm^{-1}');
```