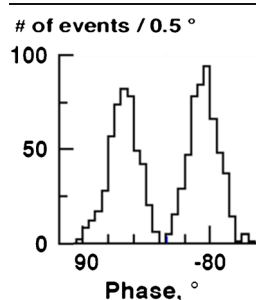




RESEARCH ARTICLE

Least-Squares Fitting of Time-Domain Signals for Fourier Transform Mass Spectrometry

Tagir Aushev,¹ Anton N. Kozhinov,² Yury O. Tsybin²¹Institute for Theoretical and Experimental Physics, 117218 Moscow, Russia²Biomolecular Mass Spectrometry Laboratory, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

Abstract. To advance Fourier transform mass spectrometry (FTMS)-based molecular structure analysis, corresponding development of the FTMS signal processing methods and instrumentation is required. Here, we demonstrate utility of a least-squares fitting (LSF) method for analysis of FTMS time-domain (transient) signals. We evaluate the LSF method in the analysis of single- and multiple-component experimental and simulated ion cyclotron resonance (ICR) and Orbitrap FTMS transient signals. Overall, the LSF method allows one to estimate the analytical limits of the conventional instrumentation and signal processing methods in FTMS. Particularly, LSF provides accurate information on initial phases of sinusoidal components in a given transient. For instance, the

phase distribution obtained for a statistical set of experimental transients reveals the effect of the first data-point problem in FT-ICR MS. Additionally, LSF might be useful to improve the implementation of the absorption-mode FT spectral representation for FTMS applications. Finally, LSF can find utility in characterization and development of filter-diagonalization method (FDM) MS.

Key words: Fourier transform (FT), Fourier transform mass spectrometry (FTMS), Transient signal, Least-squares fitting (LSF), Filter-diagonalization method (FDM), Absorption mode, Phase correction, First data-point problem, Ion cyclotron resonance (ICR), Orbitrap

Received: 1 November 2013/Revised: 17 March 2014/Accepted: 17 March 2014/Published online: 1 May 2014

Introduction

Structural analysis of complex biological and environmental samples greatly benefits from high resolution and high mass accuracy provided by Fourier transform mass spectrometry (FTMS) [1–4]. However, further improvement of FTMS analytical characteristics is required. Advanced signal processing in the time domain has recently gained a particular attention in FTMS. Fourier transform (FT) with absorption-mode spectral representation is, perhaps, the most striking example of recent progress in applied advanced signal processing aimed at increasing the resolving power compared to that of the magnitude-mode FT employed typically, for transients with the same lengths. Specifically, the maximum gain in resolving power of the absorption mode vs. the magnitude mode equals two. The principle of absorption-mode FTMS originated in Fourier transform ion

cyclotron resonance mass spectrometry (FT-ICR MS) and FT nuclear magnetic resonance (NMR) spectroscopy [5–7]. Recently, the absorption mode has received a particular attention in FT-ICR MS method development and applications [8–12]. In Orbitrap FTMS, the absorption mode is of a high interest as well and has recently been implemented as a principal part of the algorithm known as enhanced FT (eFT) [4, 13, 14]. Unlike FT-ICR MS, Orbitrap FTMS allows for a straightforward implementation of the absorption-mode spectral representation since there exists a time point at which the time-dependent phases of all ions trapped in the orbitrap are equal in the first-order approximation. Additionally, to correct for higher-order phase deviations, an accurate estimate of the actual phase function for particular experimental conditions is required. The implementation of the absorption-mode spectral representation on other electrostatic ion traps is similar to that on Orbitrap FTMS and relatively easy in comparison to FT-ICR MS [15].

Methods of non-FT signal processing refer to diverse spectral and parameter estimators and are similar to those applied for data processing in FT NMR spectroscopy. These include the maximum entropy method (MEM), linear

Electronic supplementary material The online version of this article (doi:10.1007/s13361-014-0888-x) contains supplementary material, which is available to authorized users.

Correspondence to: Yury O. Tsybin; e-mail: yury.tsybin@epfl.ch

prediction method (LPM), and other methods. However, earlier attempts to implement these methods in FTMS have suffered from numerical limitations and drawbacks familiar from the FT NMR spectroscopy. Among the more modern methods of non-FT signal processing, filter diagonalization method (FDM), an optimized parameter estimator, has demonstrated a particularly robust performance for 1D NMR. Nevertheless, the FDM-based processing of experimental ICR transients for analytical applications has been accomplished only recently [16]. In MS applications, due to its uncertainty principle, the FDM is able to provide the required resolution performance for shorter transients compared to FT-based signal processing. To take advantage of the FDM's uncertainty principle, it is required that ion packets corresponding to different mass-to-charge ratios, m/z , are sufficiently coherent on the time scale of interest. That is, on the phase plane the characteristic size of the ion packets does not exceed the distance between them at the end of ion detection [17]. Further characterization of FDM MS is thus of an interest and shall benefit from implementation of a reference signal processing method such as the least-squares fitting (LSF).

The LSF method is used in various branches of science and its robust numerical implementation is a must in scientific software packages. For instance, the LSF method is often applied implicitly as in various routine problems involving extraction of mean values and estimation of errors from a statistical set of measurements. Moreover, the LSF method's applications include comprehensive fitting problems where experimental data is to be fitted with a theoretical function in order to measure certain quantities of interest. For instance, the LSF is useful in high energy physics, where scarce statistics requires robust and sensitive methods of signal retrieval, and the number of signals to locate is limited [18]. In the MS field, the applications range from improving accuracy of molecular mass measurements to ion mobility mass spectrometry. Specifically, in FTMS the LSF-type methods have been applied previously to improve the peak shape representation of the m/z (or frequency), already Fourier transformed, data. However, the peak-shape fitting in the frequency domain may imply reduced performance due to non-linear spectral interference effects and manual restriction of the fitted frequency range. Alternatively, fitting the raw data, *viz.* the transient signal, is justified. Indeed, recall that in FTMS the metrological basis for measurements of mass-to-charge ratios of ions is provided by the theory of the transient signal [19–21]. For a given ion, the transient signal consists of several sinusoidal components that correspond to three eigen frequencies of ion motion, as well as their harmonics and interharmonics. Among these, the component with reduced cyclotron frequency is normally made to contain the most of spectral energy. For example, it is due to this theoretical form that the ion's m/z can be obtained from a corresponding peak maximum in the Fourier spectrum plotted in the magnitude or absorption mode. Similarly, the theoretical form enables

other methods to be applied to a transient signal in order to obtain the frequencies, abundances, and phases of analyzed ions. The LSF method is a reasonable choice as it is consistent with the fitting problem of interest and is able to provide the required performance. To the best of our knowledge, application of LSF methods for processing of experimental time-domain signals in FTMS has not been reported yet in the peer-reviewed literature, although advantages of its implementation have been named in patent applications [22, 23] as well as the LPM algorithm, an LSF-type method, has shown utility for frequency chasing of simulated FT-ICR MS transients [24].

In the context of the least-squares fitting of FTMS transient signals, possible deviation of experimental transients from the theoretical form requires an additional discussion. Specifically, the effect of signal damping, which develops during the ion detection, can be detrimental if present and not taken into account when an experimental transient is fitted. In modern FTMS, the signal damping is likely due to inharmonicity of the electric field in the mass analyzer, rather than other effects such as insufficient level of vacuum in the mass analyzer as in early FT-ICR mass spectrometers. This conclusion follows from the contemporary studies where the harmonization of the electric field in a mass analyzer leads to a significant increase of the transient's lifetime relative to that typically obtained for regular mass analyzers [25]. Nowadays, the signal damping effect cannot be properly taken into account in LSF calculations because the theoretical basis of the damping mechanism is not well-developed. However, regardless of the exact damping mechanism, in the current work the theoretical form with sinusoidal components is applicable since the damping effect is not noticeable at the time scale of interest.

Using standard software for scientific computing, here we implement the least-squares fitting of experimental transients in FTMS. Performance of the LSF method is evaluated on single and five-component transients. Comparisons between the conventional magnitude-mode FT processing and LSF are made for relatively short transients, emphasizing the potential use of LSF as a reference method for the development of FDM MS. Advantages and limitations of LSF processing are outlined.

Experimental, Materials, and Methods

Sample Preparation

Peptides Substance P and MRFA were obtained from Sigma-Aldrich (Buchs, Switzerland). LC-MS grade acetonitrile and water were obtained from Fluka (Buchs, Switzerland). Formic acid was obtained from Merck (Zug, Switzerland). Peptide solutions were prepared in 1:1 (v/v) water/acetonitrile solvent mixtures containing 1% (v/v) of formic acid.

Mass Spectrometry

The MS experiments were performed on a hybrid linear ion trap Fourier transform ion cyclotron resonance mass spectrometer with a 10 T superconducting magnet (LTQ FT-ICR MS Ultra; Thermo Scientific, Bremen, Germany) and a hybrid dual linear ion trap Orbitrap Fourier transform mass spectrometer with a high-field compact Orbitrap mass analyzer (Orbitrap Elite FTMS; Thermo Scientific). Ions were produced with a nano-electrospray ionization (nESI) ion source (TriVersa Nanomate; Advion Biosciences, Ithaca, NY, USA). Standard instrument control and data acquisition systems were utilized (Thermo Scientific). The total number of charges injected into the ICR cell or the Orbitrap was controlled with the automatic gain control (AGC) function. For both mass spectrometers, the target total charge (AGC value) was set to the default value of 2×10^5 . The transient signals were acquired in MIDAS data format. For both mass spectrometers, the selected sampling frequency, f_s , of the built-in analog-to-digital converter (ADC) corresponded to the typical broadband frequency (mass-to-charge) range and was equal to 2.7306(6) MHz (16.384/6 MHz). For the purpose of the current article, the measurements for two types of analytes were taken: isolated monoisotopic ions of singly charged peptide MRFA, m/z 524 (single peak, or a singlet), and an isolated isotopic distribution of a doubly charged peptide Substance P, m/z 674 (five peaks, or a multiplet). Approximately 1000 single-scan transients were acquired for each set of measurements. The length of transient signals was 96 ms; further, each transient was cut to the length, T , of 24 ms unless stated otherwise. The criterion for the choice of the transient length was based on two limiting factors: its length should be sufficient for FT signal processing to baseline-resolve the five isotopic peaks of Substance P for both mass spectrometers employed; it should not provide excessive resolution of the five peaks. Specifically, 24 ms was the minimum transient length required to baseline-resolve the isotopic envelope of Substance P analyzed on the 10 T FT-ICR MS. Additionally, the chosen transient length is consistent with the requirement to evaluate the LSF method as a reference method for further FDM MS characterization.

Data Analysis

FT processing of transient signals was performed using the framework pyFTMS, which was developed in-house and written in Python and C/C++ programming languages. The transients were processed following the conventional FTMS workflow. Specifically, the transient signals first were apodized with the von Hann window and zero-filled once. Next, Fourier transform was applied to convert the time-domain signals into the frequency-domain spectra with magnitude-mode spectral representation. Finally, the spectra were peak-picked using the standard three-point parabolic interpolation of local maxima [26–28].

To implement a parametric function that is to be minimized for a given transient signal and to automate the minimization procedure, a custom software written in C++ programming language was used. The transient signal, being a sequence of instant voltages v_n , Figure 1 and Supplementary Figure S1 (Supporting Information), where $n = 0, \dots, N-1$, and N is the total number of sampled points: $N = f_s T$, was parameterized by the fitting function F_n defined as a sum of K sinusoidal components:

$$F_n = \sum_{k=1}^K A_k \sin(2\pi f_k n t_s + \varphi_k), n = 0, \dots, N-1, \quad (1)$$

where t_s is the sample time, $t_s = 1/f_s$. The parameters to find are frequency f_k , amplitude A_k , and initial phase φ_k of the k^{th} sinusoidal component present in the transient. The initial phases corresponded to the custom-defined beginning of the transient signal. Since the lifetime of a transient signal was typically much greater than the transient duration analyzed, any detrimental effects (e.g., ion packet incoherence that develops with time) were not taken into account when constructing the function F_n , Equation (1).

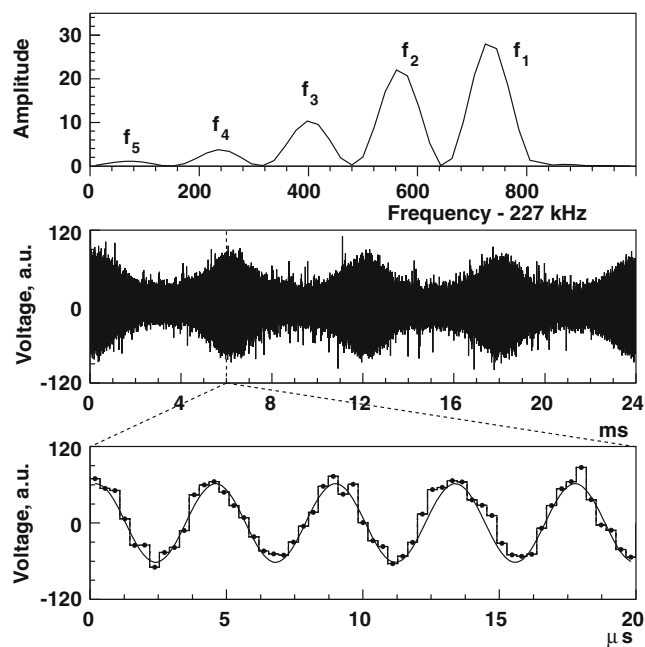


Figure 1. The principle of least-squares fitting (LSF) of transient signals. Top panel shows the magnitude-mode Fourier spectrum of a 24 ms-long experimental transient signal shown in the middle panel. Bottom panel shows an expanded view of the transient with sampled points and a curve corresponding to the fitting function. The sampled points are connected with a stair-step line for visual convenience only. The transient signal was obtained in the analysis of doubly charged peptide Substance P on the 10 T FT-ICR MS

The function to minimize was the χ^2 function defined as:

$$\chi^2 = \sum_{n=0}^{N-1} (F_n - v_n)^2. \quad (2)$$

For minimization purpose, the MINUIT package was used [29]. The parameters f_k , A_k , and φ_k were varied to locate the minimum deviation of the function, F_n , from the digitized transient signal, v_n , $n = 0, \dots, N-1$, in terms of the minimum value of χ^2 , Figure 1 and Supplementary Figure S1 (in Supporting Information). Initial seeds for amplitudes and frequencies were obtained with the FT signal processing described above, while the seed values for initial phases were set to zeros. The fit provided the set of parameters f_k , A_k , and φ_k for each sinusoidal component included in the fit function F_n as well as the minimum value of χ^2 . The calculations were carried out on a standard laptop computer. Typical LSF processing time for a single transient was on the order of 1 s.

Modeling of Transient Signals

Modeled transients were constructed using sinusoidal signals with given frequencies, amplitudes, and initial phases. When required, random noise with a given standard deviation σ was added to a transient signal. All these parameters were obtained from the experimental data unless stated otherwise. Specifically, the employed parameters of frequencies, amplitudes, and initial phases were the mean values of the corresponding experimental distributions; the standard deviation σ was obtained as $(\chi^2/N)^{1/2}$. The modeled transients were processed with FT and LSF methods, which were implemented and applied as described above for the analysis of experimental transient signals.

Results and Discussion

Least-Squares Fitting of a Single-Component Transient Signal

As the first step, 1000 single-scan ICR transients (an example is shown in Supplementary Figure S1, in Supporting Information) and 1000 single-scan Orbitrap transients containing the singlet were processed using LSF and FT methods, Figure 2. For the set of ICR transients, the amplitude, frequency, and phase distributions obtained with LSF and FT methods are shown in Figure 2, top panel; the corresponding mean values and standard deviations are listed in Table 1, top section. Here, while the amplitude distributions obtained with LSF and FT are similar, the frequency distribution obtained with LSF method is narrower than the one obtained with FT processing. Additionally, the mean values of the frequency distributions differ for LSF and FT processing.

To verify which method provided more accurate values of ion frequencies, the LSF and FT calculations were performed for a set of 1000 modeled FT-ICR MS transient signals, which were generated using the mean amplitude, frequency, and initial phase values, as well as the noise standard deviation obtained from the LSF processing of the ICR transients (the values in question are shown with red lines in Figure 2, middle panel and Table 1, middle section of the rightmost column). Since that as the LSF processing revealed two modes in the phase distribution of the ICR transients, *vide infra*, the initial phases for the modeled FT-ICR MS transients were randomly set to either of the two phase values obtained for experimental transients.

The amplitude, frequency, and phase distributions obtained with FT and LSF methods for the modeled transients are shown in Figure 2, middle panel; their mean values and standard deviations are given in Table 1, middle section. The frequency distributions obtained for the modeled FT-ICR MS transients are in agreement with those obtained for the ICR transients. Specifically, for both experimental and simulated sets of transients, the same gain in frequency precision (ratio of standard deviations), ~ 1.6 , of the LSF method relative to the FT signal processing was obtained. Ergo, in Figure 2, top panel, the increased standard deviation of the FT frequency distributions can be assigned to the processing loss of the FT-based signal processing workflow, whereas the LSF method can be said to provide more precise frequency values.

Next, consider the mean frequency values obtained for the ICR and modeled transients, Figure 2 and Table 1. For the modeled FT-ICR MS transients, the mean frequency obtained with the LSF method is unbiased compared with the mean frequency obtained with the FT processing, which gives a noticeable shift from the true frequency value. Therefore, for the frequency distributions obtained for the ICR transients, Figure 2, top, one may conclude that the LSF frequency distribution is unbiased, whereas FT processing produced shifted frequency values. We consider this deviation not to be due to physical effects of ion motion or ion detection but rather to expected artifacts of the conventional signal processing workflow in FTMS [28].

To identify the exact numerical cause of this bias, the following numerical experiment was performed. Recall that the discrete Fourier transform (DFT), which is employed in practice to analyze sampled transient signals, returns a replicated and, most importantly, sampled version of a continuous Fourier spectrum, whose maxima would provide the frequency values and abundances of interest. Specifically, the acquisition time T and single zero filling result in frequency spacing of $\Delta_j f = 1/(2T)$ in the sampled spectrum (in our case $T = 24$ ms and hence $\Delta_j f \approx 21$ Hz). Additionally, the full width of a spectral peak of interest, which is defined by the Hann apodization window and magnitude-mode spectral representation in

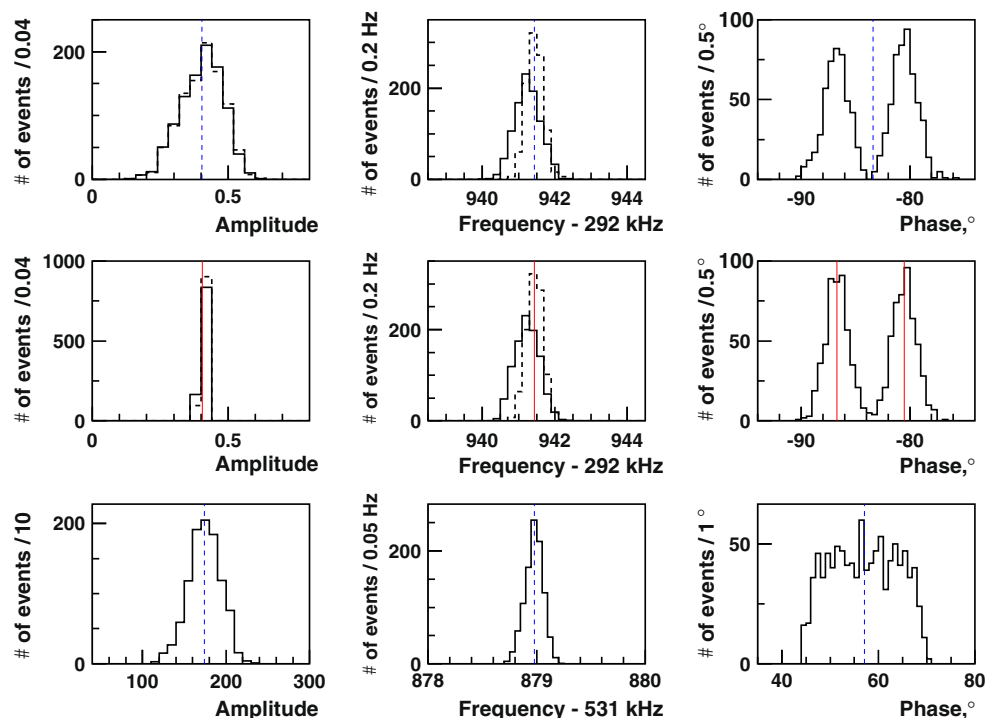


Figure 2. Results of LSF (solid lines) and FT (dashed lines) processing of single-component transients: (top panels) 1000 ICR transients; (middle panels) 1000 simulated transients with frequency $f = 292941.44$ Hz, amplitude $A = 0.405$, standard deviation of noise $\sigma = 0.65$, randomly assigned phases $\varphi = -86.7^\circ$ and $\varphi = -80.5^\circ$; and (bottom panel) 1000 Orbitrap transients. The analysis was performed for a monoisotopic ion of peptide MRFA (m/z 524). Red lines show the parameters used for the transient signal modeling, see Table 1 middle section of the rightmost column. Blue lines show the results of FT processing obtained from the averaged set of 1000 single-scan 96 ms long experimental transients, see Table 1 top and bottom sections of the rightmost column

our case, is $\Delta_{\text{wnd}}f = 4/T$. Therefore, in the sampled spectrum, the peak of interest has $\Delta_{\text{wnd}}f / \Delta f = 4 \cdot 2 = 8$ points per its full width. Among other effects, the accuracy of frequency and abundance values is governed by the peak picking procedure, which is the standard three-point parabolic interpolation of the peak maximum in our case. To investigate the influence of the peak picking procedure on the accuracy of frequency values, we generated 10 modeled transient signals with lengths $T =$

24 ms and frequencies f_n , $n = 0, \dots, 9$, spanning the interval $f_0 \leq f_n \leq f_0 + \Delta f$, where $f_0 = 292941.44$ Hz is the mean frequency obtained with LSF processing of the ICR transients, Table 1, top section. This way, the values $\Delta_{\text{wnd}}f$ were equal for these modeled transients, whereas the eight spectral points were differently distributed over the peak shape in their magnitude-mode Fourier spectra. For these transients, the frequencies of peak maxima, f , measured using the FT signal processing workflow deviated from the corresponding values f_n as the

Table 1. Mean Values and Standard Deviations of Frequencies, Amplitudes, and Initial Phases Obtained with LSF and FT Methods in the Analysis of the Single-Component ICR Transients (Figure 2 top), Simulated Transients (Figure 2 middle), and Orbitrap Transients (Figure 2 bottom). The Rightmost Column Shows Values Obtained with LSF for the Averaged Set of 1000 Experimental Transients of 96 ms Length and True Values Used for Transient Signal Modeling

Parameters	LSF mean	LSF std	FT mean	FT std	FT 96 ms (averaged) /true
ICR experimental data					
Amplitude	0.405	0.076	0.408	0.080	0.404
Frequency, Hz	292941.44	0.22	292941.25	0.35	292941.44
Phase, $^\circ$	-86.7/-80.5	1.1	N/A	N/A	-83.447
ICR simulated data					
Amplitude	0.405	0.004	0.405	0.005	0.405
Frequency, Hz	292941.44	0.22	292941.24	0.34	292941.44
Phase, $^\circ$	-86.7/-80.5	1.1	N/A	N/A	-86.7/-80.5
Orbitrap experimental data					
Amplitude	175	19	173	19	173.65
Frequency, Hz	531878.97	0.09	531878.70	0.08	531878.97
Phase, $^\circ$	57	7	N/A	N/A	57.098

peak shape relocates over the eight frequency points, which are fixed. That is, since the three-point interpolation is only an approximation of the actual peak shape, which is the Hann spectral function in our case, the peak picking procedure leads to the systematic errors in frequency measurements as follows from the correlation between the measurement error $\Delta f = f - f_n$ and the true frequency f_n , Supplementary Figure S2 (in Supporting Information). Thus, we conclude that the frequency bias in question, Figure 2, top panel, demonstrates the limited accuracy of the peak interpolation procedure of the conventional FT signal processing workflow. Moreover, this conclusion is also in agreement with the results obtained in a separate FT analysis of the averaged set of longer, $T = 96$ ms, 1000 ICR transients, as shown with blue lines in Figure 2, top panel: since the full width $\Delta_{\text{wnd}f}$ is inversely proportional to the transient's length, T , the frequency bias in question reduces as the time T increases from 24 ms to 96 ms.

Additionally, the amplitude distributions obtained for the modeled FT-ICR MS transients, Figure 2, middle panel, are similar for the both methods, as well as the two obtained previously for the ICR transients, Figure 2, top panel. However, the former are narrower than the latter. Hence, unlike the numerical effects in frequency distributions discussed above, broadening in the amplitude distributions obtained for the ICR transients, Figure 2, top panel, indicates influence of physical effects (e.g., the scan-to-scan variation in the number of ions injected into the mass analyzer).

Importantly, in addition to amplitude and frequency values, the LSF method provides the initial phases of the sinusoidal components, shown as distributions in Figure 2, right panels. The corresponding mean values and standard deviations are given in Table 1. Interestingly, for the ICR transients the phase distribution obtained with LSF processing contains two prominent peaks instead of one. That is, the initial phase of each transient belongs either to the first or second modes of the phase distribution. The observed phase difference between the two modes corresponds to $\sim 1/16$ μs , which translates into ~ 16 MHz in the frequency scale. Notably, a multiple of the latter equals the frequency of the quartz generator, 32.768 MHz, of the analog-to-digital converter employed. Therefore, the two-mode phase distribution is likely due to the finite precision of the electronics synchronization. We conclude this phenomenon to be of a random nature since no correlation with the scan number has been found, as demonstrated below for the analysis of five-component transients. Thus, we assign the appearance of the two modes to the “first data-point problem,” which includes a number of effects in digitization of transient signals. For instance, in FT NMR spectroscopy, it has been previously reported that the “first data-point problem” resulted in peak shape artifacts in absorption-mode FT NMR spectra.

For the set of Orbitrap transients, the amplitude, frequency, and phase distributions obtained with LSF and FT methods are shown in Figure 2, bottom panel; the corresponding mean values and standard deviations are given in Table 1, bottom section. The frequency distributions do not demonstrate a noticeable difference in their standard deviations, Table 1. Hence, we conclude that in contrast to the analysis of the ICR transients, here the experimental deviation of ion frequencies is the major contribution to the obtained standard deviations, whereas the processing loss of the FT workflow was less significant. The obtained amplitude distributions are analogous to those of the FT-ICR MS data analysis, Figure 2, top panel. The phase distribution is, however, substantially different from the FT-ICR MS results: a single rather wide phase distribution with a flat top is observed, *vide infra*.

Least-Squares Fitting of a Multiple-Component Transient Signal

As the next step, we analyze experimental transients containing the isotopologues from five isotopic fine-structure clusters of peptide Substance P, Figure 1. Given the time scale of interest, the LSF method should resolve the five isotopic peaks, whereas the isotopologues of each fine-structure cluster are likely beyond the resolution performance of the LSF method. Therefore, the following numerical experiment with a modeled transient signal was performed in order to test the LSF's applicability to transients with fine-structure ions. Using the m/z values and abundances of all isotopologues from the five fine-structure clusters of doubly protonated Substance P, we generated a 24 ms-long transient signal composed of sinusoidal components with corresponding amplitudes and frequencies, Supplementary Table S1 (in Supporting Information) left columns. For each cluster of isotopologues, the differences in phases accumulated by the ions during the delay between the end of the ion excitation and beginning of the ion detection can be neglected. Hence, for the sinusoidal components of the generated transient, the initial phases were set to zero.

In LSF analysis of this transient signal, for each cluster of isotopologues the provided frequency should be about the average frequency calculated in the sense of weighted average with weights equal to the corresponding abundances. In turn, the amplitudes are expected to be close to the sum of abundances of the isotopologues in a given cluster. These expected values of frequencies and amplitudes are listed in Supplementary Table S1, center columns 9, in Supporting information). The frequencies and amplitudes obtained in the LSF analysis of the modeled transient confirm these points, Supplementary Table S1, right columns (in Supporting Information) and, hence, validate the LSF's applicability to the experimental transients with fine structure ions.

We thus proceed to the analysis of the experimental transients in question. Figure 3 shows frequency, amplitude, and phase distributions obtained with LSF and FT processing of 1000 ICR transient signals containing the ions of five isotopic fine-structure clusters of Substance P. Table 2 details the corresponding mean and standard deviation values. For all five components, the LSF processing provided better frequency precision than that of the FT processing: the frequency distributions obtained with LSF are narrower compared with those obtained with FT signal processing, Figure 3 top and Table 2, similar to the analysis of the ICR transients containing the singlet from the previous subsection.

Additionally, systematic errors were observed in the FT results: the frequency distributions obtained with the FT processing shift toward higher frequencies (e.g., peak f_3 at ~ 227400 Hz), or toward lower frequencies (e.g., peak f_1 at 227737 Hz), as follows from Figure 3, top panel and Table 2. These systematic errors negatively influence the overall accuracy of ion frequency measurement and would translate into the corresponding mass errors upon calibration. These deviations are presumably due to the spectral artifacts of the conventional FT signal processing workflow. These

include the limited accuracy of the peak maximum interpolation procedure and the nonlinear interference of spectral components [28]. The former was discussed in the above analysis of the transients containing the singlet. As for the latter, note that the resolving power was just the minimum required to baseline-resolve the peaks in question, Figure 1, top panel. Therefore, the maximum of a given peak is somewhat influenced by the tails of other peak shapes [30]. One could expect that similar to the singlet data, accurate interpolation of each peak shape would solve the problem of spectral interference. However, for the magnitude-mode spectra, numerical implementation of such procedure is complicated as it implies a nonlinear summation of different peak shapes at a given frequency. On the other hand, as the acquisition time increases, a better accuracy of the peak interpolation procedure is achieved as discussed in the previous subsection, and the spectral interference effect reduces because of the increased resolution of peaks [28, 30, 31]. A separate FT analysis of a set of longer, $T = 96$ ms 1000 ICR transients indicates the following results, as shown with blue lines in Figure 3, top panel and Table 2, rightmost column. Relative to the frequencies obtained with LSF processing of 24-ms transients, the frequencies obtained

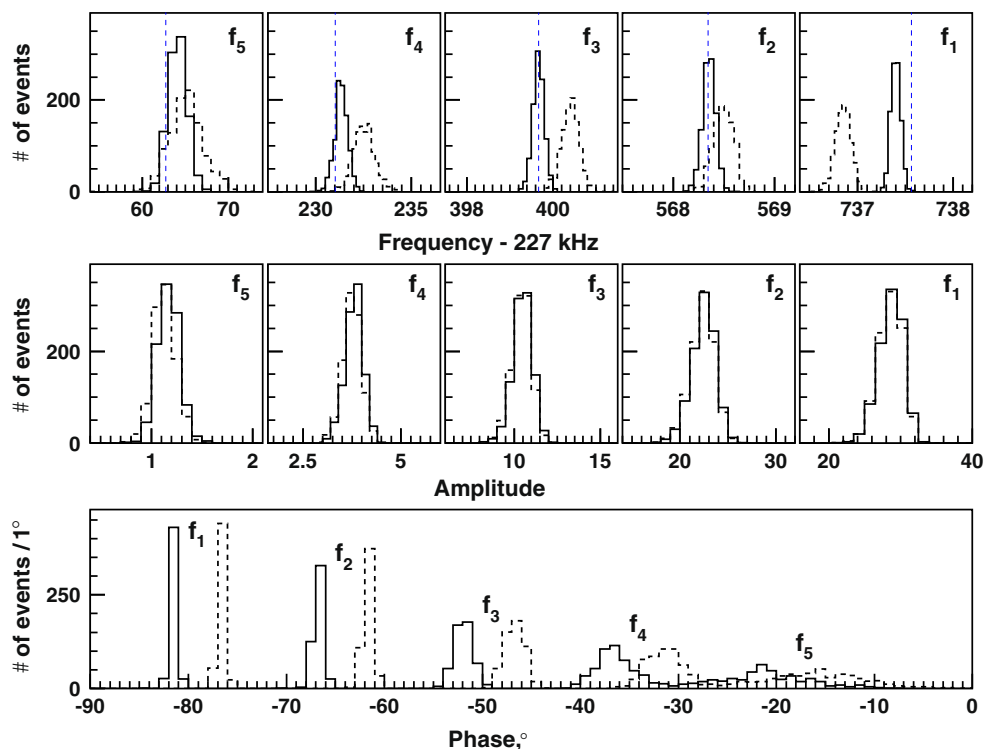


Figure 3. Results of LSF and FT processing of ICR multiple-component transients. Top and middle panels show frequency and amplitude distributions obtained with LSF (solid lines) and FT (dashed lines). Bottom panel shows phase distributions obtained with LSF; solid and dashed lines correspond to the two-mode split of the phase distributions. Bin widths for frequency distributions (top panel, from left to right): 1.0, 0.2, 0.1, 0.05, and 0.04 Hz; for amplitude distributions (middle panel, from left to right): 0.1, 0.2, 0.5, 1.0, and 1.5; and phase distribution (bottom panel): 1° . The set of 1000 transients with five components of an isotopic distribution of a doubly charged peptide Substance P was obtained experimentally on the 10 T FT-ICR MS, see Figure 1. Blue lines show the mean values obtained with FT processing for the set of longer, $T = 96$ ms, 1000 ICR transients, see Table 2 rightmost column

Table 2. Mean Values and Standard Deviations of Frequencies, Amplitudes, and Initial Phases Obtained with LSF and FT Methods for the ICR Five-Component Transients (Figure 3). The Rightmost Column Shows the Mean and Standard Deviation Values Obtained with FT Processing for the Set of Longer, $T=96$ ms, 1000 ICR Transients

Parameters	LSF mean	LSF std	FT mean	FT std	FT 96 ms
Amplitude 1	28.8	1.6	28.6	1.6	26.8±1.6
Frequency 1, Hz	227737.40	0.05	227736.85	0.08	227737.56±0.03
Phase 1, ^o	-81.5/-79.6	0.3	N/A	N/A	N/A
Amplitude 2	22.4	1.2	22.3	1.2	20.5±1.2
Frequency 2, Hz	227568.35	0.06	227568.49	0.10	227568.34±0.03
Phase 2, ^o	-66.7/-61.6	0.5	N/A	N/A	N/A
Amplitude 3	10.5	0.5	10.4	0.5	9.2±0.5
Frequency 3, Hz	227399.67	0.13	227400.40	0.19	227399.67±0.05
Phase 3, ^o	-51.9/-46.7	0.9	N/A	N/A	N/A
Amplitude 4	3.81	0.22	3.76	0.23	3.2±0.2
Frequency 4, Hz	227231.32	0.34	227232.55	0.52	227231.03±0.11
Phase 4, ^o	-36.6/-31.4	1.8	N/A	N/A	N/A
Amplitude 5	1.17	0.10	1.13	0.11	0.9±0.1
Frequency 5, Hz	227064.1	1.0	227065.1	1.9	227062.8±0.3
Phase 5, ^o	-21.9/-16.4	5.0	N/A	N/A	N/A

with FT processing of 96-ms transients deviate less than the frequencies obtained with FT processing of 24-ms transients. Hence, this agreement with the theoretical aspects given above confirms the systematic deviations in question, Figure 2, top and Table 2, to be due to the spectral artifacts of the conventional FT signal processing. Although the increase of the acquisition time can potentially lead to variation of the ion frequency, *vide infra*, the increased frequency error attributable to the latter apparently was less than the decreased frequency error due to the spectral artifacts under consideration.

The amplitude distributions obtained for the considered transient signals, Figure 3, middle panel and Table 2, are consistent with the theoretical abundances of unresolved isotopic clusters of doubly protonated Substance P, Supplementary Table S1, center columns (in Supporting Information). The obtained distributions of initial phases are shown in Figure 3, bottom panel. Here, the phase distribution for each of the sinusoidal components is split between the two families (solid and dashed lines), similar to the analysis of the ICR single-component transients described in the previous subsection, Figure 2, top panel. As previously, comparison of the quartz frequency and the distances between the phase distributions of these five components confirms the role of the “first data point problem.” Supplementary Figure S3 (Supporting Information) indicates a random nature of phase allocation to the first and second modes. Another striking feature of the phase distributions shown in Figure 3 is the high phase coherence, especially evident for the most abundant monoisotopic and A+1 isotopic peaks, as follows from their narrow distributions.

Additionally, a set of modeled FT-ICR MS transients containing the multiplet was generated using the mean values of the distributions in question, Figure 3 and Table 2. The results of the FT and LSF processing for these transients are shown in Supplementary Figure S4 and Supplementary Table S2 (in Supporting Information) and are similar to those obtained for the experimental transients, Figure 3.

Figure 4 demonstrates application of LSF processing to the set of 1,000 Orbitrap FTMS transients containing the multiplet. Table 3 lists the corresponding mean values and standard deviations. The amplitude distributions correlate with those of FT-ICR MS data analysis, Figure 3. For the frequency distributions, the behavior of their mean values is analogous to that of the results for the ICR transients containing the multiplet, Figure 3, whereas the standard deviations are the same for the two methods, similar to the analysis of the Orbitrap FTMS transients containing the singlet, Figure 2, bottom panel. The phase distributions, Figure 3 bottom panel, also follow the behavior found for the single-component Orbitrap FTMS transients, Figure 2, bottom panel. The spreads of phases for all five isotopic peaks are comparable. Additionally, the FT and LSF results obtained on modeled Orbitrap FTMS transients containing the multiplet and generated using the parameters from Table 3 are presented in Supplementary Table S3 (in Supporting Information).

Least-Squares Fitting for FTMS Characterization and Development

The phase information provided by LSF may be particularly useful for correction of phase functions and improving the accuracy of absorption mode FT spectral representation. The nature of the reported phase coherence in FT-ICR MS is, apparently, due to the ion excitation process by rf electric field and low level of phase de-coherence of ion packets on the considered time scale. The observed phase coherence of ion packets and its negligible scan-to-scan variation are remarkable.

Contrary to FT-ICR MS, ion excitation in Orbitrap FTMS is done upon ion injection into the Orbitrap mass analyzer from the C-trap. As a result, finite jitter between ion ejection from the C-trap and ion capturing into stable trajectories inside of the Orbitrap (achieved upon pulsing of central electrode potential) may lead to a scan-to-scan variation of

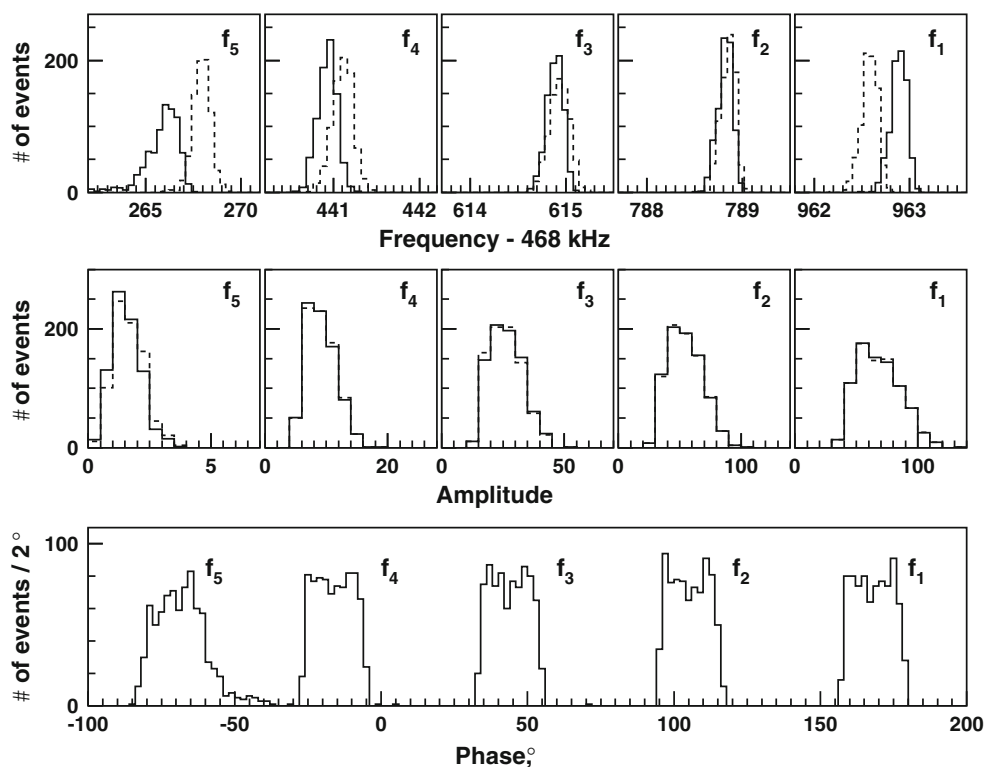


Figure 4. Frequency, amplitude, and phase distributions obtained with LSF (solid lines) and FT (dashed lines) processing of multiple-component Orbitrap transients. The set of 1000 transients with five components of an isotopic distribution of a doubly charged peptide Substance P was obtained experimentally on the Orbitrap Elite FTMS. Bin widths for frequency distributions (top panel, from left to right): 0.3, 0.08, 0.06, 0.06, and 0.06 Hz; for amplitude distributions (middle panel, from left to right): 0.5, 2, 5, 10, and 10; and phase distribution (bottom panel): 2°

the phases of ion packets. Significant reduction of the jitter time on more recent Exactive series Orbitrap FTMS, down to 20–30 ns, provides correspondingly narrower phase distributions [17]. Additionally, of equal importance is the precision of the electronics synchronization between the ion injection event and the signal acquisition event. Therefore, provided that the jitter time is sufficiently low and the data

acquisition trigger is sufficiently precise, the phase distribution of ions in the Orbitrap mass analyzer can presumably be narrower and exhibit the ultimate two-mode behavior shown by the FT-ICR MS data. Nevertheless, the obtained phase distributions are well-defined for successful transient signal averaging and absorption-mode phasing for both ICR and Orbitrap FTMS (including the eFT algorithm for the latter).

Complementary applications of the LSF method include the analysis of time-dependent quantities of the transient signal. Specifically, LSF can be used in studies of deviations of the ion frequency along the transient or, equivalently, the nonlinear variation of the total phase accumulated by an ion during the detection period. For example, the LSF analysis shows that the frequency (to be exact, the frequency averaged over the detection period), depends on the transient duration, Supplementary Figure S5 (in Supporting Information). The variation of the average frequency implies the corresponding change in the instantaneous frequency along the transient signal. This finding correlates with the study by Aizikov et al. [32]. Preliminary results demonstrate that the current implementation of LSF allows for reduction of transient duration to 12 ms (i.e., two isotopic beats for the multiplet, Figure 1) without drawbacks for method performance, whereas shortening transient duration further to 6–12 ms, for the transients in question, results in noticeable deviations in parameters returned by LSF. Presumably, this

Table 3. Mean Values and Standard Deviations of Frequencies, Amplitudes, and Initial Phases Obtained with LSF and FT Methods for the Orbitrap Five-Component Transients as Presented in Figure 4

Parameters	LSF mean	LSF std	FT mean	FT std
Amplitude 1	68	17	68	17
Frequency 1, Hz	468962.89	0.08	468962.58	0.08
Phase 1, $^\circ$	168	6	N/A	N/A
Amplitude 2	54	14	54	14
Frequency 2, Hz	468788.82	0.08	468788.85	0.08
Phase 2, $^\circ$	105	6	N/A	N/A
Amplitude 3	26	7	26	7
Frequency 3, Hz	468614.89	0.08	468614.93	0.11
Phase 3, $^\circ$	44	6	N/A	N/A
Amplitude 4	9.1	2.4	9.2	2.4
Frequency 4, Hz	468440.92	0.11	468441.10	0.12
Phase 4, $^\circ$	-16	6	N/A	N/A
Amplitude 5	1.6	0.6	1.7	0.6
Frequency 5, Hz	468265.97	0.67	468267.92	0.44
Phase 5, $^\circ$	-67	13	N/A	N/A

behavior shows restrictions of the LSF's uncertainty principle for the case of a transient signal composed of sinusoidal components with regularly spaced frequencies.

Advantages and Limitations of LSF Processing for FTMS

To summarize, advantages of the LSF processing for FTMS based on the examples in the current study can be seen both for the FTMS hardware and signal processing characterization as well as improvement of FTMS analytical performance. Specifically, LSF provides information on the phases of the transient components, presumably more accurate than the phases typically available from the FT processing. The LSF method also returns more precise values for frequencies, whereas the FT processing may lead to an increased standard deviation as a processing loss. It thus might be envisioned that these advantages translate into improved resolution (including influence of the improved phase function) and mass accuracy (via improved frequency and m/z precision) in FTMS. Additionally, since it considers the full-scale harmonic inversion problem, Equation (1), as is, the LSF method can be used as a reference method for development and characterization of the filter diagonalization method (FDM) MS.

A large number of fitting parameters complicates the LSF to converge. Therefore, the limitations of routine LSF implementation in FTMS are primarily due to the upper limit of the number of sinusoidal components to be fitted. As of today, this upper limit is at around 40 components per transient for the LSF implementation employed in this work. Another particular aspect of LSF processing is its requirement of the seed parameters (e.g., approximate frequencies). Therefore, LSF can be most efficiently applied only in combination with other signal processing methods (e.g., FT and FDM).

Conclusions

We described and evaluated the application of the least-squares fitting (LSF) method for analysis of experimental transient signals in FTMS when the transient signal consists of a few sinusoidal components. In comparisons to the typically employed FT processing, the LSF processing yielded superior precision in determination of the frequencies of sinusoidal components of transient signals. For low abundance ions present in an isotopic distribution, the LSF method provided information on their amplitudes with precision not worse than that achieved by FT. Complementary to FT processing, the LSF provides the phases of the sinusoidal components with high precision so that even the "first data-point problem" has been revealed in experimental FT-ICR MS data. Thus, the LSF method is particularly useful in studies addressing the current analytical limits in FTMS and aiming to advance the analytical performance of FTMS via corresponding development of instrumentation

and signal processing methods. Owing to numerical limitations, the application area of the LSF method in FTMS should not be confused with the analysis of transient signals obtained for measurements of broadband mass spectra corresponding to hundreds or thousands of ions with different m/z ratios.

Acknowledgments

The authors express their gratitude to Thermo Fisher Scientific Inc. for providing access under license to Orbitrap transient signals. They are grateful to Alexander Makarov for critical comments on the manuscript. They appreciate the financial support through the Joint Russia–Switzerland Research Program (grant agreement 128357), and the European Research Council (ERC Starting grant 280271 to YOT).

References

1. Marshall, A.G., Hendrickson, C.L.: High-resolution mass spectrometers. *Annu. Rev. Anal. Chem.* **1**, 579–599 (2008)
2. Scigelova, M., Hornshaw, M., Giannakopoulos, A., Makarov, A.: Fourier transform mass spectrometry. *Mol. Cell. Proteom.* **10**, M111.009431 (2011)
3. Xian, F., Hendrickson, C.L., Marshall, A.G.: High resolution mass spectrometry. *Anal. Chem.* **84**, 708–719 (2012)
4. Zubarev, R.A., Makarov, A.: Orbitrap mass spectrometry. *Anal. Chem.* **85**, 5288–5296 (2013)
5. Comisarow, M.B., Marshall, A.G.: Selective-phase ion-cyclotron resonance spectroscopy. *Can. J. Chem.* **52**, 1997–1999 (1974)
6. Craig, E.C., Santos, I., Marshall, A.G., Nibbering, N.M.M.: Dispersion versus absorption (DISPA) method for automatic phasing of Fourier transform ion cyclotron resonance mass spectra. *Rapid Commun. Mass Spectrom.* **1**, 33–37 (1987)
7. Vining, B.A., Bossio, R.E., Marshall, A.G.: Phase correction for collision model analysis and enhanced resolving power of Fourier transform ion cyclotron resonance mass spectra. *Anal. Chem.* **71**, 460–467 (1999)
8. Beu, S.C., Blakney, G.T., Quinn, J.P., Hendrickson, C.L., Marshall, A.G.: Broadband phase correction of FT-ICR mass spectra via simultaneous excitation and detection. *Anal. Chem.* **76**, 5756–5761 (2004)
9. Xian, F., Hendrickson, C.L., Blakney, G.T., Beu, S.C., Marshall, A.G.: Automated Broadband Phase Correction of Fourier Transform Ion Cyclotron Resonance Mass Spectra. *Anal. Chem.* **82**, 8807–8812 (2010)
10. Qi, Y.L., Barrow, M.P., Li, H.L., Meier, J.E., Van Orden, S.L., Thompson, C.J., O'Connor, P.B.: Absorption-mode: the next generation of fourier transform mass spectra. *Anal. Chem.* **84**, 2923–2929 (2012)
11. Xian, F., Corilo, Y.E., Hendrickson, C.L., Marshall, A.G.: Baseline correction of absorption-mode Fourier transform ion cyclotron resonance mass spectra. *Int. J. Mass Spectrom.* **325**, 67–72 (2012)
12. Kilgour, D.P.A., Wills, R., Qi, Y.L., O'Connor, P.B.: Autophaser: an algorithm for automated generation of absorption mode spectra for FT-ICR MS. *Anal. Chem.* **85**, 3903–3911 (2013)
13. Lange, O.: Methods and apparatus for producing a mass spectrum. US Patent 2011/0240841 A1 (2011)
14. Lange, O., Damoc, E., Wieghaus, A., Makarov, A.: Enhanced Fourier transform for Orbitrap mass spectrometry. *Proceeding of the 59th ASMS Conference on Mass Spectrometry and Allied Topics*, Denver, CO, June 5–9 (2011)
15. Hilger, R.T., Wyss, P.J., Santini, R.E., McLuckey, S.A.: Absorption mode Fourier transform electrostatic linear ion trap mass spectrometry. *Anal. Chem.* **85**, 8075–8079 (2013)
16. Kozhinov, A.N., Tsybin, Y.O.: Filter diagonalization method-based mass spectrometry for molecular and macromolecular structure analysis. *Anal. Chem.* **84**, 2850–2856 (2012)

17. Kozhinov, A.N., Aushev, T., Tsybin, Y.O.: Advanced signal processing methods for FTMS: implementation and characterization. Proceedings of the 61st ASMS Conference on Mass Spectrometry and Allied Topics, Minneapolis, MN, (2013)
18. Aushev, T., Adachi, I., Arinstein, K.: Study of the decays $B \rightarrow Ds1(2536)^+(D)\overline{\nu}$. *Phys. Rev. D* **83**, 051102-1–051102-6 (2011)
19. Nikolaev, E.N., Gorshkov, M.V.: Dynamics of ion motion in an elongated cylindrical cell of an ICR spectrometer and the shape of the signal registered. *Int. J. Mass Spectrom. Ion Process* **64**, 115–125 (1985)
20. Grosshans, P.B., Shields, P.J., Marshall, A.G.: Comprehensive theory of the Fourier transform ion cyclotron resonance signal for all ion trap geometries. *J. Chem. Phys.* **94**, 5341–5352 (1991)
21. Miladinović, S.M., Kozhinov, A.N., Tsybin, O.Y., Tsybin, Y.O.: Sidebands in Fourier transform ion cyclotron resonance mass spectra. *Int. J. Mass Spectrom.* **325/327**, 10–18 (2012)
22. Grothe, R.A.: Estimation of ion cyclotron resonance parameters in Fourier transform mass spectrometry. US Patent 8431886, B2 (2013)
23. Grothe, R.A.: Mass spectrometry systems. US Patent 2013/0013274 A1 (2013)
24. Nikolaev, E.N., Heeren, R.M.A., Popov, A.M., Pozdnev, A.V., Chingin, K.S.: Realistic modeling of ion cloud motion in a Fourier transform ion cyclotron resonance cell by use of a particle-in-cell approach. *Rapid Commun. Mass Spectrom.* **21**, 3527–3546 (2007)
25. Nikolaev, E., Boldin, I., Jertz, R., Baykut, G.: Initial experimental characterization of a new ultra-high resolution FTICR cell with dynamic harmonization. *J. Am. Soc. Mass Spectrom.* **22**, 1125–1133 (2011)
26. Zhurov, K.O., Kozhinov, A.N., Tsybin, Y.O.: Evaluation of high-field Orbitrap Fourier transform mass spectrometer for petroleomics. *Energy Fuels* **27**, 2974–2983 (2013)
27. Kozhinov, A.N., Zhurov, K.O., Tsybin, Y.O.: Iterative method for mass spectra recalibration via empirical estimation of the mass calibration function for Fourier transform mass spectrometry-based petroleomics. *Anal. Chem.* **85**, 6437–6445 (2013)
28. Kozhinov, A.N., Miladinovic, S.M., Tsybin, Y.O.: Spectral errors in Fourier transform mass spectrometry revisited. Proceedings of the 59th ASMS Conference on Mass Spectrometry and Allied Topics, Denver, CO, (2011)
29. James, F., Roos, M.: Minuit—a system for function minimization and analysis of the parameter errors and correlations. *Comput. Phys. Commun.* **10**, 343–367 (1975)
30. Tolmachev, A.V., Masselon, C.D., Anderson, G.A., Udseth, H.R., Smith, R.D.: Frequency shifts due to the interference of resolved peaks in magnitude-mode fourier-transform ion cyclotron resonance mass spectra. *J. Am. Soc. Mass Spectrom.* **13**, 387–401 (2002)
31. Easterling, M.L., Amster, I.J., van Rooij, G.J., Heeren, R.M.A.: Isotope beating effects in the analysis of polymer distributions by Fourier transform mass spectrometry. *J. Am. Soc. Mass Spectrom.* **10**, 1074–1082 (1999)
32. Aizikov, K., O'Connor, P.B.: Use of the filter diagonalization method in the study of space charge related frequency modulation in Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectrom.* **17**, 836–843 (2006)