

Fourier transform mass spectrometry at the uncertainty principle limit for improved qualitative and quantitative molecular analyses

THÈSE N° 6384 (2015)

PRÉSENTÉE LE 20 FÉVRIER 2015

À LA FACULTÉ DES SCIENCES DE BASE
LABORATOIRE DE SPECTROMÉTRIE DE MASSE DE BIOMOLÉCULES
PROGRAMME DOCTORAL EN CHIMIE ET GÉNIE CHIMIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Anton KOZHINOV

acceptée sur proposition du jury:

Dr M. Drabbels, président du jury
Prof. Y. Tsybin, directeur de thèse
Prof. R. M. Heeren, rapporteur
Prof. A. A. Makarov, rapporteur
Prof. J. Vanicek, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2015

Acknowledgements

The present thesis would not have come into being without the contribution, support, and help of many people. First, I would like to express my gratitude to my advisor, Prof. Yury O. Tsybin, whose research laboratory I was so fortunate to join four years ago. I would also like to thank Christine Kupper for taking full control of many administrative processes in the lab. I highly appreciate the contribution of the current and former lab members to this thesis, the people with such outstanding personalities: Aleksey Vorobyev with his never-ending source of fun; Saša Miladinović, who really had his ways of handling the ICRs including the older one; Krzysztof Piech and his awesome idea on the lab's awesomeness; Ŭni Laskay with her positive attitude; Luca Fornelli, both a doctor of philosophy and a philosopher; Daniel Ayoub, who really helps recall a good, twist-genre song on a stylish orange tie (which is in Russian though); Kostya Zhurov, who is always ready to help, especially if you ask him to God-bless someone right at the moment; Kostya Nagornov, a welcoming host at the Nagornovs'; and Kristina Srzentić with her hilarity being always on. My appreciation is also extended to the excellent services and their staff here at the EPFL, including Patrick Favre, Harald Holze, Steve Couturier, and André Fattet of the electronic and mechanical workshops, as well as the EPFL library and the BCH magasin. Many people, including Prof. Oleg Yu. Tsybin of SPbSPU, Tagir Aushev of IThEP, Marion Girod and Jérôme Lemoine of Université Lyon 1, Marco Seynen, Marco Konijnenburg, and Don Smith of AMOLF, Nicolas Pasquier of NI, Prof. Steven Gygi and Martin Wuehr of Harvard University, Prof. Joshua Coon of University of Wisconsin, Richard Knochenmuss of Tofwerk, Steven Beu of S C Beu Consulting, Konstantin Aizikov of Thermo Scientific are gratefully acknowledged for inspiring discussions, their expertise, technical support, provided samples, and experimental assistance. I am highly grateful to my committee members Prof. Jiri Vanicek, Prof. Ron M.A. Heeren, Prof. Alexander A. Makarov, and Dr. MER Marcel Drabbels for their valuable time and work. I'd also like to emphasize my special thanks to several people from my earlier years, viz. Dr. Mikhail V. Gorshkov, who opened the world of FTMS for me; Prof. Alexander K. Chibisov, my MSc project advisor; and my high school teachers Elena L. Aivazyan and Lidia M. Ivanova. The financial support from the Scientific and Technological Cooperation Programme Switzerland-Russia (grant agreement 128357), the Swiss National Science Foundation (SNSF project 200021-125147/1), and the European Research Council (ERC Starting grant 280271) is gratefully appreciated.

Lausanne, 18 September 2014

A. K.

Abstract

Nowadays, among the instrumentation park of mass spectrometry, Fourier transform mass spectrometers (FTMS), including ion cyclotron resonance (ICR) and Orbitrap FTMS, provide the highest analytical performance for accurate measurements and mass resolution. Nevertheless, molecular analysis in currently challenging research areas, such as life and environmental sciences, necessitates further improvement of these analytical characteristics. In recent decades, a regular approach to address the problem behind the analytical performance was using increased electromagnetic fields. However, per a given increase of their magnitudes, that approach is currently requiring more and more resources, hence demonstrating its limited feasibility for tomorrow. Therefore, only a qualitative breakthrough in the underlying methodology may then lead to the next series of developments enabling improved molecular analysis. The present research is dedicated to the fundamental question behind the analytical performance in FTMS. Specifically, this thesis represents an interdisciplinary study aimed at improved molecular analysis in FTMS-based applications, achieved via better comprehension of the uncertainty principle in FTMS, viz. its dependence on the measurement scheme, including ion traps, signal processing, and data analysis, as well as its influence on achievable analytical characteristics in FTMS. To start, the uncertainty principle for measurements in FTMS has been investigated, asserting a limit to the precision with which complementary physical quantities in FTMS, e.g. detection period and scale of frequency details to resolve, can be measured. Specifically, two corollaries of the uncertainty principle are considered, viz. resolution performance and performance for accurate measurements in FTMS. Importantly, the uncertainty principle shows dependence on the particular measurement scheme employed. For instance, signal detection, signal processing, and data analysis of the standard measurement scheme impose their own restrictions to the resulting uncertainty principle, thus leading to the current limitations. However, the ultimate limitations in the analytical characteristics in question are defined by the uncertainty principle limit due to physics of ion motion in the mass analyzer. Hence, with corresponding developments in data analysis methods, methods for signal processing, and designs of ion traps, novel measurement schemes have been implemented, where the quantitative form of the uncertainty principle is modified towards the ultimate limit. Finally, the implemented solutions have been evaluated in the context of FTMS-based analysis of crude oil fractions, protein identification and characterization, quantitative proteomics, and analysis of isotopic fine structures of peptides. To conclude, the achieved success of this work should considerably contribute to the currently challenging analytical applications of FTMS.

Abstract

Key words:

mass spectrometry, MS; Fourier transform mass spectrometry, FTMS; ion cyclotron resonance, ICR; orbitrap; uncertainty principle; signal processing; ion trap; super-resolution methods, SR; unperturbed cyclotron frequency; transient signal

Résumé

Depuis les expériences fondamentales menées par J. J. Thomson sur les rayons cathodiques et anodiques au début du XX^{ème} siècle, la spectrométrie de masse s'est étendue pour devenir une technique d'analyse indispensable à différentes applications, de la physique atomique aux sciences de la vie. Un certain nombre de découvertes importantes ont permis de révolutionner l'instrumentation de la spectrométrie de masse au cours des cent dernières années, permettant ainsi l'analyse compréhensive d'échantillons variés. Aujourd'hui, le spectromètre de masse à transformée de Fourier (SMTF), qui inclut la résonance cyclotron ionique (RCI) et le SMTF Orbitrap, permet la meilleure performance analytique parmi la palette d'instruments disponibles pour la spectrométrie de masse. Cependant, l'analyse moléculaire dans certains domaines de recherche difficiles, comme les sciences de la vie et la durabilité énergétique, nécessite une amélioration de la performance analytique des SMTF. En général, le problème de fond de la performance analytique peut être traité en utilisant deux approches complémentaires incluant des champs électromagnétiques plus intenses et une meilleure compréhension du principe d'incertitude des SMTF. Cette recherche étudie spécifiquement la dernière approche. Pour commencer, le principe d'incertitude pour des mesures faites avec un SMTF a été développé, et montre un lien étroit avec des caractéristiques analytiques importantes comme la performance de résolution et la précision de mesure. Le principe d'incertitude dépend surtout du schéma général de mesure employé, incluant le mouvement des ions, la détection et le traitement du signal, ainsi que l'analyse des données. Ensuite, en considérant les développements correspondants dans les méthodes d'analyse de données, les méthodes de traitement du signal, et les modèles de pièges ioniques, différents modèles de mesure ont été implémentés, dans lesquels la forme quantitative du principe d'incertitude est modifiée afin d'obtenir la performance analytique requise. Finalement, le potentiel des résultats obtenus a été évalué dans le contexte d'analyses moléculaires qualitatives et quantitatives dans les domaines de la protéomique et de la pétroléomique.

Mots clefs :

la spectrométrie de masse, SM ; la spectrométrie de masse à transformée de Fourier, SMTF ; la spectrométrie de masse à résonance cyclotron ionique à transformée de Fourier, SM RCI ; le principe d'incertitude ; le traitement du signal ; la piège ionique ; la protéomique ; la pétroléomique.

List of Papers

The present thesis is based on the following selected papers, which will be referred to in the text by their Roman numerals. This research was also presented in parts in conference presentations, a selected list of which is given below. Two provisional patent applications have been submitted. Finally, papers with related work omitted from this thesis are listed separately.

Selected papers

- I. Miladinović S.M.*, **Kozhinov A.N.***, Tsybin O.Y., Tsybin Y.O. Sidebands in Fourier transform ion cyclotron resonance mass spectra. *Int. Journ. Mass Spectrom.*, **2012**, 325–327, 10–18, doi: 10.1016/j.ijms.2012.08.009. **these authors contributed equally.*
- II. Zhurov K.O., **Kozhinov A.N.**, Fornelli L., Tsybin Y.O. Distinguishing analyte from noise components in mass spectra of complex samples: Where to cut the noise? (*Appendix*) *Anal. Chem.*, **2014**, 86, 3308–3316, doi: 10.1021/ac403278t.
- III. **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.*, **2013**, 85, 6437–6445, doi: 10.1021/ac400972y.
- IV. Aushev T., **Kozhinov A.N.**, Tsybin Y.O. Least-squares fitting of time-domain signals for Fourier transform mass spectrometry. *Journ. Amer. Soc. Mass Spectrom.*, **2014**, 25, 1263–1273, doi: 10.1007/s13361-014-0888-x.
- V. **Kozhinov A.N.**, Tsybin Y.O. Filter diagonalization method-based mass spectrometry for molecular and macromolecular structure analysis. *Anal. Chem.*, **2012**, 84, 2850–2856, doi: 10.1021/ac203391z.
- VI. Miladinović S.M., **Kozhinov A.N.**, Gorshkov M.V., Tsybin Y.O. On the utility of isotopic fine structure mass spectrometry in protein identification. *Anal. Chem.*, **2012**, 84, 4042–4051, doi: 10.1021/ac2034584.
- VII. Nagornov K.O., Gorshkov M.V., **Kozhinov A.N.**, Tsybin Y.O. High-resolution Fourier transform ion cyclotron resonance mass spectrometry with increased throughput for biomolecular analysis. *Anal. Chem.*, **2014**, 86, 9020–9028, doi: 10.1021/ac501579h.

List of Papers

- VIII. Nagornov K.O., **Kozhinov A.N.**, Tsybin Y.O. Paradigm shift in ion trap design and capabilities for Fourier transform ion cyclotron resonance mass spectrometry. **2014**, *submitted for publication*.

Conference contribution

Selected oral presentations:

- C-I. **Kozhinov A.N.**, Nagornov K.O., Ayoub D., Tsybin Y.O. Unexplored reserves of resolution in Fourier transform mass spectrometry. *Proceedings of the 62nd ASMS Conference on Mass Spectrometry and Allied Topics*. Baltimore, MD, USA, June 15–19, **2014**.
- C-II. **Kozhinov A.N.** High-speed data acquisition system with LabVIEW PXIe. *NIDays 2014*. Bern, Switzerland, May 5, **2014**.
- C-III. Tsybin Y.O., Laskay Ü.A., Fornelli L., Zhurov K.O., **Kozhinov A.N.** Enhanced Fourier transform and filter diagonalization method mass spectrometry for top-down analysis of antibodies and petroleomics. *Proceedings of the 19th International Mass Spectrometry Conference*. Kyoto, Japan, September 15–21, **2012**.
- C-IV. **Kozhinov A.N.**, Laskay Ü.A., Zhurov K.O., Fornelli L., Tsybin Y.O. Filter diagonalization method-based mass spectrometry. *Proceedings of the SCS Fall Meeting 2012*. Zürich, Switzerland, September 13, **2012**.

Selected poster presentations:

- C-V. **Kozhinov A.N.**, Nagornov K.O., Zhurov K.O., Tsybin Y.O. New methods of data analysis for FTMS with improved analytical performance. *Proceedings of the 20th International Mass Spectrometry Conference*. Geneva, Switzerland, August 24–29, **2014**.
- C-VI. **Kozhinov A.N.**, Aushev T., Tsybin Y.O. Signal processing and instrumentation for Fourier transform mass spectrometry with improved analytical performance. *Proceedings of the SCS Fall Meeting 2013*. Lausanne, Switzerland, September 6, **2013**.
- C-VII. **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, USA, June 9–13, **2013**.
- C-VIII. **Kozhinov A.N.**, Miladinović S.M., Tsybin O.Y., Tsybin Y.O. Harmonics and sidebands in Fourier transform ion cyclotron resonance mass spectra. *Proceedings of the 6th Summer Course for Mass Spectrometry in Biotechnology and Medicine*. Dubrovnik, Croatia, July 8–14, **2012**.
- C-IX. **Kozhinov A.N.**, Laskay Ü.A., Fornelli L., Zhurov K.O., Tsybin Y.O. A Python library for advanced signal processing in Fourier transform mass spectrometry. *Proceedings of the 60th ASMS Conference on Mass Spectrometry and Allied Topics*. Vancouver, BC, Canada, May 20–24, **2012**.

- C-X. **Kozhinov A.N.**, Miladinović S.M., Gorshkov M.V., Tsybin Y.O. The role of isotopic fine structures in protein identification by high resolution mass spectrometry. *Proceedings of the SCS Fall Meeting 2011*. Lausanne, Switzerland, September 9, **2011**.
- C-XI. **Kozhinov A.N.**, Miladinović 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, USA, June 5–9, **2011**.

Patents

- P-I. Tsybin Y.O., **Kozhinov A.N.**, Nagornov, K.O. A device and method for ion cyclotron resonance mass spectrometry. Patent application PCT/IB2014/060709.
- P-II. **Kozhinov A.N.**, Tsybin Y.O. An electrostatic ion trap for Fourier transform mass spectrometry. Patent application PCT/IB2014/064793.

Related papers not included in the thesis

- R-I. Zhurov K.O., **Kozhinov A.N.**, Tsybin Y.O. Evaluation of high-field orbitrap Fourier transform mass spectrometer for petroleomics. *Energy & Fuels*, **2013**, 27, 2974–2983, doi: 10.1021/ef400203g.
- R-II. Zhurov K.O., **Kozhinov A.N.**, Tsybin Y.O. Hexagonal class representation for fingerprinting and facile comparison of petroleomic samples. *Anal. Chem.*, **2013**, 85, 5311–5315, doi: 10.1021/ac400418j.
- R-III. Laskay Ū.A., Srzentić K., Fornelli L., Upir O., **Kozhinov A.N.**, Monod M., Tsybin Y.O. Practical considerations for improving the productivity of mass spectrometry-based proteomics. *Chimia*, **2013**, 67, 244–249, doi: 10.2533/chimia.2013.244.
- R-IV. Tsybin Y.O., Fornelli L., **Kozhinov A.N.**, Vorobyev A., Miladinović S.M. High-resolution and tandem mass spectrometry — the indispensable tools of the XXI century. *Chimia*, **2011**, 65, 641–645, doi: 10.2533/chimia.2011.641.
- R-V. Tsybin O.Y., Nagornov K.O., **Kozhinov A.N.**, Tsybin Y.O. Transient calculation algorithms in Fourier transform ion cyclotron resonance mass spectrometry. **2014**, submitted for publication.

List of Abbreviations and Notations

Abbreviations

AWG	arbitrary waveform generator
CPU	central processing unit
DAQ	data acquisition [system]
DFT	discrete Fourier transform
DSP	digital signal processing
eFT	enhanced FT [algorithm]
ESI	electrospray ionization
FIFO	first in-first out [data buffer]
FDM	filter-diagonalization method
FFT	fast Fourier transform
FT	Fourier transform
FT-ICR MS	Fourier transform ion cyclotron resonance mass spectrometry
FTMS	Fourier transform mass spectrometry
FWHM	full width at half maximum
ICR	ion cyclotron resonance
LC	liquid chromatography
LC-MS	liquid chromatography coupled to mass spectrometry
LSF	least-squares fitting
LTQ	linear trap quadrupole
MALDI	matrix-assisted laser desorption/ionization
MS	mass spectrometry
NADEL	narrow aperture detection electrodes [cell]
PDF	probability density function
PCI	peripheral component interconnect
PXI	PCI extensions for instrumentation
PXIe	PXI Express
RAM	random-access memory
RF	radio-frequency [signal]
SNR	signal-to-noise ratio
SR	super-resolution [method for signal processing]
ZF	zero-filling

List of Abbreviations and Notations

Notations

$\Re[z]$ and $\Im[z]$	real and imaginary parts of a complex number z
A	ion abundance
j	unit imaginary number, $\sqrt{-1}$
m/z	mass-to-charge ratio
o	little- o notation in Taylor expansions
\mathcal{Z}	the Zahlen symbol denoting the set of all integer numbers
$\lfloor x \rfloor$	the floor function of a real number x
$\lceil x \rceil$	the ceiling function of a real number x

Contents

Title page	i
Acknowledgements	iii
Abstract	v
Résumé	vii
List of Papers	ix
List of Abbreviations and Notations	xiii
Contents	xv
Introduction	1
1 Background	5
1.1 Subject of mass spectrometry	5
1.1.1 Substance as a form of matter	5
1.1.2 Chemical elements, atoms, and isotopes	6
1.1.3 Molecular formulas, molecules and isotopologues	7
1.2 Analytical questions in mass spectrometry	9
1.3 Mass spectrometry instrumentation	9
1.3.1 Brief review of history and milestones	9
1.3.2 ICR and Orbitrap FTMS	11
1.3.3 Measurement setup	13
1.4 Uncertainty principle	13
2 Preliminary developments	15
2.1 Measurement theory	15
2.1.1 Phase portrait of ion motion	15
2.1.2 Two sides to the uncertainty principle limit	19
2.1.3 Transient signal as a response function	20
2.1.4 Limitations of FT signal processing	23
2.2 High-performance data acquisition system	25
2.2.1 PXI Express hardware platform	27

xv

Contents

2.2.2	Data streaming software	29
2.3	pyFTMS software framework	30
3	Data analysis	33
3.1	Signal processing with FT	33
3.1.1	Spectral discreteness	35
3.1.2	Non-linear spectral interference	38
3.1.3	Correlated errors	41
3.2	Noise influence	43
3.2.1	Logarithmic distributions	44
3.2.2	Effect of digital noise	48
3.3	Mass calibration function	49
3.3.1	Recalibration method	51
3.3.2	Performance evaluation	51
4	Signal processing	55
4.1	FT and phase intersection points	56
4.1.1	No information on ion phases	56
4.1.2	Single phase intersection point	57
4.1.3	Multiple phase intersection points	59
4.1.4	Resolution performance comparison	61
4.2	Least-squares fitting	64
4.2.1	Method implementation	65
4.2.2	Performance evaluation	66
4.2.3	Implications for development of FTMS	72
4.2.4	Advantages and limitations	73
4.2.5	LSF-based workflow for quantitative proteomics	73
4.3	Filter-diagonalization method	76
4.3.1	Methodology	76
4.3.2	Super-resolution mode in FTMS	79
4.3.3	SNR limited resolution	80
4.3.4	Spectral irregularity	80
5	Instrumentation	83
5.1	4X cell	83
5.1.1	Design of the 4X cell	84
5.1.2	Resolution performance	85
5.1.3	Evaluation for applications	87
5.2	NADEL cell	88
5.2.1	Design of the NADEL cell	89
5.2.2	Conventional regime	91
5.2.3	Increased resolution performance	92
5.2.4	Accurate mass measurements regime	94

Conclusions	99
A Papers	103
A.1 Paper I	103
A.2 Paper II	113
A.3 Paper III	125
A.4 Paper IV	135
A.5 Paper V	155
A.6 Paper VI	169
A.7 Paper VII	188
A.8 Paper VIII	205
Bibliography	264
Curriculum Vitae	265

Introduction

Having started a little longer than a hundred years ago, numerous developments in the physics of yesteryear have led to a new era in physical science. With this progress, mankind's knowledge has been significantly revised and further developed for various physical phenomena involving substance and field, time and space, life and the Universe. This revolution has exceedingly contributed into various contemporary scientific fields, ranging from engineering and fundamental physics to life and environmental sciences. Particularly, along with other key developments, the first discovery of a subatomic particle, the electron, made in the late XIX century by Sir J. J. Thomson[1] spurred a flurry of research at the subatomic and supra-atomic levels, giving origin to the new physical disciplines of atomic physics and molecular chemistry, respectively, as well as many derivatives of them formed and developed to date. Notably, shortly following this discovery, isotopes of several chemical elements were revealed to exist[2, 3] and the measurement technique employed was recognized as a powerful means of analyzing chemical compounds and was subsequently developed into a general method of mass spectrometry.[4, 5]

Naturally, the focus of mass spectrometry has gradually been shifting from fundamental questions involving chemical elements and their isotopes to various analytical problems ranging from regular refinement of isotopic masses and relative abundances to molecular structure analysis and quantitation. Continuous developments in its instrumentation including various ion sources, techniques for ion fragmentation, and mass analyzers, have allowed mass spectrometry of today to perform a comprehensive analysis of virtually any chemical compound even when given at extremely low levels of concentration. Currently, mass spectrometry is a key analytical technique in obtaining the required molecular information from minute amounts of samples present in the gas, solid or liquid phase. It allows one to identify the structure of molecules from samples of interest and quantify the amounts of molecules of different types. This information is deduced from accurately measured mass-to-charge ratios and abundances of ionized molecules and their fragments. In this regard, mass spectrometry is so fundamental to substance analysis that when synthesizing a new chemical compound, providing its mass spectrum along with characterization by other relevant methods is considered a must.

The instrumentation park of mass spectrometry includes a number of instruments differing by their levels of analytical performance. Nowadays, Fourier transform mass spectrometers

Introduction

(FTMS), including ion cyclotron resonance (ICR), which was first developed back in the 1970s by Alan G. Marshall and Melvin B. Comisarow,[6, 7] and Orbitrap FTMS introduced in the early 2000s by Alexander Makarov,[8] is a group of instruments providing the highest analytical characteristics such as resolving power and mass measurement accuracy, especially those instruments that implement a hybrid architecture comprising a variant of radio-frequency ion trap along with the ICR cell or the orbitrap, respectively.

Although FTMS has been significantly revolutionized in the past, new developments in its methodology are vital for further progress in FTMS-based applications. Indeed, qualitative and quantitative molecular analyses in currently challenging research areas necessitate improved analytical performance. To name a few, consider the sustainability of the energy and fuels production, where processing and transport are extremely molecular content-dependent. The shift in fossil fuel production from the light crude oils toward the heavy ones has already revealed an important environmental and economical problem of accurate specification and quantitation of heteroatoms, e.g. sulfur. Likewise, the field of proteomics is currently facing extreme challenges in the analysis of complex biological samples. However, the required level of analytical performance of FTMS is not readily provided today. Thus, molecular structure analysis and molecular quantitation in proteomics and petroleomics present a substantial challenge for state-of-the-art MS, indicating the need for further development in mass spectrometry in order to address the growing analytical interest in complex materials of biological and chemical origin and their interaction pathways.

The aim of research

Generally, among the most important unsolved questions in state-of-the-art FTMS, the problem behind the limited analytical performance is related to the uncertainty principle in FTMS. This problem can be subdivided into the following two questions. The first unsolved question includes detrimental effects of motion of ion packets in the mass analyzer, colloquially known as ion packet decoherence (dephasing) and coalescence, whose nature has not yet been well-understood due to lack of a comprehensive theory accounting for static and time-dependent perturbations of the trapping electromagnetic fields. For instance, these effects of ion motion cause limited accuracy of mass measurements. Indeed, it has already long been around that mass accuracy achievable with the standard measurement approach cannot breakthrough the level of 1 ppm or, to say exactly, an order of it. The second unsolved question is a better realizing that the uncertainty principle in FTMS does depend on the particular methodology employed. For instance, there exists a stereotype that it is the resolution performance of Fourier transform that imposes a fundamental limit to the achievable resolution in FTMS. Contrastingly, whilst Fourier transform indeed defines the apparent resolution in mass spectra, the fundamental limit is rather imposed by the finite size of ion packets, as governed by the decoherence effect mentioned above.

From the viewpoint of the recent years of FTMS, an apparent approach to improve the limitations mentioned above is using stronger electromagnetic fields. However, given that a unit increase of their magnitudes is currently requiring more and more resources, this option is slowly approaching the saturation level as of today.

In contrast, the present thesis is a complementary way towards advances in FTMS-based molecular analysis via improved analytical characteristics. Specifically, the present research addresses the problem of analytical performance via a qualitative transition in the very paradigm of FTMS, rather than implementing a quantitative way with increased electromagnetic fields. To state in a single sentence, the aim of the present thesis has been to develop the question on the uncertainty principle in FTMS followed by implementation of several approaches for improved analytical performance, with a specific focus made to accuracies for measurements of ion masses and abundances, as well as resolution performance for distinguishing ion species with close mass-to-charge ratios, in order to enable comprehensive molecular structure analysis and molecular quantitation in the currently challenging FTMS-based applications, with the particular areas of interest being proteomics and petroleomics.

Thesis outline

This thesis contains five enumerated chapters and one appendix. Chapter 1, *Background*, reviews mass spectrometry as a field of science by means of considering its aspects from different points of view, including its subject, analytical questions, and apparatus.

Chapter 2, *Preliminary developments*, creates the basis for achieving the aim of this research. The concept of phase portraits for the dynamic system of ions trapped in a mass analyzer is introduced. Then, the theory for the ICR transient signal is revisited. Next, the uncertainty principle in FTMS is developed and shown to depend on the measurement scheme including ion motion, signal detection (generation), signal processing, and data analysis. Importantly, here we show the uncertainty principle to incorporate the two questions on measurement accuracy and resolution performance. Additionally, a limiting case is considered when the uncertainty principle is governed by limitations in confinement of ions in the mass analyzer. The chapter concludes by describing software framework and data acquisition system as the tools specifically implemented for tackling the research problem.

Based on these preliminary developments, Chapter 3, *Data analysis*, presents the first step towards improved analytical performance of FTMS achieved in the present work. This chapter is meant to prepare for the following research by means of separate consideration of the data analysis part of the total measurement scheme. This chapter describes the data analysis methods aimed at increased measurement accuracy of ion mass-to-charge ratios and abundances, as well as dynamic range of ion abundances. As a particular analytical case, a mass calibration method is discussed, enabling comprehensive analysis of petroleum crude oil fractions, at least of moderate complexity, with Orbitrap FTMS.

Introduction

In Chapter 4, *Signal processing*, the standard measurement scheme is modified in accordance with the considerations developed in Chapter 2. Specifically, this chapter is focused on the signal processing part of the total measurement scheme. Here, instead of the standard FT, other methods for signal processing were implemented, including an FT-based method with double phase correction, least-squares fitting (LSF), and filter-diagonalization method (FDM), all aimed at increased performance for resolution. The application areas considered here include protein structure analysis in proteomics, peptide identification using their isotopic fine structures, and quantitative proteomics. Importantly, apart from pure applications as a signal processing part of the measurement scheme, the methods presented in this chapter are also intended for integrated approaches involving hardware aspects discussed in the following chapter.

Chapter 5, *Instrumentation*, considers achieving improved analytical performance on the hardware level. Here, in accordance with the developments in Chapter 2, we integrate the developments described separately in the previous chapters and complement them with new designs of two ICR cells, referred to as 4X cell and NADEL cell. The 4X cell provides four-fold gain in resolution performance when employed instead of the standard ICR cell. In turn, the NADEL cell installed on the standard 10 T FT-ICR MS provides three distinct regimes, including the regime of a standard ICR cell, a regime for increased resolution performance, and a regime for increased mass measurement accuracy. For the second regime, allied signal processing aimed to convert the increased spatial resolution of analyzed ion packets into increased frequency resolution of the ion packets is presented. Additionally, advantage of the third regime is demonstrated in analysis of petroleomic samples.

Finally, in *Conclusions* the fulfilled goals are summarized, their advantages for molecular analyses are discussed, and the future perspectives in FTMS are outlined. Appendix A, *Papers*, contains papers included in this thesis.

1 Background

This chapter aims to reveal the essence of mass spectrometry by means of showing what properties it manifests in the following different contexts. In Section 1.1, the subject of mass spectrometry is discussed. The fundamental concept of matter is briefly reviewed to determine the scope of mass spectrometry in the physical world. Atoms and molecules are considered as high-order systems of one of the forms of matter, with characterizing the mass and abundance properties of isotopes and isotopologues being the very species detected in a mass analyzer. Then, Section 1.2 summarizes the analytical questions addressing by mass spectrometry, hence defining the *raison d'être* for this field of science. Direct and inverse analytical problems are distinguished, leading to quantitative and qualitative kinds of analysis performed in mass spectrometry. Finally, the instrumentation aspects of mass spectrometry are covered in Section 1.3. History and milestones of the instrument development are briefly reviewed. The instrumentation park of Fourier transform mass spectrometry, including Fourier transform ion cyclotron resonance MS and Orbitrap FTMS, is introduced. This is followed by the description of the standard approaches for signal detection, as well as analog and digital signal processing in FTMS.

1.1 Subject of mass spectrometry

1.1.1 Substance as a form of matter

The subject of physics is *matter*. As such, the subject of mass spectrometry as a branch of physical sciences should be matter as well, possibly with an option to be a subset of it. Matter is one of the two most fundamental concepts in physics, with the reference frame, such as space-time continuum or space of quantum states, being the other. Typically, the two concepts allow a clear distinction, with the former being considered the objective physical reality and the latter serving as a convenient mathematical abstraction.

With numerous developments in physics, the concept of matter has been refined many times. Historically, there used to be a fundamental distinction between matter and field. The modern

Chapter 1. Background

physics eliminates the key difference between matter and field that was considered in the past, thus making for improved understanding of the nature via a more coherent scientific framework. As such, with current developments in physics, it is rather appropriate to think of matter as excitations of various quantum fields of fundamental nature. The fundamental quantum fields are represented by a number of elementary and composite particles, which can be classified by a set of quantum numbers defining their state.

All known elementary particles are either bosons or fermions. The elementary bosons can occupy the same quantum state, while the elementary fermions cannot. With these properties, elementary bosons are often carriers of fundamental interactions such as electromagnetic, strong, weak, and gravitational fields, e.g. the photon as a carrier for electromagnetism. In turn, elementary fermions are considered as the constituents of *substance*, e.g. the electron as an elementary particle and the nucleons as composite particles. Indeed, the "bulky" property implied for substance can be thought of as the result of different quantum states of the elementary particles composing a given system, whilst this property would not rather be characteristic were these particles at the same quantum state.

When organized into higher-order systems, such as atoms and molecules (since these are the kind of substance prevailing on Earth and in the Solar system), substance as a form of matter represents the subject of mass spectrometry.

1.1.2 Chemical elements, atoms, and isotopes

It is interesting to notice that the Latin word *elementum* reminds the ordered set of Latin letters *l*, *m*, *n* and *t* followed by the common ending *um*, thus recalling the ancient proverb: "As words consist of letters, so bodies consist of elements". Nowadays, chemical elements are similarly considered as a part of substance. The modern concept of *chemical elements* is deeply due to the classification of substance based on the periodicity of its properties.

The periodic table proposed by Mendeleev, where the columns contain elements of substance grouped by many major physico-chemical properties while the rows represent alike periods, was developed such that the order of the elements naturally suggested by the atomic masses was occasionally ignored in favor of more coherent grouping of properties, effectively listing the elements in order of increasing nuclear charge. With further developments in physics and chemistry, it was realized that the physico-chemical properties of such elements are mainly due to electrons distributed over a set of available quantum states in an *atom*, supporting the classification based on the increasing nuclear charge as it equals to the number of electrons in an uncharged atom. Thus, it is effectively the periodic table that defines the concept of chemical element as a collection of atoms with the same nuclear charge, with the latter being also known as atomic number.

While the atomic number does uniquely identify a chemical element, atoms of a particular chemical element may differ in numbers of neutrons. Such atoms are known as *isotopes*, with

this term literally meaning that these sorts of atoms occupy the same place in the periodic table as their physico-chemical properties are only weakly dependent on interactions at the nuclear level. As such, isotopes of a given chemical element have the same atomic number but different masses. The mass of an isotope is determined mainly by the sum of masses of all the constituent particles such as protons, neutrons, and electrons, whereas the binding energy of the nucleus and the binding energy of electrons provide the first and second order corrections, respectively, commonly known as the mass defect in bound systems. Due to the mass defect, the mass difference of two isotopes of a given chemical element is close to the mass of the neutron times the number difference of neutrons in these isotopes, but also shows little dependence on the difference in binding energies of the two systems.

The theory of nuclear stability distinguishes isotopes that are theoretically stable to all known forms of decay and isotopes that are prone to either of the known forms of decay. From the experimental point of view, an extended group of so-called observationally stable isotopes is often of interest, composed of theoretically stable isotopes, isotopes that are theoretically unstable but have extremely long half-life periods predicted by the theory, and isotopes that are theoretically unstable to some form of decay but that have not been observed to undergo decay experimentally.

In the context of observationally stable isotopes, a chemical element can be characterized with relative occurrence of its isotopes, Figure 1.1. When speaking of natural isotopic abundances, i.e. relative numbers of isotopes as naturally encountered on Earth, for a given chemical element one distinguishes the isotope with minimum allowed number of neutrons, isotopes with extra neutrons if any, and the most abundant isotope among these. Most of the chemical elements have a number of isotopes, e.g. hydrogen (H), carbon (C), nitrogen (N), oxygen (O), sulfur (S), etc., whereas only some of the chemical elements, e.g. phosphorus (P), fluorine (F), sodium (Na), cesium (Cs), gold (Au), have only a single stable isotope and therefore are called monoisotopic chemical elements.[10]

1.1.3 Molecular formulas, molecules and isotopologues

Being systems of atoms, molecules and their systems such as molecular complexes allow generalizing the aspects of isotopic mass defects and isotopic abundances. As a chemical element may represent a number of isotopes, i.e. atoms differing only in their numbers of neutrons, so may a molecular formula represent a number of isotopologues, i.e. molecules differing only in their isotopic composition.

Different isotopologues of a given molecular formula inherits the mass and abundance properties of the isotopes in the following way. Specifically, in mass spectrometry, the mass of an isotopologue is given as a sum of masses of all constituting isotopes. That is, the additive rule is simply applied, while the second order correction due to mass defect is typically ignored as such a difference is normally beyond the analytical performance of mass spectrometers. In turn, relative abundances of isotopologues as systems of isotopes are given by the multinomial

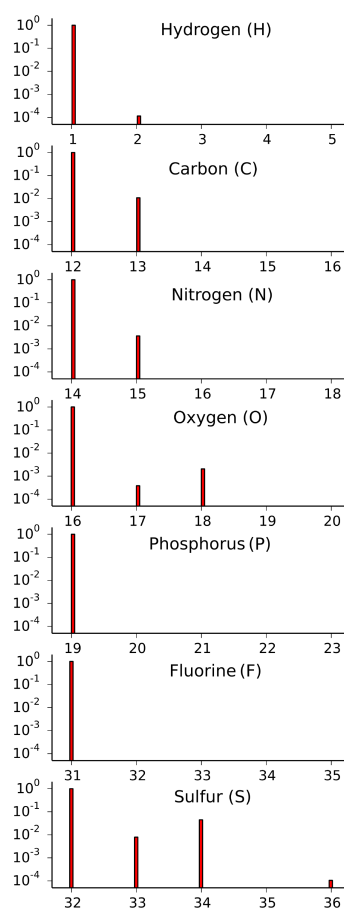


Figure 1.1: Natural relative abundance (vertical axis) vs. exact mass (horizontal axis) plot for isotopes of several chemical elements commonly present in biological molecules. Notice that fluorine and phosphorus are monoisotopic chemical elements. The data is represented based on the NIST table of isotopes.[9]

expansion based on isotopic relative abundances.[11] Finally, the first order correction in mass defect of isotopes results in so-called isotopic fine structures, Figure 1.2, where each fine structure is a set or isotopologues differing only by distribution of a given number of extra neutrons over the atoms of the lightest isotopologue.

In the context of generalization of properties of isotopes to molecules as systems composed of them, it would make perfect sense to define the monoisotopic molecular formula as the one composed only of monoisotopic chemical elements. Nevertheless, by historical definition, the monoisotopic mass or monoisotopic peak in a mass spectrum is that corresponding to an isotopologue composed of the most abundant isotopes of each element present in the molecular formula of interest.[12] Hence, this definition might be seen as only an approximation stating that given that all except the most abundant isotopes are neglected for each of the chemical elements in a given molecular formula, the isotopologue composed of these atoms would give the monoisotopic mass in question.

1.2. Analytical questions in mass spectrometry

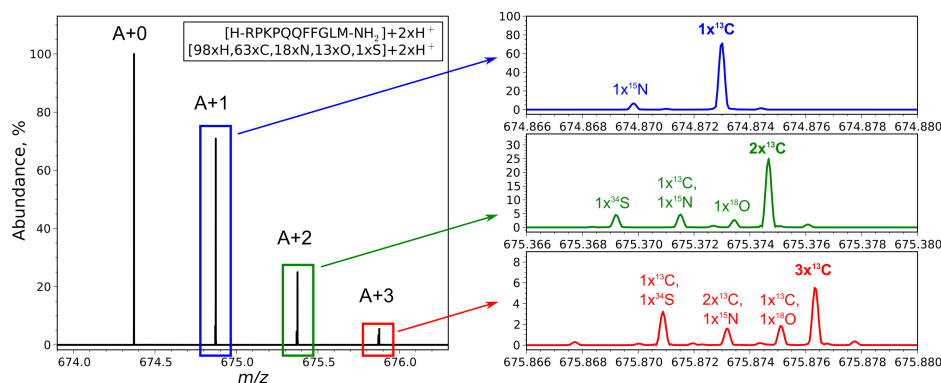


Figure 1.2: Isotopologues of peptide substance P. Left: A simulated mass spectrum of doubly charged peptide substance P (positive ESI mode). Right: Insets showing the (top) A+1, (middle) A+2, and (bottom) A+3 isotopic fine structures of ionized substance P.

1.2 Analytical questions in mass spectrometry

A mass spectrometer (MS) measures the mass-to-charge ratios m/z and relative abundances A of ionized molecules or atoms. The mass-to-charge ratio is a dimensionless ratio of the mass expressed in atomic mass units (a.m.u.), or Daltons (Da) depending only on the context, which corresponds to the mass of a ^{12}C atom being equal to 12 Da, to the number of elemental charges of the ion. The relative abundance characterizes the relative number of ions of given m/z .

Hence, one may conventionally distinguish direct and indirect analyses. Direct analysis is distinct by its aim, which consists in measurements of ion m/z values of ions and hence masses of molecules, as well as ion abundances and hence relative amounts of molecules. For example, direct analysis may include precise measurements of isotopic masses and abundances, as well as molecular quantitation. In contrast, indirect analysis aims to determine some extra information on the analyte using the measured values m/z and A . For example, this may include qualitative molecular analysis in which measured m/z values of the analyte or its fragments are used to determine the corresponding masses, which next are employed for molecular structure identification.

1.3 Mass spectrometry instrumentation

1.3.1 Brief review of history and milestones

Oftentimes, when one discusses the history of mass spectrometry, due consideration is given to events starting from early XIX century with groundbreaking work by J. J. Thomson in 1912–1913.[2, 3] This is exemplified by the celebration of 100 years of mass spectrometry at the largest global mass spectrometry oriented annual conference, the American Society for Mass

Chapter 1. Background

Spectrometry Conference, in Minneapolis, USA, in 2013. However, an honorary mention must go to his earlier work from 1897, when J. J. Thomson discovered and reported the m/z value of an electron.[1]

After receiving the Nobel Prize for this discovery in 1906, in 1912 Thomson proceeded to obtain what are considered the first mass spectra of several small gas molecules, such as nitrogen, CO₂ and oxygen by using a mass spectrograph.[2] The next year, Thomson obtained the mass spectrum of neon and therefore produced first evidence for the existence of elemental isotopes.[3] Towards the end of the decade, F.W.Aston begins his monumental work of determining metastable isotopes of most stable elements known at that time. This leads Aston to measurements of mass defects for various isotopes by early 1920s, which a decade later allows K. T. Bainbridge to prove Einstein's $E = mc^2$ postulation.[13]

From 1930s onwards, the field went through a number of revolutions in the instrumentation, including ionization techniques, fragmentation of ions, mass analyzers, and hybrid instruments, enabling analysis of more and more complex analyte mixtures and larger analytes. Specifically, since the late 1950s, gas chromatography coupled with mass spectrometry, which was developed by F.W. McLafferty, has significantly contributed to the analysis of volatile compounds. Discovery and development of matrix assisted laser desorption ionization (MALDI)[14] and electrospray ionization (ESI)[15] during the 1980s permitted to significantly expand the range of ionizable analytes with non-volatile compounds, including biological molecules.

In turn, *in situ* structural characterization of the analyte was significantly enhanced by introduction of techniques leading to inclusion of tandem mass spectrometry experiments, which allowed to isolate the analyte of interest, fragment it, then detect the product ions. Of particular prominence are: collision induced dissociation (CID), infrared multiphoton dissociation (IRMPD) and electron capture/transfer dissociation (ECD/ETD), which were introduced in late XX–early XXI centuries.

Finally, key developments in mass analyzers included the construction and commercialization of mass spectrometers based on various principles of mass separation, such as time-of-flight, quadrupole, radio-frequency ion trap, ion cyclotron resonance and orbitrap mass analyzers. In 1946, Stephens proposes the concept of time-of-flight mass spectrometer. Next decade, W. Paul develops the radio-frequency ion trap[16] and Dehmelt builds the Penning ion trap based on the cyclotron accelerator invented by Lawrence in 1931. In 1972, V.I. Karataev and B.A. Mamyrin develop reflectron time-of-flight mass spectrometer.[17] In 1974, A. G. Marshall and M. B. Comisarow develop Fourier transform ion cyclotron resonance mass spectrometer.[6, 7] And, based on the electrostatic trap developed by Kingdon in the early 1920s, A. A. Makarov invents the orbitrap mass analyzer at the turn of XXI century.[8]

Presently, there exist a lot of good references about the history of MS and FTMS. Particularly, the reviews and tutorials on FTMS can be good first steps for the curious reader to explore further.[18, 19, 20, 21, 22, 23]

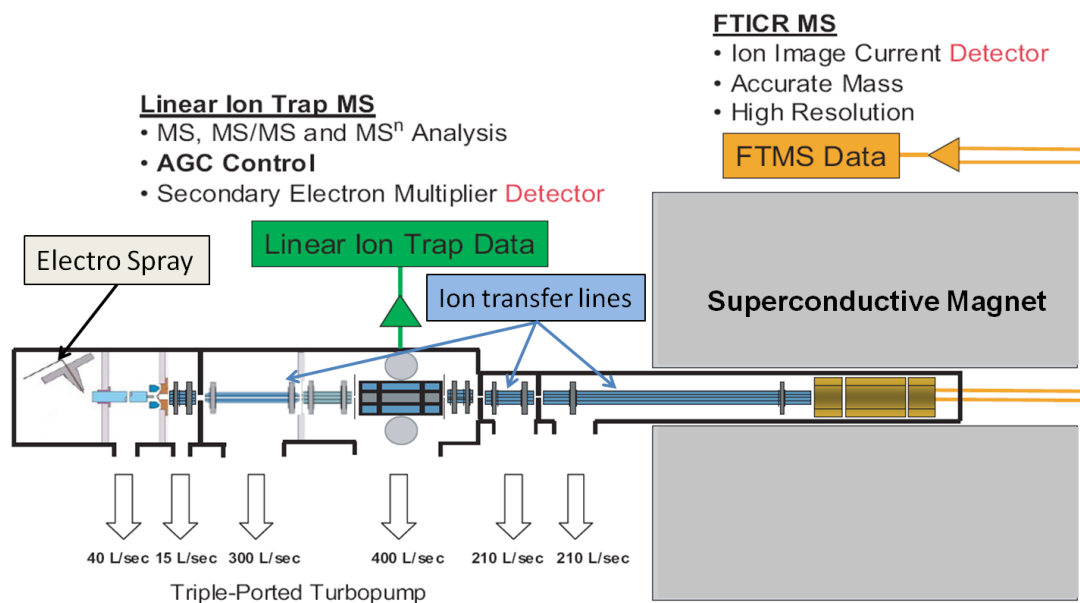


Figure 1.3: Schematic diagram of 10 T FT-ICR MS. Left: Atmospheric pressure ion source. Middle: Linear ion trap. Right: 10 T superconducting magnet with ICR cell. Figure adapted from the user manual LTQ FT Ultra.

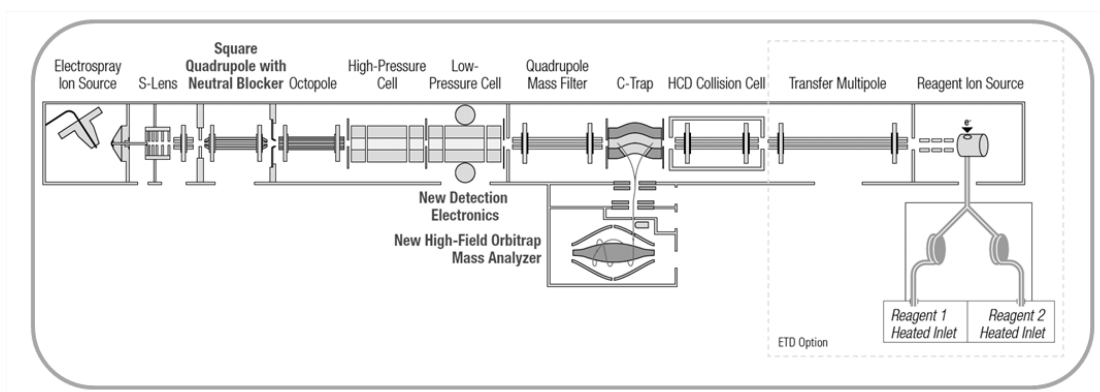
1.3.2 ICR and Orbitrap FTMS

In this work, the most of experiments were conducted on two hybrid instruments: 10 T ESI LTQ FT-ICR MS equipped with standard and customized ICR cells, as well as Orbitrap Elite FTMS (Thermo Scientific, Bremen, Germany). The schematics of FT-ICR MS is shown in Figure 1.3. From left to right by major section: ionization source followed by several ion guides (multipoles); in center, the linear ion trap, LTQ; made up of four electrodes set to RF electric potentials; ion transfer octapoles; the ICR cell. The ICR cell is located inside a 10 T homogeneous magnetic field parallel to the axis of ion transfer.

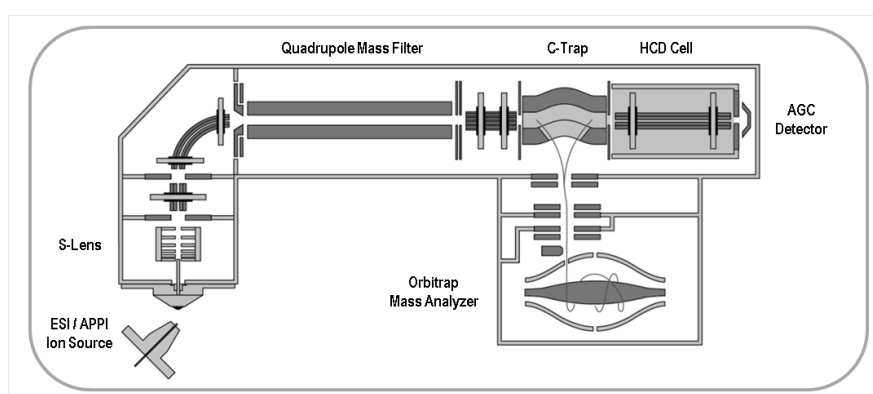
ESI method was employed for sample ionization. The following ESI ion sources were employed: the standard ESI source ($\mu\text{L}/\text{min}$ flow); a nano-ESI robotic chip-based source (nL/min flow), TriVersa Nanomate (Advion Biosciences, Ithaca, USA). The linear ion trap (LTQ) was used to isolate ions within a user specified m/z range (e.g. 2-10 m/z) which can then be followed by transfer of analytes into the ICR cell via the transfer octapoles. Additionally, the LTQ implements the automatic gain control (AGC) function to control total charge of ions injected into the ICR cell. After transferred into the ICR cell, the ions are axially trapped using electric fields on the trapping electrodes of the ICR cell, and radial confinement occurs due to the magnetic field.

Figure 1.4a shows schematics of the Orbitrap Elite FTMS. Similarly to the FT-ICR MS, the

Chapter 1. Background



(a) Schematic diagram of Orbitrap Elite FTMS.



(b) Schematic diagram of Q Exactive Orbitrap FTMS..

Figure 1.4: Schematic diagrams of Orbitraps. Figures adapted from <http://planetorbitrap.com>.

front end allows for interfacing identical atmospheric pressure ionization sources. The LTQ is interfaced with the C-trap, which is C-shape quadrupole intended for accumulating the ions and injecting them into the orbitrap. The electrodes of orbitrap are shaped such that the quadro-logarithmic trapping electric potential is created inside the trap. The orbitrap consists of a central spindle-like electrode and a barrel-like outer detection electrode split into two parts. The ion packet is injected into the orbitrap via a slot located in one half of the outer detection electrode, with additional help of a small deflection electrode situated nearby. As the ions injected into the orbitrap, a high voltage is applied to the central electrode (e.g. 3.5 kV).

Several experiments were carried out using Q Exactive FTMS (Thermo Scientific), Figure 1.4b. The Q Exactive Orbitrap FTMS differs by the presence of the quadrupolar mass filter instead of the LTQ; functionality of the other annotated components of its diagram is similar to those of the Orbitrap Elite FTMS.

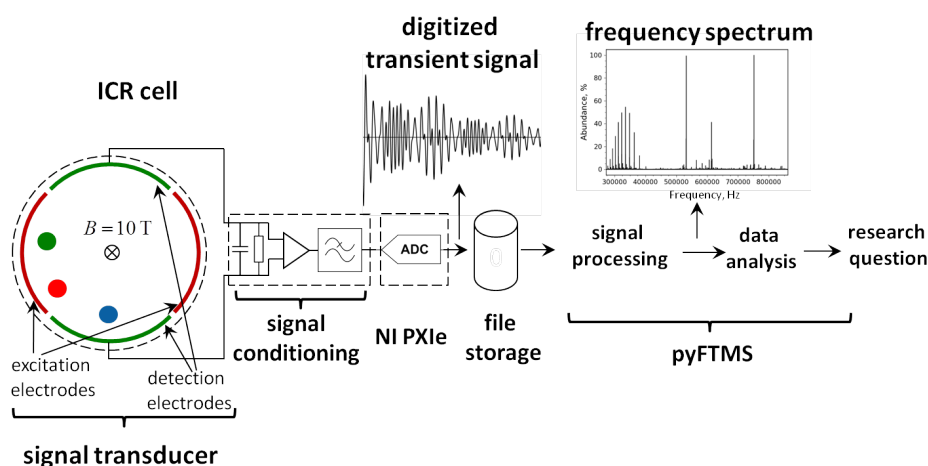


Figure 1.5: The measurement scheme in FT-ICR MS. The annotations NI PXIe and pyFTMS refer to specific data acquisition system and software implemented in this work.

1.3.3 Measurement setup

Figure 1.5 depicts the general measurement scheme in FTMS, as an example for FT-ICR MS. From left to right by major section: ion trap (ICR cell) as a signal transducer representing a stimulus-response system and intended for encoding the ion trajectory into electrical signal; signal conditioning electronics (including a preamplifier) meant for signal detection, amplification, and filtering; signal digitization (digitizer module), file storage, signal processing and data analysis.

The standard measurement scheme includes a standard mass analyzer, e.g. standard ICR cell or orbitrap, built-in digitizer, standard FT-based signal processing, viz. magnitude mode or absorption mode FT of a transient signal previously zero-filled and apodized; and data analysis, including peak picking procedure to locate the peak maxima in the Fourier spectrum, frequency-to- m/z conversion (mass calibration), and further analysis of m/z and A values in the context of the particular analytical question.[18]

The modified measurement schemes of the present research are distinct by developments in data analysis methods, methods for signal processing, and designs of ion traps. Additionally, a high-performance data acquisition system (NI PXIe) and a software framework for signal processing and data analysis (pyFTMS) were implemented as a software and hardware tool sets for the purposes of this work, Figure 1.5.

1.4 Uncertainty principle

In physics, the uncertainty principle is known due to Werner Heisenberg, who originally introduced it for position x and momentum p of a quantum particle.[24] Specifically, in

quantum mechanics the uncertainty principle states that the more precisely the position of a particle is determined, the less precisely its momentum can be known. The opposite holds true as well: the more precisely the momentum of a particle is determined, the less precisely its position can be known. The inequality relating the standard deviation of position σ_x and the standard deviation of momentum σ_p is given via the reduced Planck constant: $\sigma_x\sigma_p \geq \frac{\hbar}{2}$. Equivalently, the position and momentum of a particle may be referred to as complementary values of physical quantities, meaning that these quantities cannot be measured accurately at the same time and the limit to the measurement precision is given by the uncertainty principle's inequality.

The uncertainty principle does not relate only to position and momentum of a particle. In a broad sense, the uncertainty principle is applicable for any pair of complementary physical quantities. Indeed, the uncertainty principle in this case is nothing than any mathematical inequality asserting a limit to the precision with which certain pairs of physical quantities can be measured simultaneously. The particular lower bound limit in such inequalities, e.g. $\frac{\hbar}{2}$ for quantum mechanics, depends on the nature of the quantities involved or, equivalently, the state of the system that is described with these quantities. For instance, in signal processing the uncertainty principle holds that a function cannot be both time-limited and band-limited.[25]

In the context of FTMS, one corollary of the uncertainty principle relates to resolution performance and would hold that one cannot achieve high frequency resolution Δf at low detection period T : $\Delta f T \geq C$, where C is a constant value that depends on the state of the system of interest or, following the terminology of the present thesis, on the measurement scheme, Figure 1.5. To be exact, while the latter equation is correct as a qualitative dependence between the detection period and the scale of frequency details of interest, its particular form may vary. For instance, for value Δf of the frequency details scale, it may incorporate average frequency difference (e.g. FDM processing) or minimum frequency difference instead (e.g. FT processing). Moreover, the uncertainty principle can also be modified quantitatively by changing the constant C (e.g. FT processing, new designs of mass analyzers). Similarly, another corollary relates to limited performance for accurate measurements of a quantity q , viz. m/z or abundance values of ions, at finite signal-to-noise ratios (SNR), $\text{SNR} = \frac{q}{\sigma_q}$, with the latter defined in a broad sense, i.e. as the ratio between measured value q and its standard deviation σ_q . [26]

Thus, these two corollaries are what is referred to by the uncertainty principle in the present thesis. With the present thesis' developments aimed at modification of the uncertainty principles of each of the three main elements in the measurement scheme, namely signal generation, signal processing, and methods of data analysis, the corresponding uncertainty principle's form is modified such that improved analytical performance of FTMS is achieved.

2 Preliminary developments

2.1 Measurement theory

We start by formalizing the fundamental measurement principle of FTMS in order to develop the concept of the uncertainty principle in FTMS, which in turn governs achievable analytical performance for a given measurement scheme and hence will next be modified accordingly to achieve improved analytical performance.

2.1.1 Phase portrait of ion motion

To start, let us derive the phase portrait for ion motion in a given mass analyzer, e.g. an ICR cell or an orbitrap. Firstly, we consider a physical model describing motion of a single ion in uniform magnetic field and quadratic electrostatic field of an ICR cell. The following analysis is likewise applicable to the case of quadro-logarithmic field of an orbitrap, as discussed below in the text. With the Lorentz force of the electric field \vec{E} and magnetic field \vec{B} , acting on the ion of mass m and charge q , the ion motion equation is a differential equation of the second order:

$$m\ddot{\vec{r}} = q\vec{E} + q\dot{\vec{r}} \times \vec{B} \quad (2.1)$$

We choose a cylindrical coordinate system, $\{\rho, \varphi, z\}$, such that the magnetic field and the ion radius-vector are given as: $\vec{B} = \|0, 0, B_z\|$ and $\vec{r} = \|\rho, \varphi, z\|$, respectively, Figure 2.1. As there is no other charges considered, e.g. "space charge" and electrode surface charges, the electric field is solely defined by the distribution of electric potential, Φ , in the mass analyzer: $\vec{E} = -\nabla\Phi$. In turn, as the electric potential is defined by the ion trap's geometry and potentials applied to the ion trap's electrodes, Φ must satisfy Laplace's equation with the boundary conditions,

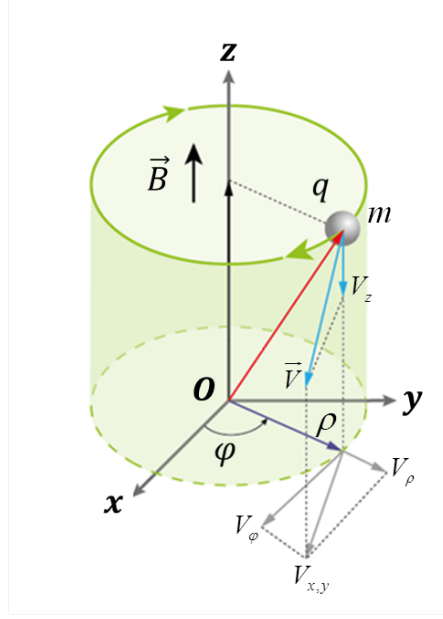


Figure 2.1: Representation of ion motion in the ICR cell, shown in a cylindrical coordinate system. Static homogeneous magnetic field is directed along the z -axis: $\vec{B} = \|0, 0, B_z\|$. The radius-vector describing the ion trajectory is $\vec{r} = \|\rho, \varphi, z\|$.

which can be given in the Dirichlet form:

$$\begin{aligned} \nabla^2 \Phi &= 0 \\ \Phi &= \begin{cases} V_t, & \text{on trapping electrodes} \\ 0, & \text{on other electrodes} \end{cases} \end{aligned} \quad (2.2)$$

Generally, Eq. 2.2 can be solved numerically, e.g. using ion optics simulation packages such as SIMION,[27] or derived analytically. For example, for the Penning trap, the electric potential is a simple quadratic form: $\Phi = A(2z^2 - x^2 - y^2) + B$, whilst the exact solution for ion traps employed in FT-ICR MS can be somewhat more involved. For example, the following equation is an analytical form for the electric potential of a cubic trap with edge d and trapping electrodes set to V_t while the other electrodes are kept at zero potential:[28]

$$\Phi(x, y, z) = \frac{16V_t}{\pi^2} \sum_{n,m=0}^{\infty} \frac{(-1)^{n+m} \cos\left[(2n+1)\frac{\pi x}{d}\right] \cos\left[(2m+1)\frac{\pi y}{d}\right] \cosh\left[\lambda_{mn}\frac{\pi z}{d}\right]}{(2n+1)(2m+1) \cosh\left[\lambda_{mn}\frac{\pi}{2}\right]} \quad (2.3)$$

where $\lambda_{mn} = [(2n+1)^2 + (2m+1)^2]^{1/2}$. Its Taylor's expansion about the trap center, $\Phi = \Phi^{(0)} + \Phi^{(2)} + o((\rho/d)^3)$, allows quadratic approximation: $\Phi = V \left[\delta + \frac{\alpha}{d^2} (2z^2 - x^2 - y^2) \right]$, where α and δ are dimensionless constants. In the context of a conventional cylindrical ICR cell, a similar form for quadratic approximation is expected with a different meaning of the metric parameter d . We thus proceed with the harmonic approximation of the cubic cell, while

assuming that the obtained results will also be valid for the harmonic approximation of the conventional cylindrical ICR cell of interest.

Notice that the equation for ion motion along the z -axis can be decoupled in Eq. (2.1):

$$\begin{cases} m\ddot{z} = -q\nabla_z\Phi \\ m\ddot{\vec{\rho}} = -q\nabla_{\vec{\rho}}\Phi + q\dot{\vec{\rho}} \times \vec{B} \end{cases} \quad (2.4)$$

With the following notations $a_0 = b_0 = 2\alpha qV/md^2$, $a_1 = -b_1 = -qB/m = -\omega_c$, and $c_0 = -4\alpha qV/md^2$, Eq. (2.4) takes the form:

$$\begin{cases} \ddot{x} = a_0x + b_1y \\ \ddot{y} = b_0y + a_1x \\ \ddot{z} = c_0z \end{cases} \quad (2.5)$$

From the theorem for existence and uniqueness of a solution for a system of linear differential equations with real-valued parameters, it follows that the solution of Eq. (2.5) is:

$$\begin{cases} x = A\sin(-\sqrt{\lambda_1}t + C) + B\sin(-\sqrt{\lambda_2}t + D) \\ y = A\cos(-\sqrt{\lambda_1}t + C) + B\cos(-\sqrt{\lambda_2}t + D) \\ z = E\cos(-\sqrt{\lambda_3}t + F) \end{cases} \quad (2.6)$$

where A, B, C, D, E , and F are real-valued constants defined by initial coordinates $\vec{r}(0)$ and velocities $\dot{\vec{r}}(0)$; λ_3 is a real-valued root of the characteristic equation $\lambda^2 - c_0 = 0$; and λ_1 and λ_2 are real-valued roots of the characteristic equation:

$$\begin{vmatrix} \lambda^2 - a_0 & -b_1\lambda \\ -a_1\lambda & \lambda^2 - b_0 \end{vmatrix} = 0 \quad (2.7)$$

Thus, we obtained three eigen modes of ion motion, known as cyclotron mode, magnetron mode, and axial mode, each characterized with one of the eigen values λ_i .

Next, to find the normal coordinates, we decouple the first and second equations in Eq. 2.6 by using the following variable: $\hat{\rho} = jx + y$, which is evaluated by Eq. 2.6 to: $\hat{\rho} = \hat{\rho}_+ + \hat{\rho}_-$, where

$$\hat{\rho}_+ = Ae^{j(-\sqrt{\lambda_1}t+C)} \text{ and } \hat{\rho}_- = Be^{j(-\sqrt{\lambda_2}t+D)}.$$

In the standard mode of operation of the ICR cell, the cyclotron mode is excited, while it is aimed to keep the magnetron mode non-excited. Hence, $\hat{\rho} = \hat{\rho}_+ = Ae^{j(-\sqrt{\lambda_1}t+C)}$. Thus, under these conditions, $\hat{\rho}$ satisfies to the following equation:

$$\ddot{\hat{\rho}} = \lambda_1\hat{\rho} \quad (2.8)$$

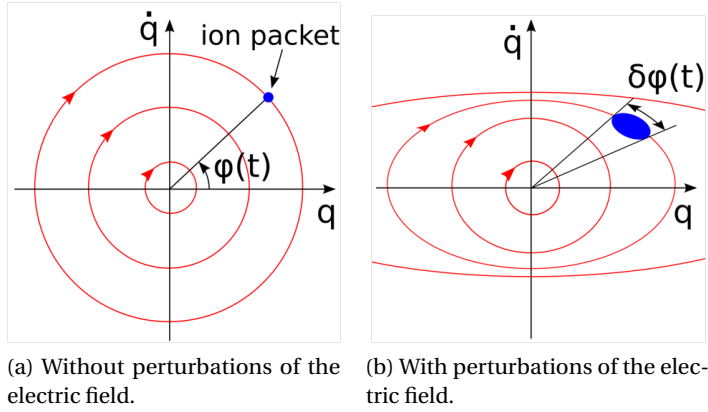


Figure 2.2: Illustration of phase portraits and ion packet sizes in the phase space. The phase space corresponds to normal coordinates for ion motion equations. These considerations are applicable for both FT-ICR MS and Orbitrap FTMS.

which is a canonical form of differential equation. An analogous equation would hold true for $\hat{\rho}_-$ as well, were the magnetron mode excited. Hence, we obtained the normal coordinates $\hat{\rho}_+$ and $\hat{\rho}_-$.

In the context of Orbitrap FTMS, the derivation of the 3rd equation of Eq. 2.5 is applicable also to the orbitrap mass analyzer as it is the axial mode of ion motion that is of interest there. That is, for the orbitrap the ion motion equation is the same type as Eq. 2.8.

From Eq. 2.8 it follows that the dynamic system "ion in a mass analyzer" considered here is a harmonic oscillator whose phase portrait is a set of closed circular contours, Figure 2.2a. Therefore, measurements of m/z values imply measurements of angular frequencies ω with which the analyzed ions move on their phase portraits in the phase space. Additionally, in analysis of ions with close m/z , the difference ΔL of the full paths of phase trajectories of the ions, acquired during a given detection period, provides a measure for the achieved ion separation.

As the next order of approximation, consider a more comprehensive physical model that takes into account static and time-dependent perturbations of the electric potential compared to the quadratic approximation considered above. Specifically, the static perturbations are due to the actual geometry of the ion trap and finite precision of trap manufacturing. In turn, the time-dependent perturbations include Coulomb ion-ion and ion-electrode interactions. The effects in question are known to cause decoherence of ion packets as well as coalescence of ion packets. However, the theory for these effects is not currently well-developed in FTMS. Quantitative description of these effects is beyond the scope of the present work. Nevertheless, a qualitative consideration is sufficient for its purposes.

From the initial model above it follows that the effect of decoherence may lead to artifacts in ion motion, e.g. known as ion packet spreading, ion comets formation, variable ion

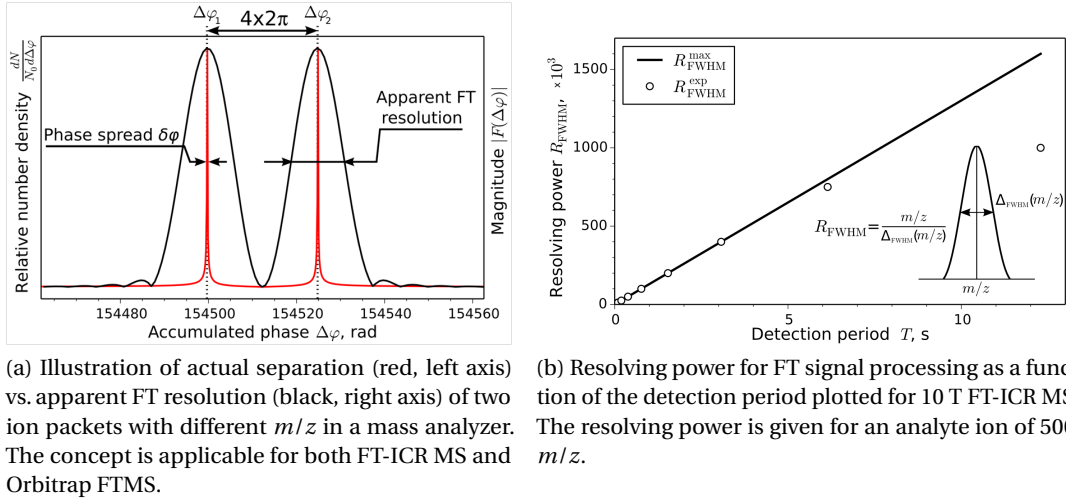


Figure 2.3: Ion separation vs. FT resolution performance in FTMS.

frequencies, [29, 30, 31, 32] and other detrimental effects distorting the phase trajectory of each ion during the detection period. That is, each ion from the ion packet now has its own phase trajectory that somewhat deviates from the closed circular contour that could be plotted for this packet if the detrimental effects were negligible, Figure 2.2b. Therefore, each ion packet corresponding to a certain m/z value has a finite size in the phase plane, δL . The size δL thus represents the physical limit to achievable ion separation.

2.1.2 Two sides to the uncertainty principle limit

Notice that the measure for ion separation may be represented as the phase spread of the ion packet, whereas measurement of ion frequency can be seen as measurement of the phase, $\Delta\varphi$, that the ion accumulates during the detection period, T . Indeed, in the initial model above, all ions of the same m/z would accumulate the same amounts of phase during the detection period, $\varphi = \omega T$. The same expression is also valid for the model allowing for perturbations if the quantity ω is read as the instantaneous angular frequency $\omega(t)$ averaged over the detection period, T : $\omega = \frac{1}{T} \int_0^T \omega(t) dt$.

This way, ion separation and m/z measurement may be represented as the distribution of the numbers N of ions over the accumulated amounts of phase, $p(\Delta\varphi) = dN/(N_0 d\Delta\varphi)$, Figure 2.3a, where N_0 is the total number of ions in the mass analyzer. Effectively, the distribution $p(\Delta\varphi)$ shows the actual separation of ions achieved in a mass analyzer, allowing also for the field perturbations introduced above. For example, the phase spreading of ion packets would be represented as nearly symmetric peaks in the distribution $p(\Delta\varphi)$, whereas the ion comet formation could lead to peaks with a noticeable asymmetry.

Thus, in $p(\Delta\varphi)$ the width of a peak represents the fundamental limit to ion separation, whereas

the mean value of phase where a peak is located provides the frequency of interest. Therefore, this consideration shows that separation of ions (resolution performance to which Chapters 4 and 5 are devoted) on one hand, and frequency measurement (measurement accuracy which will be considered in Chapters 3 and 5) on the other, are two naturally inter-mixed entities. Specifically, the two are the two corollaries of the uncertainty principle as introduced in Chapter 1. Indeed, the distribution $p(\Delta\varphi)$ represents the fundamental measurement limit in FTMS. This distribution can be considered in terms of ion separation as well as in terms of accuracies of measured frequencies. Hence, the concepts of resolution and measurement accuracy are two characteristics of the distribution $p(\Delta\varphi)$. Additionally, as it is the number density distribution, $p(\Delta\varphi)$ also comprises information on ion abundances. Hence, the abundance measurement accuracy can be given with similar probability density distributions plotted for each of the analyzed ions, Chapter 3.

Thus, in MS experiments the distribution $p(\Delta\varphi)$ should be the key characteristic of interest since it provides the required information on ion mass-to-charge ratios and abundances. Let us consider how this distribution is measured or, more specifically, estimated experimentally.

2.1.3 Transient signal as a response function

It is required to find out how exactly the phase portrait of a trapped ion is encoded into the transient signal as next this signal has to be decoded into m/z and abundance values. As a signal transducer, the mass analyzer, an ICR cell or an orbitrap, converts ion trajectories $\vec{r}(t)$ into an electric signal referred to as a transient signal $s(t)$: $s(t) = \mathcal{C}[\vec{r}(t)]$. In terms of dynamic systems, it is convenient to think of the transient signal as a transient response function of the dynamic system "ions in a mass analyzer" to an external stimulus, which is an excitation signal in case of the ICR cell and ion injection in case of the orbitrap, Chapter 1. Here we deal with the mapping \mathcal{C} of the mass analyzer. Specifically, in Paper I, we revisit the question of the ICR transient signal's composition.

Experimental characterization and potential use of other approaches than the standard detection with reduced cyclotron frequency peak have not been extensively studied in FT-ICR MS. In Paper I, firstly we examine characteristics of harmonics and their sidebands of the reduced cyclotron frequency in experimental spectra. Figure 2.4 shows a broadband FT-ICR mass spectrum of a monoisotopic ion of singly protonated MRFA peptide (nominal m/z 524) acquired on 10 T FT-ICR MS with the standard ICR cell. The spectrum contains peaks at $n\omega_+$, $n = 1, 2, 3, \text{etc.}$, corresponding to the n^{th} harmonics of the reduced cyclotron frequency ω_+ , and their magnetron sidebands at $\omega_{nk} = n\omega_+ \pm k\omega_-$, $k = 1, 2, 3, \text{etc.}$ The high order harmonics and the sidebands have low intensities compared to that of the ω_+ peak. Moreover, intensities of the sidebands ω_{nk} , relative to that of the corresponding harmonics peaks $n\omega_+$, increase with the order n of the harmonics. The high resolution FT mass spectrum (resolving power, RP > 100'000 at $m/z = 400$) reveals the magnetron sidebands around the first harmonic as symmetric pairs of peaks shifted equally above and below the ω_+ peak by frequency difference, f , equal

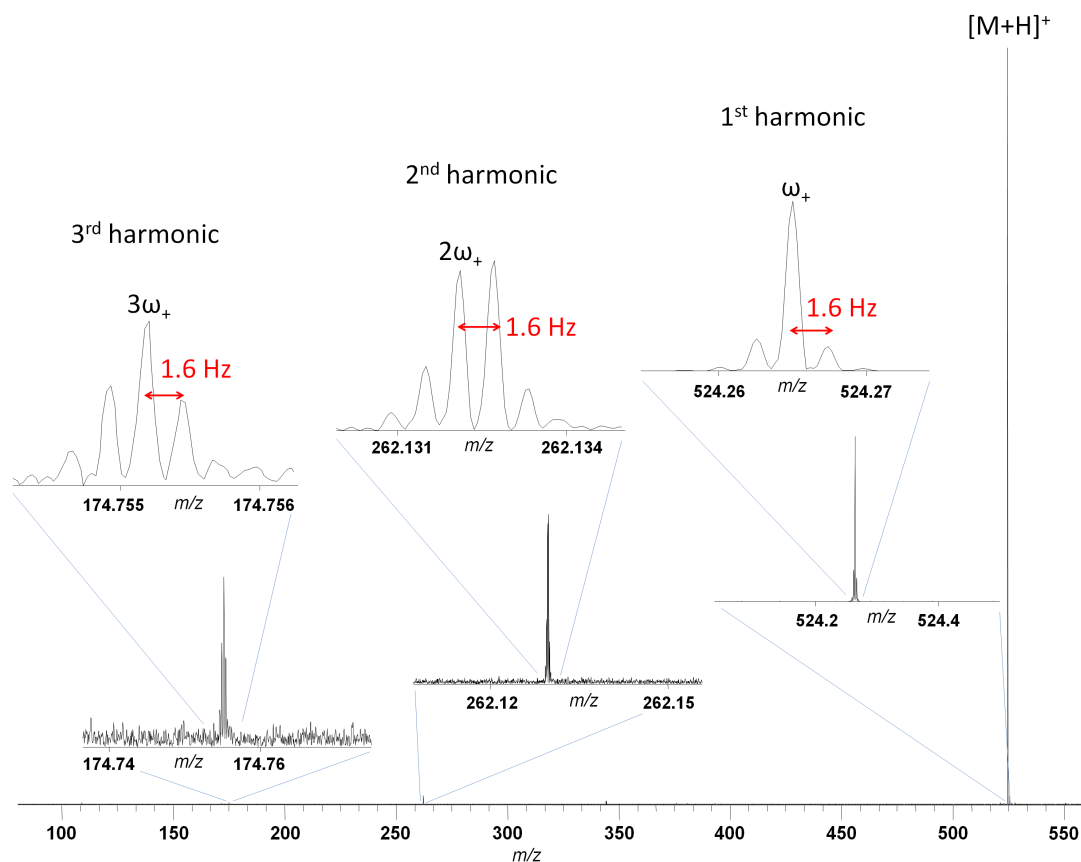


Figure 2.4: Harmonics and sidebands in a mass spectrum of monoisotopic ion of singly protonated peptide MRFA. Insets show the expanded regions around the 1st, 2nd, and 3rd harmonics of the reduced cyclotron frequency. The magnetic field was 11.5 T, the trapping potential was 0.8 V, and ion excitation amplitude was $105 V_{p-p}$ (35% of the maximum available value). The number of charges accumulated in the linear ion trap for subsequent injection into the ICR trap was set to 10^6 (AGC value). Figure adapted from Paper I.

to 1.6 Hz for the employed experimental parameters. The frequency ranges around the second and third harmonics show $2\omega_+$ and $3\omega_+$ peaks surrounded by their sidebands. The intensities distribution around the second harmonic is apparently asymmetrical in contrast to the first and third harmonics. Importantly, the sideband peak spacings f are the same for all harmonics.

Next, we develop an analytical model that contributes to the existing theory of the ICR signal by showing the origin of sidebands to be the result of the broadband amplitude-phase modulation occurring in the ICR transient signal. The initial model above with no time-dependent field

Chapter 2. Preliminary developments

perturbations is considered here. Firstly, we linearize the trajectory Eq. 2.6 to obtain:

$$\begin{aligned}\rho(t) &\simeq \rho_+ [1 + \varepsilon \cos(\Delta\omega t - \Delta\phi)] \\ \varphi(t) &\simeq \varphi_+ - \omega_+ t + \varepsilon \sin(\Delta\omega t - \Delta\phi)\end{aligned}\quad (2.9)$$

where we neglect the second and higher order terms of the small parameter $\varepsilon = \rho_- / \rho_+ \ll 1$, which represents the ratio of magnitudes for complex phasors $\hat{\rho}_-$ and $\hat{\rho}_+$, defined by the parameters A and B in Eq. 2.6. In Eq. 2.9, $\Delta\phi = \phi_+ - \phi_-$ is the difference of initial phases for the two modes of ion motion (defined by the parameters C and D in Eq. 2.6), and $\Delta\omega = \omega_+ - \omega_-$ is the difference of corresponding angular frequencies.

The obtained trajectory is a circular orbit $\rho = \rho_+, \varphi(t) = \varphi_+ - \omega_+ t$ perturbed by the curve $\Delta\rho(t) = \rho_+ \varepsilon \cos(\Delta\omega t - \Delta\phi), \Delta\varphi(t) = \varepsilon \sin(\Delta\omega t - \Delta\phi)$, Figure 2.5. Thus, the ion trajectory is a modulated circular orbit. Indeed, the ion trajectory shows azimuthal deviations $\Delta\varphi(t)$ around the $\varphi_+ - \omega_+ t$ value as well as radial deviations $\Delta\rho(t)$ around the ρ_+ value oscillating with the parametric frequency $\Delta\omega$ in both cases.

Finally, we establish the relation between the amplitude-phase modulation of the trajectory Eq. 2.9 and the resulting transient signal. In the approximation of elongated cylindrical ICR cell,[33] the following expression of the transient signal $s(t)$ was derived:

$$\begin{aligned}s(t) &\simeq \frac{q}{C} \sum_{m=0}^{\infty} a_m AM_m(t) PM_m(t) \\ AM_m(t) &= [1 + \varepsilon \cos(\Delta\omega t - \Delta\phi)]^{2m+1} \\ PM_m(t) &= \cos[(2m+1)(\varphi_+ - \omega_+ t + \varepsilon \sin(\Delta\omega t - \Delta\phi))]\end{aligned}\quad (2.10)$$

where $AM_m(t)$ and $PM_m(t)$ represent the amplitude modulation and the phase modulation terms of the m^{th} order, C is the effective input capacitance of the 1st cascade of the pre-amplifier, and a_m is an m -dependent factor. The oscillating radial and azimuthal deviations of the ion trajectory lead to the amplitude and phase modulations, respectively, occurring in the transient signal.

Thus, the physical origin of the harmonics and sidebands has been shown to be as follows: (i) finite cross-sectional dimensions of the ICR cell lead to its slightly non-linear response; (ii) if the ion magnetron motion is excited along with the cyclotron motion, the ion trajectory shows radial and azimuthal deviations oscillating with the parametric frequency around the ideal circular orbit; (iii) the oscillating radial and azimuthal deviations result in broadband amplitude and phase modulations, respectively, occurring in the ICR transient signal; and (iv) the amplitude-phase modulation produces interharmonics of the reduced cyclotron and magnetron frequencies and thus, specifically, the magnetron sidebands of the harmonics of the reduced cyclotron frequency.

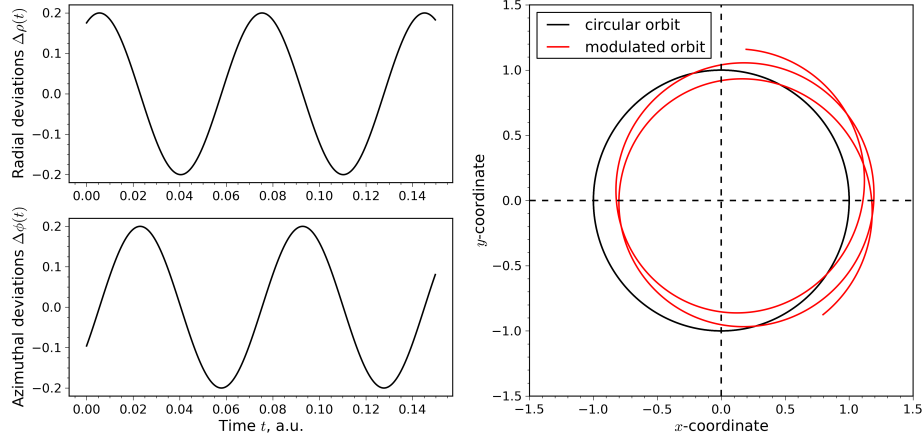


Figure 2.5: Numerical simulation of the linearized ion trajectory. Top left panel: radial deviations; bottom left panel: azimuthal deviations; right panel: the ion trajectory (red) as a result of the radial-azimuthal modulation of the circular orbit (black). Figure adapted from Paper I.

Besides, in the limiting case of $\varepsilon = 0$ (circular trajectory), the transient signal, Eq.2.10 is consistent with the previous studies considering circular ion motion:

$$s(t) = \frac{q}{C} \sum_{m=0}^{\infty} a_m \cos(2m+1)\omega_+ t \quad (2.11)$$

The corollary Eq. 2.11 shows that if the magnetron motion is quenched, the amplitude-phase modulation does not take place so that the ICR cell does not provide the interharmonics (sidebands). In turn, an implementation of a wide band transient signal generation with a number of non-zero coefficients a_m for increased spatial resolution of ion packets, translated into increased resolution of ions in the phase space and hence in the frequency domain, will be discussed in Chapter 5.

2.1.4 Limitations of FT signal processing

Let us introduce the concept of generalized measurement correlator, which extracts experimental quantities of interest, e.g. frequencies and amplitudes of sinusoidal components, from a given experimental signal. In state-of-the-art FTMS, the measurement correlator in question is implemented as standard FT signal processing workflow, Chapter 1. Indeed, the basis functions employed in FT are close, within the resolution limitation of FT, to the theoretically expected form of the transient signal of a single ion, e.g. Eq. 2.11 with $m = 0$. Hence, FT provides an estimate of the distribution $p(\Delta\varphi)$, while the difference between the estimate and the actual distribution is governed by the resolution performance of FT.

Consider a transient signal with length T acquired for an ion with mass-to-charge ratio m/z

Chapter 2. Preliminary developments

moving with the fundamental frequency (reduced cyclotron mode) f in the ICR cell. In general, here we do not limit consideration to the case of constant frequencies, i.e. the second model involving possible perturbations may be considered. Hence, the frequency f , in principle, can be time-dependent: $f = f(t)$. If these variations are small enough so that they are unresolvable by FT, the averaged frequency is then meant in the following text.

FT processing provides an estimate of the experimental distribution $p(\Delta\varphi)$, Figure 2.3a. As follows from comparison of resolution performance of different FT signal processing workflows, *vide infra*, Chapter 3, apparent FT resolution of ion packets is limited by the value:

$$\Delta_{\text{wnd}}\varphi = 2\pi C_{\text{FT}} \quad (2.12)$$

where C_{FT} is a constant defined by the apodization function and type, e.g., magnitude or absorption, of spectral representation. Indeed, in frequency domain, the corresponding broadening of peaks results from a convolution of the Fourier spectrum of the transient signal induced by ions and the spectral response of the apodization function. The spectral response has a full width (FW): $\Delta_{\text{FW}}f = C_{\text{FT}}/T$ and represents the time-vs.-frequency uncertainty principle of the FT signal processing workflow. Therefore, to baseline resolve two ions of different m/z , the required detection period T should satisfy to:

$$\Delta f T \geq C_{\text{FT}} \quad (2.13)$$

provided that the sizes of the two ion packets are less than their difference in accumulated phases. In Eq. 2.13, $\Delta f > 0$ is the difference of the frequencies of these ions. Hence, the difference of the phases that the ions accumulate during the time T should be:

$$|\Delta\varphi_1 - \Delta\varphi_2| = 2\pi\Delta f T \geq 2\pi C_{\text{FT}} \quad (2.14)$$

For instance, in the case of magnitude-mode spectral representation and the Hann apodization window, the constant C_{FT} equals 4, and, hence, the difference in question should be at least as large as 8π , i.e. 4 full cycles. However, the ion packet size in the phase space, $\delta\varphi$, might be significantly less than 4 cycles, Figure 2.3a. Hence, the apparent resolution of FT signal processing workflow may hide the actual ion separation achieved experimentally. This conclusion is also confirmed by experimental data as follows from measurements at multiples and harmonics of the fundamental frequency, Chapter 5. Thus, the FT workflow's time-vs.-frequency uncertainty principle defined for baseline resolution is found, Eq. 2.13.

From Eq. 2.13, it follows that there exists an upper bound of resolving power:

$$R_{\text{FWHM}}^{\text{max}} = \frac{m/z}{\Delta_{\text{FWHM}}(m/z)} = \frac{f}{\Delta_{\text{FWHM}}f} = \frac{fT}{C_{\text{FWHM}}} \quad (2.15)$$

The constant C_{FWHM} governs the full width at the half maximum (FWHM) of the spectral response, $\Delta_{\text{FWHM}}f = C_{\text{FWHM}}/T$, and is defined by the apodization window and spectral

representation type. For example, $C_{\text{FWHM}} = 2$ in the case of magnitude-mode spectral representation and Hann apodization window.

However, due to ion packet decoherence, resolving power achieved in experiments might be lower than provided by Eq. 2.15 for a given detection period T and mass analyzer. That is, as the ion packet decoherence develops, a longer detection period may be needed to achieve a required resolving power, and eventually the resolving power gets restricted by the ion packet spread. Therefore, in case of non-negligible ion decoherence, the experimentally achieved resolving power R_{FWHM} can be noticeably below the upper-bound $R_{\text{FWHM}}^{\text{max}}$. In Figure 2.3b, this is illustrated for a serial ICR cell (open circles).

Thus, if two coherent ion packets get separated in the phase space with respect to their accumulated phases, but the resolution performance of FT signal processing workflow is insufficient and hides the actual ion separation, an excessive data acquisition time T is required by FT. It is hence desirable either to achieve enhanced resolution for the detection period T , or to achieve the same resolution level but for a shorter detection period. These aims can be achieved with methods of super-resolution signal processing, Chapter 4, as well as with developments in instrumentation, Chapter 5, discussed in the following text.

2.2 High-performance data acquisition system

Data acquisition (DAQ) is the process of collecting data from sensors of measurement instruments. In the context of the experimental setup introduced previously, Figure 1.5, data acquisition refers to converting analog waveforms generated by the signal transducer, viz. the ICR cell or the orbitrap, into digital signals for further storage, analysis, and visualization. Generally, a DAQ system comprises DAQ devices, such as signal digitizers, to interface to signal sources, a host computer with a central processing unit (CPU) to communicate with the DAQ devices at the hardware level, and DAQ software to gain access to required functionality of the DAQ devices.[34] The DAQ devices and the CPU are connected through the host computer's bus, which in turn defines the hardware architecture of the DAQ system. The hardware architecture chosen to build a DAQ system is crucial as it imposes the limits to achievable performance in data acquisition. Specifically, the key characteristics of a computer bus are bandwidth and latency, Figure 2.6. The bandwidth measures the maximum amount of data that can be transferred through the bus in a unit of time; the latency defines the inherent delay in data transmission across the bus.

It is thus desirable that the DAQ architecture employed for an FTMS instrument is consistent with expected analytical characteristics of the total measurement setup. The previous generations of DAQ hardware for FTMS made use of custom-built electronics with commercial non-standard computers, the Industry Standard Architecture (ISA) bus,[36] the General Purpose Interface Bus (GPIB),[37] and the Versa module europa eXtension for Instrumentation (VXI) bus.[38, 39] Today, the Peripheral Component Interconnect (PCI) bus, which is a standard bus of the personal computer (PC), as well as the PCI eXtensions for Instrumentation (PXI)

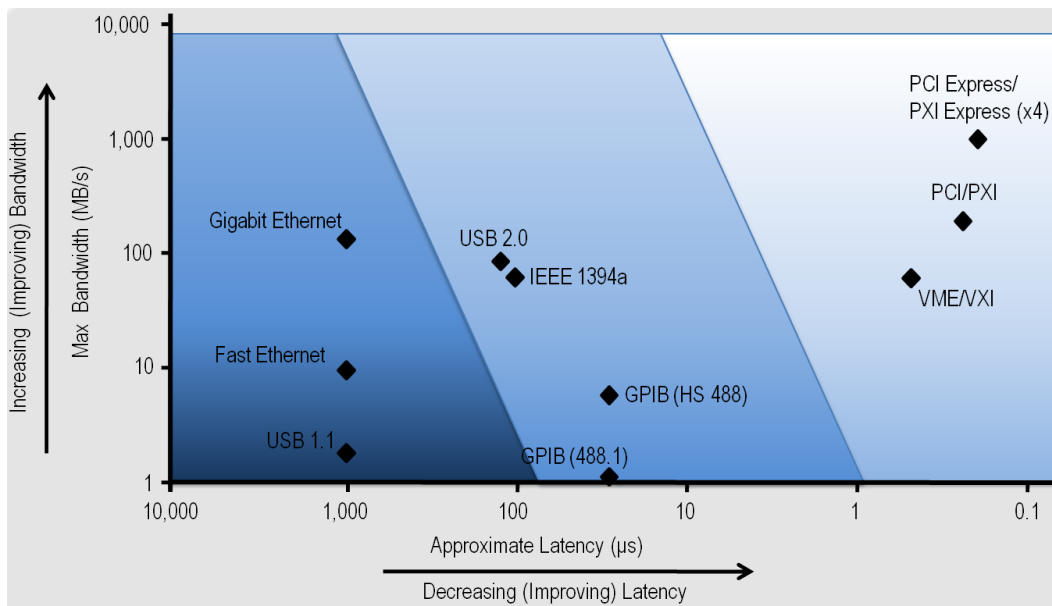


Figure 2.6: Comparison of various computer buses used in data acquisition, including GPIB and VXI buses, which were used in FTMS of previous generations, PCI/PXI buses, which are currently in use in state-of-the-art research-grade FTMS, and PXI Express, which has been employed for the purposes of the present thesis and is envisioned to gain adoption in next generations of FTMS. Figure adapted from Ref.[35]

bus, which is a PCI's derivative developed as an open industry standard for improved synchronization between different devices within the PCI-type bus, are widely employed for the most modern FTMS instruments, notably those developed in research laboratories.[40, 41, 42]

The pertinent characteristics of DAQ systems for FTMS instruments include the number of input channels, length of transient signals available for acquisition, spectral bandwidth available for signal digitizing, synchronization precision between different DAQ devices and other electronics of the FTMS instrument, vertical resolution of the digitizer imposing the digital noise-limited range of ion abundances available for analysis, and maximum duty cycle in sequential acquisition of transient signals. Most of these characteristics depend on the bandwidth of the bus employed because amounts of data to acquire scale as the number of employed channels, sampling frequency, and number of distinguishable intervals of signal amplitudes:

$$\begin{aligned} \text{Required Bandwidth (MB/s)} = & \hspace{15em} (2.16) \\ \text{Number of Channels} \times \text{Sampling frequency (MHz)} \times \text{Bytes/Sample} \end{aligned}$$

Despite the fact that the PCI/PXI architectures became the industry standard for data acquisition, data streaming capabilities of the PXI-based DAQ systems that are currently in use in FTMS are rather limited, primarily because of the restricted bandwidth.[43] Therefore,

these systems are usually programmed to implement the conventional acquisition mode with finite records, thus relying completely on the available amounts of the on-board memory of signal digitizers and generators employed.[40, 42] This mode has two crucial limitations when it comes to increased amounts of data to sample. Firstly, the maximum acquisition size is governed by the amount of on-board memory available on the digitizer. The on-board memory is expensive and amount of memory that can be installed on a digitizer is technologically limited. Secondly, the speed with which the acquired data will be transferred from the on-board memory to the host computer is governed by the bandwidth of the PCI or PXI bus. In this respect, while the PCI/PXI architectures' bandwidth, which is up to 132 MB/s, may be reasonable for many applications using the state-of-the-art research-grade FTMS instruments, higher bandwidth is required when amounts of data to transfer in a unit of time increase.

With development of the PCI Express bus in the PC industry and its derivative PXI Express, designed to offer dedicated bandwidth per instrument, higher values of bandwidth became available, e.g. 250 MB/s, 500 MB/s, and 985 MB/s per a single lane of the bus, depending on the bus generation.[35] Moreover, the PCI Express architecture allows grouping several lanes of the bus, e.g. 4, 8, or 16, into a single link, thus multiplying the bandwidth accordingly. For example, 1 GB/s dedicated bandwidth per a DAQ device can be achieved with the 4-lane PXI Express bus of the first generation (each lane 250 MB/s). Due to its bandwidth advantages, recently the PXI Express architecture has also gained attention for increasing the data throughput of the current systems in FTMS.[42, 44]

However, it is important to realize that not only does the new architecture remove the bottleneck limitations of the PCI/PXI architectures for fast data transfers, but it also enables efficient implementation of data streaming between different PXI Express devices on the bus, e.g. signal digitizers, field-programmable gate arrays (FPGA)-based modules for on-line signal processing, and signal generators. Hence, using the data streaming capability, the more sophisticated mode for signal acquisition or signal generation with circular representation of the on-board first in-first out (FIFO) buffer of a signal digitizer or signal generator may be employed so that the amounts of data to acquire and to generate are no more limited by the available on-board memory of these devices. Therefore, to meet requirements of the present thesis, a high-performance DAQ system has been implemented based on the PXI Express architecture and based on in-house developed DAQ software enabling the data streaming acquisition mode with continuous memory buffers.

2.2.1 PXI Express hardware platform

The hardware of the DAQ system implemented in the presented work is based on the PXI Express architecture and is built using commercially available PXI Express modules (National Instruments, Ennetbaden, Switzerland). The DAQ system includes a digitizer (NI PXIe-5122), a CPU-based embedded computer/controller (NI PXIe-8135), and a chassis (NI PXIe-1062Q), Figure 2.7. The digitizer provides a sampling frequency of up to 100 MHz and 14-bit resolution

Chapter 2. Preliminary developments

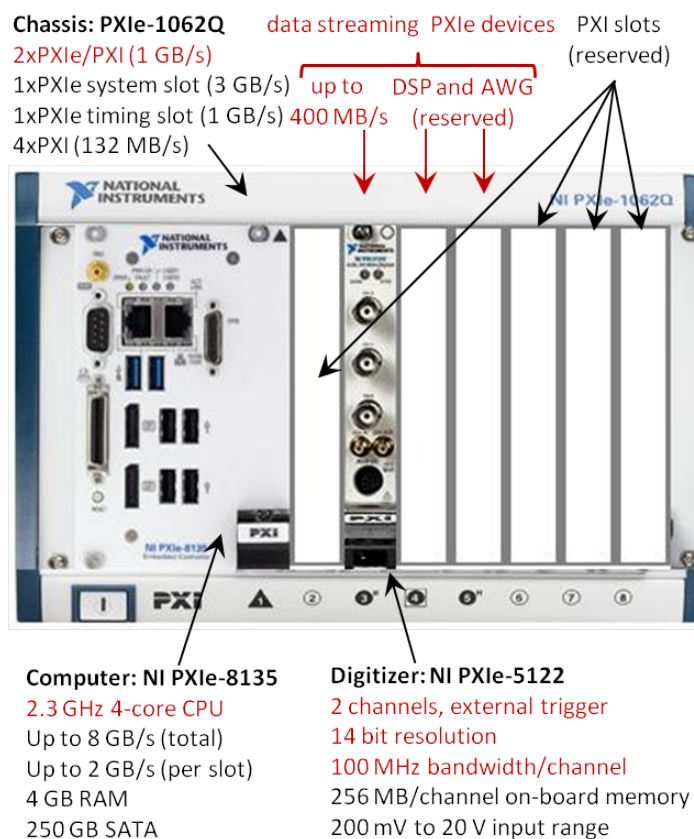


Figure 2.7: Schematic representation of hardware components of the PXI Express DAQ system implemented in the present work. The DAQ system comprises commercially available PXI Express modules, including a digitizer, an embedded computer, and a chassis, and is designed to enable the data streaming-based acquisition of transient signals.

per each channel, and is capable of data streaming at the full data rate of 400 MB/s via its 4-lane PXI Express interface.[45] The computer is based on a quad-core 2.3 GHz processor, comprises 4 GB of dual-channel random-access memory (RAM),[46] and intended to control the data acquisition process and collect the acquired data via the developed DAQ software. The chassis provides three PXI Express slots with 1 GB/s per-slot dedicated bandwidths, four PXI peripheral slots, and a system slot with approximately 3 GB/s bandwidth.[47]

Following the guest-host principle, the DAQ system acquires each transient signal by the triggering sequence generated by the host FTMS instrument, FT-ICR MS or Orbitrap FTMS. The output analog signal from the signal conditioning unit, viz. the pre-amplifier, of the experimental setup, Figure 1.5, is connected to the first input channel of the DAQ system's digitizer, while the default connection of this signal to the built-in electronics of the FTMS is also preserved. For FT-ICR MS, a custom-made adapter is employed to pick up the triggering sequence from a buffered digital trigger generated by the built-in data acquisition electronics of the instrument. Additionally, when a modified detection scheme with generation of two

transient signals on the FT-ICR MS is used, the second signal is connected to the second input channel of the digitizer.

The free slots are reserved for extension modules such as an FPGA module for on-line digital signal processing (DSP), a 2-channel arbitrary waveform generator (AWG), and signal amplifiers. All the hardware units of the DAQ system are balanced and matched to provide the necessary performance when dealing at the maximum bandwidths of each PXI Express device of data streaming capability, namely the digitizer with its full data rate of 400 MB/s, as well as the optional DSP and AWG modules. Overall, the DAQ hardware meets the following requirements of the present work: two input channels with 14-bit vertical resolution each to achieve broad digital noise-free dynamic range; high-precision synchronization with the host FTMS instrument via an analog trigger input (10 ns and better); available sampling frequency of up to 100 MHz to enable digitization of transients with wide spectral bandwidth (up to 50 MHz). As a future extension, the DAQ system can be equipped with a 2-channel signal generator to implement an option for quadrature ion excitation in future developments, Section 5.2.

2.2.2 Data streaming software

To program operation of the DAQ system's hardware, data acquisition software was developed with the LabVIEW environment and NI-SCOPE instrument drivers (National Instruments). The graphical user interface of the developed software for data acquisition is shown in Figure 2.8. The core functionality of the DAQ software is the data streaming mode for signal acquisition,[48, 49] including data streaming directly between the DAQ devices, producer-consumer acquisition loop for data transfers through the embedded computer's random-access memory (RAM), triggered acquisition for recording transients of a given length, and file output optimizations. Specific algorithms of the DAQ system's operation include recognition of encoded trigger signals, i.e. triggering sequences, from the host FTMS, viz. those for the FT-ICR MS, Orbitrap Elite FTMS, and Q Exactive Orbitrap FTMS; configuration of the acquisition sequence, including detection time per each transient, number of transients to acquire, duty cycle; and options for on-line digital signal processing as well as data visualization. Major parts of signal processing of the present thesis were performed using the pyFTMS framework developed in-house and described in the following subsection.

Overall, in the context of requirements of the present work, the developed DAQ software provides a user interface for data acquisition configuration and enables data streaming of the PXI Express architecture employed, thus allowing acquisition of transient signals with large numbers of data points without restrictions imposed by the digitizer's on-board memory amount. As a future numerical optimization, the DAQ system can be equipped with an FPGA module to replace some of performance-critical signal processing steps, which were performed in off-line mode using a multi-core CPU, with on-line computations using an FPGA chip.

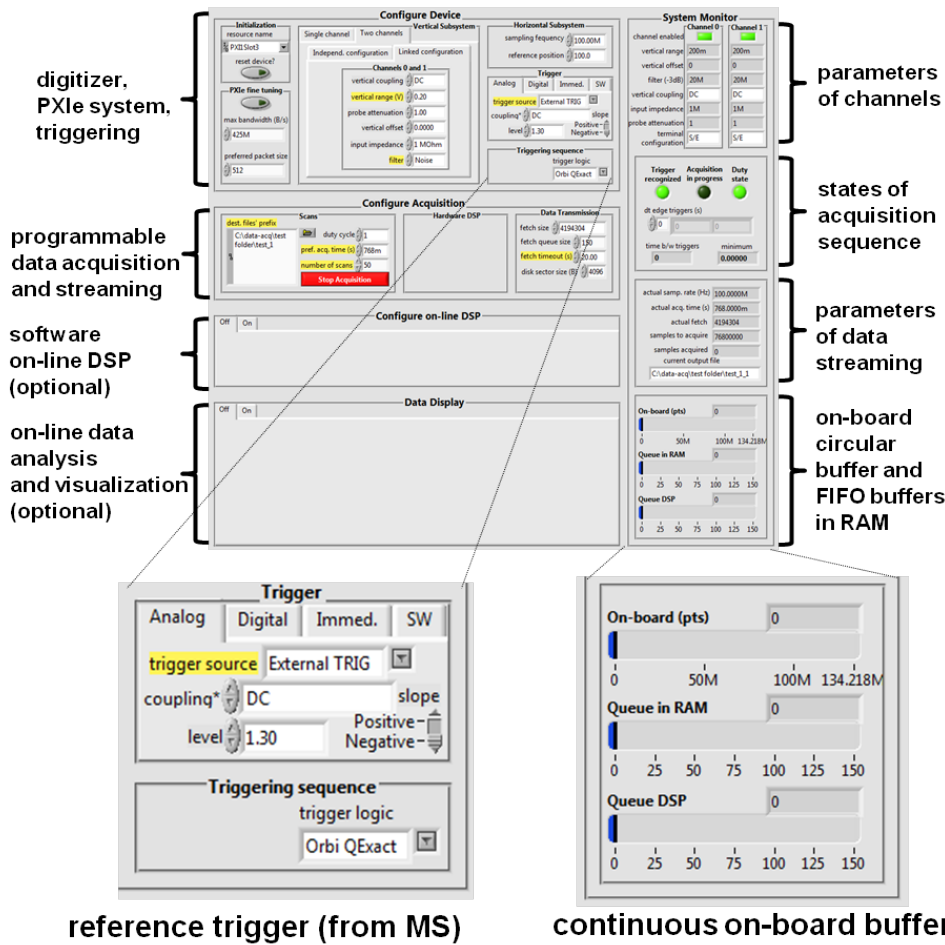


Figure 2.8: Graphical user interface of the DAQ software developed in the present work to implement the data streaming acquisition mode for PXI Express DAQ devices.

2.3 pyFTMS software framework

Traditionally, FTMS software was written using low-level statically-typed programming languages such as C, C++, Fortran, and Java, e.g. ICR-2LS,[50] Boston University Data Analysis (BUDA),[51] Arbitrary Waveform Editor (AWE) software,[40] Modular Ion cyclotron Data Acquisition System (MIDAS) software,[38, 39, 42] and MassSpecter.[41] Generally, the low-level languages are preferred in performance-critical applications because a computer program written in such a language is compiled directly into processor instructions, which are relatively fast to execute by the processor. However, the major disadvantage of these languages is that it takes considerable time to develop high-quality software, as well as to maintain or extend existing software.

In contrast, taking advantage of recent developments in dynamically-typed high-level programming languages, notably the Python language,[52] the developer's productivity and the quality of produced code can be significantly increased. First appearing 20 years ago,

today Python is a powerful general-purpose programming language with in-built software design philosophy making for code quality and hence code extensibility, as well as code development speed.[52] In the context of MS data processing, Python's advantages has recently gained interest in proteomics and related applications.[53, 54, 55, 56, 57] Importantly, by supplementing Python with third-party libraries written in a compiled language for execution speed performance, Python becomes a sophisticated tool for numeric programming. Indeed, over the past five years, numeric programming with the Python programming language armed with gold-standard C and Fortran libraries became a widely recognized and efficient approach for software development in science and engineering.[58]

Therefore, to implement required software tools in the present work, a software framework, pyFTMS, was developed with Python programming language and standard software packages for scientific computing. Namely, the methods for data analysis and signal processing developed in this work, as well as the data analysis workflows of this work, were built on top of the core functionality provided by the pyFTMS software. The framework is based on Python's extensions for scientific computing. Specifically, the NumPy extension is for fundamental numeric programming, the SciPy extension for general numerical computation, the matplotlib extension for data visualization, the FFTW library for discrete Fourier transform,[59] and a custom library for the filter-diagonalization method (FDM). Least-squares fitting (LSF) of transients is implemented by a custom library based on the MINUIT package.[60]

By design, the framework is divided into several modules, each serving different purposes, including cheminformatics of elemental and isotopic compositions, bioinformatics of peptides and proteins, file input and output (I/O) operations, signal processing, and analysis of mass and Fourier spectra. The signal processing module provides FT, LSF, FDM, as well as related signal processing elements such as apodization windows for filter design and spectral analysis, least-squares solution of matrix equations, phase correction methods, efficient implementation of finite impulse response (FIR) filtering/decimation. Based on this functionality, absorption mode FT, double phase correction FT, and FT method with extended basis were implemented. The module for analysis of mass spectra implements mass calibration, including forward and inverse m/z -vs.-frequency conversion based on the standard laws,[61, 62, 63, 8, 64] as well as the iterative mass calibration method for Orbitrap FTMS and mass calibration using ω_c regime of FT-ICR MS with a NADEL cell; and provides methods for peak picking and noise thresholding of mass spectra.

Additionally, based on the NIST isotopic masses and abundances,[9] the cheminformatics module features in-silico representations of isotopic compositions and provides mass and abundance computations of exact isotopic distributions, including isotopic fine structures. Based on elemental compositions of the 20 standard amino acids, common post-translational modifications, and cleavage rules of enzymes, the bioinformatics module supports in-silico representations of amino acid sequences including peptides and proteins. Finally, the file I/O module provides functionality for reading and writing various types of MS files including transient files, e.g. the HDF5 file format, and mass spectra files, e.g. the MGF format.

3 Data analysis

As the first step towards improved analytical performance of FTMS, this chapter reports developments made specifically in, yet not limited to, the framework of the standard measurement scheme of FTMS. Based on the considerations developed in Chapter 2, the current chapter presents different methods for data analysis providing improved analytical characteristics such as increased measurement accuracy of ion mass-to-charge ratios and abundances and maximized dynamic range of ion abundances for given experimental conditions.

Thus, the current chapter is focused on the first of the two inter-related groups introduced in Chapter 2 with the consideration of the uncertainty principle of measurements in FTMS. As shown in Chapter 2, the overall measurement scheme provides only an estimate for the actual distribution of the numbers of ions over the mass-to-charge ratios. Since each of them may introduce its own limitation, the current chapter considers the three elements of the measurement scheme separately.

Section 3.1 discusses methods to control the inherent level of measurement errors of FT signal processing workflow, including those resulting from the detrimental effects of spectral discreteness and spectral interference. Then, Section 3.2 describes detrimental effects of thermal noise and digital noise resulting from the signal transducer, signal conditioning electronics, and signal digitization. Finally, Section 3.3 presents a data analysis method for accurate estimation of the mass calibration function, thus enabling comprehensive analysis of petroleum samples with Orbitrap FTMS.

3.1 Signal processing with FT

Let us specify the following notation for signals, transforms, and functions used in the following text. Fourier transform (FT) of a continuous signal $s(t)$ is given as follows:

$$S(f) = \int_{-\infty}^{+\infty} s(t)e^{-j2\pi ft} dt \quad (3.1)$$

Chapter 3. Data analysis

The following form for the discrete Fourier transform (DFT) of a uniformly sampled signal $s_n = s(nt_s)$, $n = \overline{0, N-1}$, $t_s = 1/f_s$, where f_s is the sampling frequency and N is the number of points, is used:

$$S_m = \sum_{n=0}^{N-1} s_n e^{-j2\pi nm/N}, \quad m = \overline{0, N-1} \quad (3.2)$$

As follows from Chapter 2, in the standard measurement scheme the model for the transient signal with detection period T can be written as:

$$s(t) = \sum_{k=1}^K A_k \cos(2\pi f_k t + \varphi_k), \quad t \in [0, T] \quad (3.3)$$

where f_k , A_k , and φ_k are the frequency, amplitude, and initial phase of the k^{th} sinusoidal component in the transient signal induced by K ion packets of different m/z . Parameters A_k , and φ_k represent, respectively, the total charge (the ion abundance times the charge state of ions) and initial phase of the k^{th} packet of ions, though these parameters may also imply influence of the amplitude-frequency and phase-frequency functions of the signal conditioning electronics.

The digitized transient signal is given as a sampled version of the signal $s(t)$:

$$s_n = \sum_{k=1}^K A_k \cos(2\pi f_k t_s n + \varphi_k), \quad n = \overline{0, N-1} \quad (3.4)$$

From the convolution, linearity, and shifting properties of the DFT, as well as the DFT spectrum of a complex exponent, it follows that the DFT of the sampled transient Eq. 3.4 can be expressed as follows:

$$S_m = \sum_{k=1}^K \left[\frac{A_k}{2} e^{j\varphi_k} W_{m-m_k} + \frac{A_k}{2} e^{-j\varphi_k} W_{m+m_k} \right], \quad m_k = f_k N / f_s, \quad m = \overline{0, N-1} \quad (3.5)$$

where W_m is given as follows:

$$W_m = \frac{\sin(\pi m)}{\sin(\pi m/N)} \frac{e^{-j\pi m}}{e^{-j\pi m/N}}, \quad m \in \mathcal{Z} \quad (3.6)$$

and is the DFT of a rectangular window $H_n - H_{n-N}$ formed using the discrete Heaviside function H_n :

$$\begin{cases} H_n = 0, & n < 0, \\ H_n = 1, & n \geq 0 \end{cases} \quad (3.7)$$

As a reference, let us also derive the FT of the corresponding continuous signal Eq. 3.3:

$$S(f) = \sum_{k=1}^K \left[\frac{A_k}{2} e^{j\varphi_k} W(f - f_k) + \frac{A_k}{2} e^{-j\varphi_k} W(f + f_k) \right] \quad (3.8)$$

where $W(f)$, being the FT of the rectangular window $H(t) - H(t - T)$ formed using the continuous Heaviside function $H(t)$, can be written as follows:

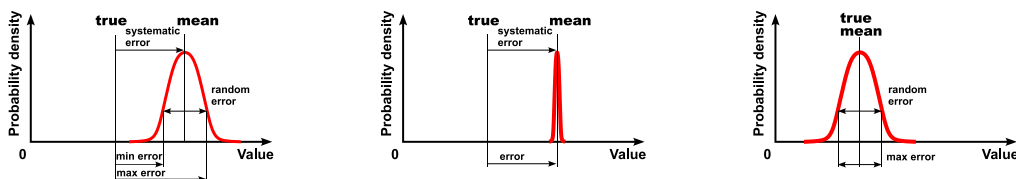
$$W(f) = T \frac{\sin(\pi f T)}{\pi f T} e^{-j\pi f T} \quad (3.9)$$

Known as the Dirichlet kernels of the FT and DFT, Eqs. 3.6 and 3.9 describe the envelopes of ion signals in the spectra S_m and $S(f)$. It is worth noting that expressions like these define uncertainty principles for frequency measurements for a given signal processing workflow based on FT, with Hann apodization and zero-filling included, be it the magnitude-mode FT, absorption-mode FT, or the signal processing method developed for double phase correction, Chapter 4.

Additionally, from comparison of Eqs. 3.6 and 3.9, one may see three key aspects of discrete Fourier spectra obtained in FTMS vs. continuous Fourier spectra provided by the theory of the transient signal, Chapter 2: (i) the discrete transform S_m is close to the continuous transform $S(f)$, $f = mf_s/N$, at the points around $m = \pm m_k$, $k = \overline{1, K}$; (ii) S_m is a sampled estimate of $S(f)$; and (iii) S_m notably deviates from the $S(f)$, $f = mf_s/N$, at points m located far from $\pm m_k$, $k = \overline{1, K}$, demonstrating spectral replication and phase variation effects of the DFT. Hence, while the DFT provides a reasonable approximation of the FT in analysis of FTMS transient signals, the DFT has specific properties, viz. cyclic spectral replication, phase variation and spectral sampling, which potentially causes measurement artifacts, when not taken into account. In the following text, the effect of spectral discreteness is considered, while cyclic spectral replication is analyzed in the context of goal of double phase correction method, Chapter 4.

3.1.1 Spectral discreteness

The following consideration is focused specifically on errors in measurements solely due to inherent limitations of the FT signal processing workflow. Typically, in experimental practice it is required to keep these errors below a given threshold, which in turn is defined by other experimental errors that are hard to exclude. As follows from Chapter 2, errors of measurements, including those considered in this chapter, are a particular case of the uncertainty principle in FTMS. That is, measurements of the m/z and abundance values for an ion can be characterized with a probability density plot, which in turn characterizes the errors in question and provides a measure for the range in which the true value is actually located, Figure 3.1.



(a) Systematic and random errors are comparable. (b) Systematic errors predominate. (c) Random errors predominate.

Figure 3.1: Random errors vs. systematic errors classification.

Here we start with errors due to discrete nature of DFT spectra. Other two detrimental effects of signal processing, i.e. spectral replication and phase variation are neglected. The effect of spectral discreteness causes inaccurate determination of peak maxima leading to frequency (m/z) and abundance measurement errors, Figure 3.2. Complementing past considerations of zero-filling technique known in the literature,[65, 66] here we perform the following: (i) we derive the values of m/z and abundance errors taking into signal apodization, e.g. Hann window; (ii) probability density functions describing the errors in question are obtained; and (iii) based on the obtained results, dependence of the errors in question is analyzed as a function of the resolution setting of a mass spectrometer, i.e. length of the transient signal.

Here, the case of magnitude-mode FT is exemplified. To take into account the effect of signal apodization, the following approach for expansion of the spectral response of a given apodization window is performed. As follows from Eq. 3.9, when no apodization is performed, the spectral peak shape is:

$$|W_{\text{rect}}(f)| = T \frac{\sin(\pi f T)}{\pi f T} \quad (3.10)$$

which can be said to be the case with a rectangular apodization window of length T . Next, the signal is apodized with a function $w(t): s(t) \cdot w(t)$. For instance, the Hann apodization function as the standard apodization in FT-ICR MS,[21] is considered. The corresponding peak shape can now be represented in terms of the spectral peak Eq. 3.9:

$$|W_{\text{hann}}(f)| = \left| \frac{0.5}{1 - (fT)^2} W(f) \right| = \left| 0.25W\left(f - \frac{1}{T}\right) + 0.5W(f) + 0.25W\left(f + \frac{1}{T}\right) \right| \quad (3.11)$$

In the context of discrete spectra, a sampled version of Eq. 3.11 is obtained as shown previously. Typically, to reduce the effect of spectral discreteness, the well-known operation of padding the signal with zeros, or zero-filling, can be applied. Zero-filling operation can be taken into account, resulting in frequency spacing $\Delta f = \frac{1}{2^n T}$ in the sampled version of Eq. 3.11, where n is the number of zero-fillings.[66] For instance, one zero-filling is typically used in the standard processing in FT-ICR MS; however, higher number of zero-fillings can also be employed when required, e.g. as discussed below.

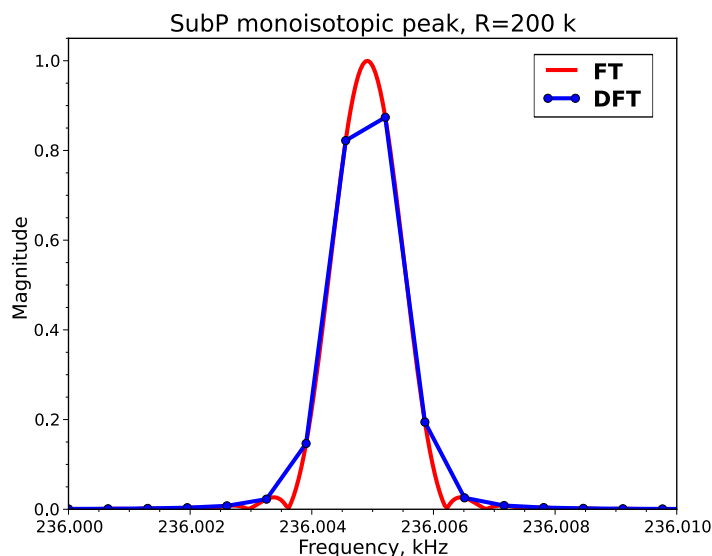


Figure 3.2: Illustration of mass and abundance errors due to spectral discreteness: a DFT spectrum obtained using the FT signal processing workflow (with the Hann apodization function, a single zero-filling, and the magnitude-mode spectral representation) vs. a corresponding continuous FT approximation. A transient signal was numerically simulated with parameters corresponding to monoisotopic peak of doubly protonated peptide substance P and 10 T FT-ICR MS instrument. The length of the transient corresponds to resolving power of 200'000 in magnitude-mode Fourier spectra. For illustration purposes the continuous FT approximation was obtained using DFT processing with sufficiently high number of zero-fillings.

Now, we proceed to obtaining the probability density functions providing the measure of the errors under consideration. For instance, for the frequency (m/z) errors:

$$\Delta f(f, T, n) = \frac{k}{2^n T} - f \quad (3.12)$$

where $k \in \mathcal{Z} : \left| \frac{k}{2^n T} - f \right| \rightarrow \min$. Hence, the probability density function in question is a delta function, Figure 3.3a, centered within $[f_0 - \Delta_{\max} f; f_0 + \Delta_{\max} f]$. It should be noted that the considered estimate, Eq.3.12, is obtained as a noise-free limit. Given finite SNR values of experimental data, the analytically useful maximum number of n is thus restricted.

A similar result holds true for the errors in abundance measurement: $[A_0 - \Delta_{\max} A; A_0]$, Figure 3.3b. Thus, these errors demonstrate systematic nature. For instance, frequency errors depend on detection time T , number of zero-fillings n , apodization window constant C_{FT} , and the true frequency value f_0 . Besides, maximum errors $\Delta_{\max} f = \frac{1}{2^n T}$ are independent of f_0 and serve as upper bound estimations for the actual errors since f_0 is unknown in usual practice unless a special experiment is designed.

When significant and not taken into account, these errors may even hide the random errors,

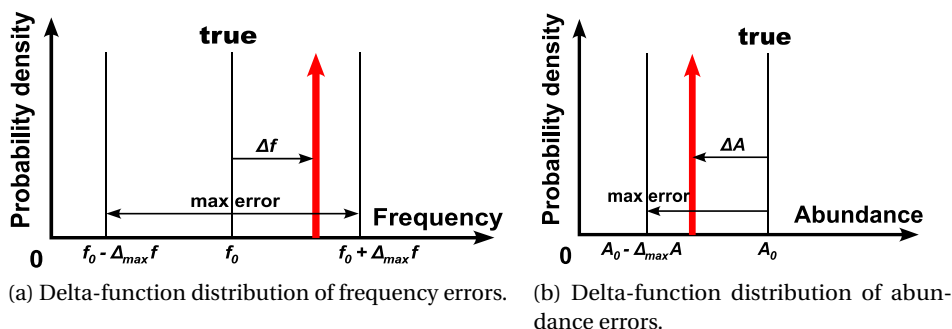


Figure 3.3: Probability density functions due to spectral discreteness. The results were obtained as a noise-free limiting case.

especially for sufficiently short transient signals. Therefore, in the next step we estimate these errors as a function of resolution setting (detection time) and numbers of zero-fillings, in order to be able to keep these errors below a given level, Figure 3.4a, Figure 3.4b, when dealing with Fourier transforms of short signals, e.g. 100 ms. For instance, the relative error of frequency (m/z) is estimated in the following way:

$$\frac{\delta_{\max f}}{f} = \frac{1}{2^n} \frac{1}{fT} = \frac{1}{2^n} \frac{1}{C_{\text{FWHM}} R} \quad (3.13)$$

where R is the resolving power setting (for FT-ICR MS) and C_{FWHM} is a constant defined by apodization window and type of spectral representation, Chapter 2. For example, $C_{\text{FWHM}}=2.00$ in case of magnitude-mode FT with Hann apodization window. For Orbitrap FTMS, the value R here equals the resolution setting times two.

3.1.2 Non-linear spectral interference

Previously, spectral interference was considered for exponentially damped as well as non-apodized signals[67]. In the context of FT errors, we extend the consideration of spectral interference effects in the context of magnitude and absorption modes and with enabled apodization.

To exemplify, notice that as follows from Eqs. 3.5 and 3.6, a magnitude-mode spectrum with close spectral components (corresponding to different ion packets) has an inherent non-linearity, which results in interference of these components. Indeed, consider the power spectrum:

$$P_m = |S_m|^2 = \sum_{k=1}^K S_k \sum_{l=1}^K S_l^* = \sum_{k,l=1}^K S_k S_l^* \quad (3.14)$$

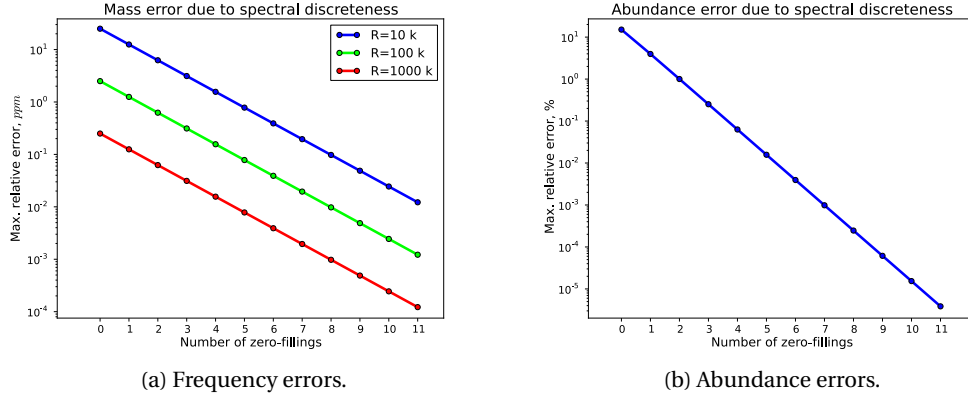


Figure 3.4: Estimated errors due to spectral discreteness as a function of number of zero-fillings and the setting of resolving power.

In terms of the peak shape W_m , Eq. 3.6, P_m can be expressed as the sum $P_m^+ + P_m^- + P_m^{+/-}$, where:

$$P_m^+ = \sum_{k,l=1}^K \frac{A_k A_l}{4} e^{jm_k} W_{m-m_k} e^{-jm_l} W_{m+m_l}^* \quad (3.15)$$

$$P_m^- = \sum_{k,l=1}^K \frac{A_k A_l}{4} e^{-jm_k} W_{m+m_k} e^{jm_l} W_{m-m_l}^* \quad (3.16)$$

$$P_m^{+/-} = \sum_{k,l=1}^K \frac{A_k A_l}{4} \left[e^{jm_k} W_{m-m_k} e^{jm_l} W_{m+m_l}^* + e^{-jm_k} W_{m+m_k} e^{-jm_l} W_{m-m_l}^* \right] \quad (3.17)$$

Each of these contains a term representing mutual interference of spectral components due to different ion packets. For instance, P_m^+ is:

$$P_m^+ = \sum_k \frac{A_k^2}{4} |W_{m-m_k}|^2 + \sum_{k,l=1, l>k}^K \frac{A_k A_l}{4} 2\Re \left[e^{jm_k} W_{m-m_k} e^{-jm_l} W_{m+m_l}^* \right] \quad (3.18)$$

where the first term represents the peak located at $m = m_k$ and with ion abundance $A = A_k$, whilst the second term represents a perturbation resulting in shifted peak maximum in the spectrum. The obtained result, Eq. 3.18, quantitatively describes the effect of spectral interference in magnitude-mode FT, Chapter 4.

Interference of spectral components shifts peak maxima in spectra from their expected locations. This introduces errors in frequency (m/z) and abundance measurements. The limiting case is when the detection time is sufficiently small so that the interference is only defined by the difference of initial phases of ions. Depending on the analyte and experimental setup, the phase difference can be negligible and thus lead to constructive interference only, Figure 3.5a, so that the apparent abundance is given by the sum of abundances, and the

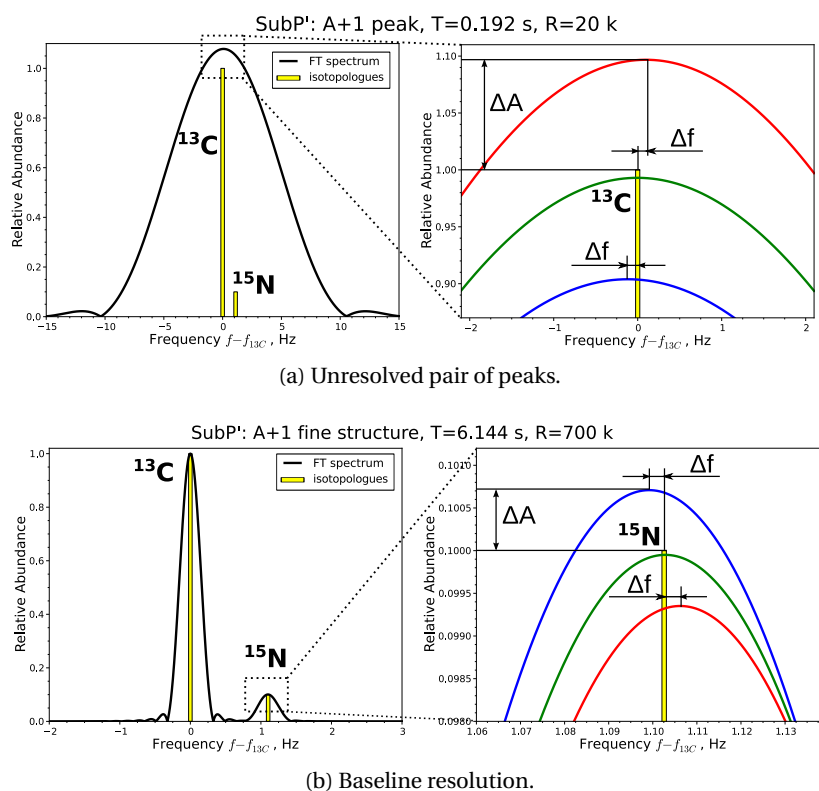


Figure 3.5: Frequency and abundance errors due to interference of spectral components in a numerical experiment using A+1 isotopic fine structure of substance P. Transient signals were numerically simulated with parameters corresponding to ^{13}C and ^{15}N isotopologues of doubly protonated peptide substance P and 10 T FT-ICR MS instrument. The case of unresolved pair of peaks corresponds to a transient length of 0.192 s (equivalent resolving power 20'000). The case of baseline resolved pair of peaks corresponds to a transient length of 6.144 s (equivalent resolving power 700'000).

apparent frequency is given as a weighted average. Intermediate cases around the baseline resolution threshold, Figure 3.5b, are such that the effect of spectral interference allows a linear approximation where the measured values vary harmonically as functions of their frequency difference, phase difference, and the detection time. Finally, the effect of spectral interference decreases when detection period is sufficiently large so that significantly excessive baseline resolution is achieved.

Analytical estimations for the spectral errors have been obtained and validated. These errors are systematic and defined by detection time, frequency and initial phase differences of interfering signals, relative abundances and apodization window. Spectral interference in the 1st order approximation shows harmonic dependence of the errors on the detection time or equivalently, difference of the initial phases, Figure 3.6a, Figure 3.6b.

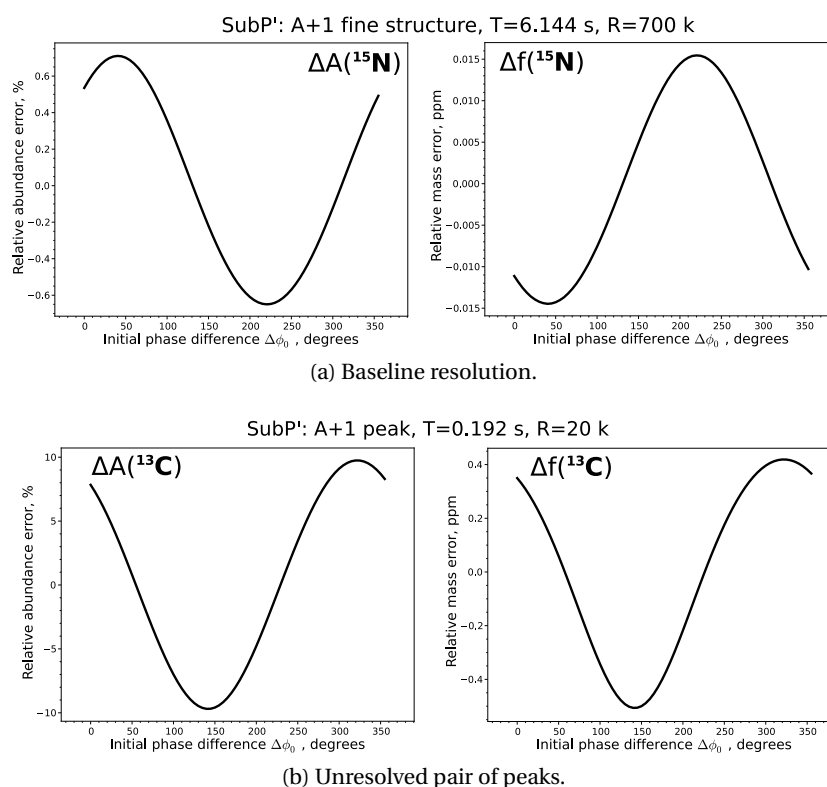


Figure 3.6: Errors due to spectral discreteness as a function of initial phase difference in numerical experiments using A+1 isotopic fine structure of substance P. Transient signals were numerically simulated with parameters corresponding to ^{13}C and ^{15}N isotopologues of doubly protonated peptide substance P and 10 T FT-ICR MS instrument. The case of unresolved pair of peaks corresponds to a transient length of 0.192 s (equivalent resolving power 20'000). The case of baseline resolved pair of peaks corresponds to a transient length of 6.144 s (equivalent resolving power 700'000).

3.1.3 Correlated errors

As the next stage after DFT processing with a given apodization function, a given number of zero-fillings, and a given type of spectral representation, the signal processing workflow includes a peak picking procedure, which is aimed to reduce the detrimental effect of spectral discreteness on measured frequency (m/z) and abundance values, as exemplified in Figure 3.7.

For instance, the standard signal processing workflow includes a peak-picking procedure based on parabolic approximation of three spectral points around a local maximum, Figure 3.7. Here we analyze errors remaining after the peak picking procedure.

In Paper IV discussed in Chapter 4, *vide infra*, the frequency distributions obtained for a set of 1000 ICR transients of isolated monoisotopic ion of peptide MRFA (524 m/z), Figure 4.6, top panel, with the standard FT signal processing workflow on one hand and with the least-squares fitting of transients, on the other, differ in their mean values, as an experimental

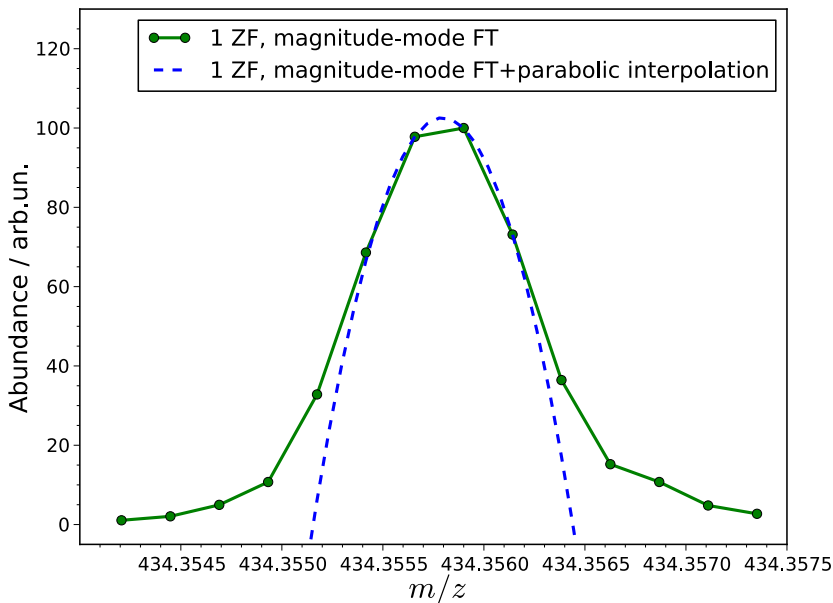


Figure 3.7: Illustration of three-points parabolic interpolation in the analysis of resin crude oil sample on Orbitrap Elite FTMS.

demonstration of the errors in question. In that study, additional numerical experiments showed that the LSF frequency distribution is unbiased, whereas FT processing produced shifted frequency values, as discussed in Chapter 4. To identify the exact numerical cause of this bias, the following numerical experiment was performed. Recall that detection time T and single zero filling result in frequency spacing of $\Delta_d f = 1/(2T)$ in the discrete Fourier spectrum. Specifically, the detection time of those experimental transients was $T = 24$ ms and hence $\Delta_d f \approx 21$ Hz. Additionally, the full width of a spectral peak of interest, which is defined by the Hann apodization function and magnitude-mode spectral representation in our case, is $\Delta_{\text{wnd}} f = 4/T$. Therefore, in the sampled spectrum, the peak of interest has $\Delta_{\text{wnd}} f / \Delta_d f = 4 \cdot 2 = 8$ 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. Therefore, 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 $f_0 \leq f_n \leq f_0 + \Delta_d f$, where $f_0 = 292941.44$ Hz is the mean frequency obtained with LSF processing of the transients. This way, the values $\Delta_{\text{wnd}} f$ were all 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 numerical transients, the frequencies of peak maxima, f , measured using the FT signal processing workflow, deviated from the corresponding true values f_n as the peak shape relocates over the eight frequency points, which in turn are fixed, Figure 3.8.

That is, since the three-point interpolation is only a second-order approximation of the actual

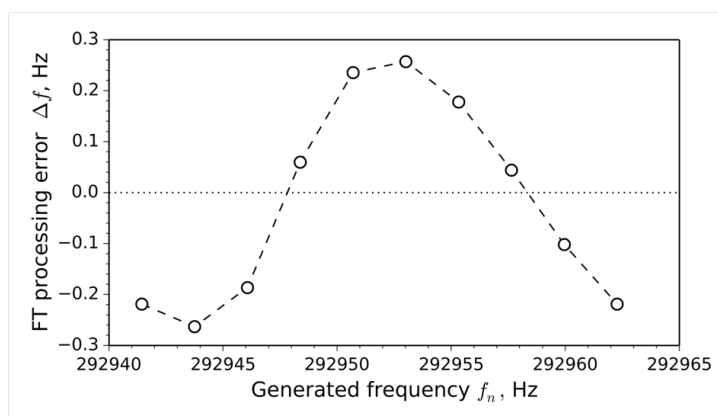


Figure 3.8: Correlation between the measurement error of FT signal processing and the true frequency in numerical simulations of frequency measurements with the standard FT workflow including Hann apodization window, single zero-filing, and three points parabolic interpolation. Out of $N = 10$ points on the plot, the n^{th} point, $n = 0, \dots, N - 1$, shows the analysis of the transient signal with frequency $f_n = f_0 + n\Delta f / (N - 1)$, where $\Delta f = 1 / (2T)$ is the frequency spacing in the discrete Fourier spectra of these signals. The following parameters were used: length $T = 24$ ms, frequency $f_0 = 292,941.44$ Hz (the leftmost point on the plot), amplitude $A = 0.405$, and phase $\varphi = -86.70$ degrees. Additional results obtained for the other phase $\varphi = -80.50$ degrees did not differ noticeably with respect to the effect under consideration (data not shown). Figure adapted from Paper IV.

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 , Figure 3.8. Thus, the frequency bias in question, Figure 4.6, top panel, demonstrates the limited accuracy of the peak interpolation procedure of the standard FT signal processing workflow, while the errors in question Δf are correlated with the true frequency f_n . Additionally, 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 4.6, top panel: since the full width $\Delta_{\text{wnd}}f$ is inversely proportional to the transient's length, T , the correlated error of frequency reduces as the time T increases from 24 ms to 96 ms.

3.2 Noise influence

The next considered part of the measurement scheme is the signal conditioning and signal digitization units. Here, thermal and digital noise are introduced to a digitized transient signal, thus influencing Eqs. 3.6 and 3.9. In result, peak maxima in mass spectra can be shifted from their expected positions predicated by Eqs. 3.6 and 3.9, hence causing frequency (m/z) and abundance measurement errors, Figure 3.9.

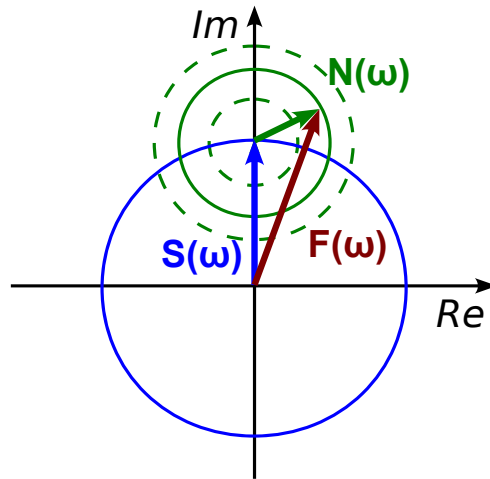


Figure 3.9: Illustration of the additive model of noise in frequency domain: $F(\omega) = S(\omega) + N(\omega)$, where F , S , and N are, respectively noisy spectrum of signal, noise-free spectrum of signal, and noise spectrum.

3.2.1 Logarithmic distributions

In the context of noisy transient signals, Paper II describes a method for noise thresholding aimed to maximize the range of assignable analytes in complex mass spectra, including analysis of crude oil fraction on Orbitrap Elite FTMS.[68] The method is based on logarithmic intensity plots and developed for determination of the noise level in mass spectral of complex samples such as crude oils and top-down MS spectra. Generally, the method is not included in the present thesis, but the noise theory developed in Paper II, Appendix is of particular interest here.

Specifically, in Paper II, Appendix, for a discrete Fourier spectrum of a noisy signal we establish a relationship between the standard deviation of the distribution of spectral amplitudes of noise in the spectrum and the local maximum of the distribution of logarithmic spectral amplitudes, Figure 3.10. The problem is solved separately for the absorption-mode and magnitude-mode FT spectral representations. For the sake of brevity, apodization and zero-fillings of the transient signal are not considered and the derivations are given in terms of continuous-variable functions.

First, let us find the distributions of the noise spectral amplitudes. Let the experimental transient signal be represented as:

$$f(t) = s(t) + n(t) \tag{3.19}$$

where $s(t)$ is the signal induced by the ions moving in the mass analyzer and $n(t)$ is a single realization of the noise, which is modeled as random noise with a Gaussian distribution. From here on, we may ignore the signal, viz., it is assumed that $s(t) \equiv 0$. Due to the linearity property of the Fourier transform, this assumption does not change the final result in the case

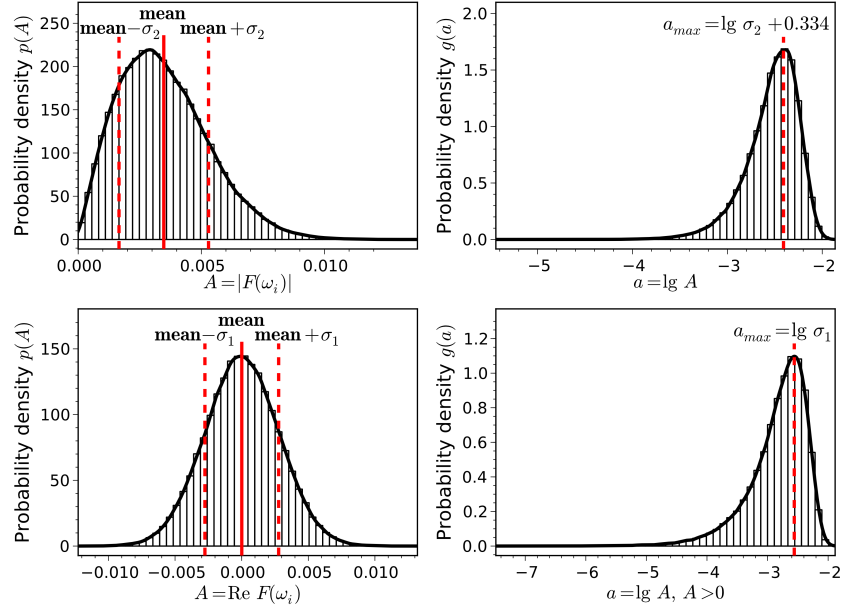


Figure 3.10: Relation between the distribution of spectral amplitudes and the distribution of logarithmic spectral amplitudes for noise-only Fourier spectra represented in magnitude (top) and absorption (bottom) modes. Left panels: distributions of spectral amplitudes. Right panels: distributions of logarithmic spectral amplitudes; the local maxima provide the standard deviation of interest, σ . Figure adapted from Paper II.

of absorption-mode spectral representation; for the magnitude mode the result will be an approximation with reasonable accuracy due to non-linear spectral interference discussed previously in this chapter.

As follows from Eq. 3.1, the Fourier transform of $f(t)$ can be then given as:

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-j\omega t} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} n(t) e^{-j\omega t} d\omega \quad (3.20)$$

We introduce the random variables $A = \Re[F(\omega)]$ and $A = |F(\omega)|$, Figure 3.10, left panels, for the cases of the absorption-mode and magnitude-mode spectral representations, respectively. Let $p(A)$ denote the probability density function for the corresponding variable A . In the case of absorption mode, $p(A)$ is the Gaussian distribution with zero mean value and some standard deviation σ_1 :

$$p(A) = k_1 \exp\left(-\frac{A^2}{2\sigma_1^2}\right) \quad (3.21)$$

In turn, the variable $A = |F(\omega)|$ is the square root of the sum of squares of two variables having

Chapter 3. Data analysis

the same distribution, Eq. 3.21:

$$A = |F(\omega)| = \sqrt{\Re^2[F(\omega)] + \Im^2[F(\omega)]} \quad (3.22)$$

Therefore, for magnitude mode, the function $p(A)$ is the Rayleigh distribution (χ -distribution with two degrees of freedom):

$$p(A) = k_2 \frac{A}{\sigma_1} \exp\left(-\frac{A^2}{2\sigma_1^2}\right) \quad (3.23)$$

with the following standard deviation expressed in terms of the standard deviation of the distribution Eq. 3.21:

$$\sigma_2 = \sigma_1 \sqrt{2 - 2\Gamma^2(3/2)} \approx 0.655\sigma_1 \quad (3.24)$$

where $\Gamma(x)$ is the gamma function.

Now, consider the following change of variables:

$$a = \lg A, A > 0 \quad (3.25)$$

$$a = \lg A \quad (3.26)$$

for absorption and magnitude modes, respectively. Let us find the corresponding probability density functions $g(a)$ and their local maxima, Figure 3.10, right panels. From the change of variables theorem[26] and Eqs. 3.21, 3.23, and 3.24 it follows that both functions $g(a)$ can be written as

$$g(A) = p(A) \frac{dA}{da} = p(A) \frac{A}{\lg e} \quad (3.27)$$

where $p(A)$ is given by Eq. 3.21, or Eq. 3.23, depending on the spectral representation. The following condition defines the points a_{\max} of local maxima of the functions $g(a)$:

$$\left. \frac{dg}{da} \right|_{a^*=a_{\max}} = \left. \frac{dg}{d \lg A} \right|_{\lg A^*=a_{\max}} = 0 \quad (3.28)$$

which can be reduced via equivalent transformations:

$$\left. \frac{dg}{d \lg A} \right|_{\lg A^*=a_{\max}} = \frac{1}{\lg^2 e} \left(\frac{d(pA)}{dA} \frac{dA}{d \lg A} \right) \Big|_{\lg A^*=a_{\max}} = \frac{1}{\lg^2 e} \left[A \frac{dp}{dA} + p \right] \Big|_{\lg A^*=a_{\max}} \quad (3.29)$$

to the following form:

$$A^* \frac{dp(A^*)}{dA} + p(A^*) = 0 \quad (3.30)$$

Using Eqs. 3.21 and 3.23, we obtain the derivative dp/dA for the absorption and magnitude modes, respectively:

$$\frac{dp_1}{dA} = -\frac{A}{\sigma^2} k_1 \exp\left(-\frac{A^2}{2\sigma^2}\right) = -\frac{A}{\sigma^2} p_1 \quad (3.31)$$

$$\frac{dp_2}{dA} = -\left(\frac{A}{\sigma^2} + \frac{1}{A}\right) p_2 \quad (3.32)$$

From the change of variables Eqs. 3.25 and 3.26, and obtained derivatives Eqs. 3.31 and 3.32, we find the following solutions for Eq. 3.30 in case of absorption and magnitude modes, respectively:

$$\sigma_1 = A^* = 10^{a_{\max}} \quad (3.33)$$

$$\sigma_1 = \frac{1}{\sqrt{2}} A^* = \frac{1}{\sqrt{2}} 10^{a_{\max}} \quad (3.34)$$

Finally, rewriting Eq. 3.34 in terms of the standard deviation Eq. 3.24 of the magnitude-mode noise distribution gives:

$$\sigma_2 \simeq 0.463 A^* = 0.463 \times 10^{a_{\max}} \quad (3.35)$$

Thus, we have found the expressions of interest: Eq. 3.33 and Eq. 3.35. For convenience, they can be alternatively represented as follows:

$$a_{\max} = \lg \sigma_1 \quad (3.36)$$

$$a_{\max} \simeq \lg \sigma_2 + 0.334 \quad (3.37)$$

As can be seen from Figure 3.10, the point of local maximum of the noise distribution plotted for logarithmic intensity provides the standard deviation, σ , of noise. Importantly, the standard deviation can also be obtained in the case when the mass spectrum is acquired in the absorption mode with negative amplitudes cut out, i.e., when the conventional "N sigma" approach is not directly applicable. For that, the obtained threshold values can be recalculated into corresponding values of N via the relation between the standard deviation and the noise distribution maximum, depicted in Figure 3.10.

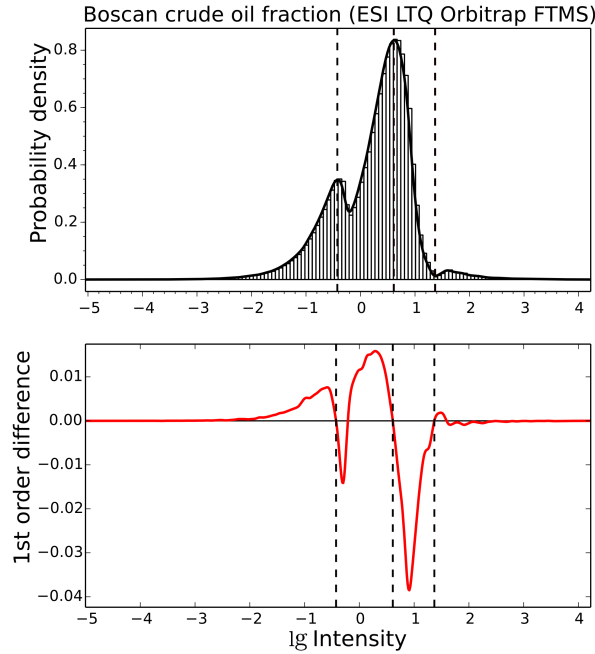


Figure 3.11: Noise analysis of a mass spectrum acquired for boscan crude oil fraction on Orbitrap Elite FTMS. Top panel: The probability density function was plotted for logarithmic intensities from the mass spectrum. Bottom panel: Local minima and maxima of interest are automatically determined using 1st order difference functional applied to the probability density function.

3.2.2 Effect of digital noise

Additionally, effect of digital noise resulting from quantization of digital data should be quantitatively described to prevent introducing excessive amounts of digital noise when a digital transient signal is acquired. Specifically, in the context of the logarithmic plots, Figure 3.11 shows the noise level analysis of a crude oil fraction's mass spectrum acquired on Orbitrap FTMS. Apart from the local minimum, which is employed for noise thresholding purposes,[68] the plot reveals two types of noise distributions. Since the points of local maxima provide the corresponding standard deviations, the two maxima in this plot may be assigned to two separate noise distributions with different standard deviations. The noise distribution in the higher intensity range is considered to be due to the thermal noise of the detection circuit, whereas the noise distribution in the lower intensity range is presumably due to the quantization effects behind the digital representation of the experimental data.

Another source of digital noise is due to quantized levels of a digital transient signal. An estimate of the equivalent standard deviation, σ_d , is known in digital signal processing as:

$$\sigma_d = \frac{q}{\sqrt{12}} \quad (3.38)$$

Here, q is the least significant bit voltage, i.e. $q = \Delta V / 2^n$, where ΔV and n are the full scale voltage range and number of bits of an ADC employed. In the context of FTMS measurements, when the full scale voltage range of the ADC is matched to the voltage range of the transient signal going from the signal conditioning unit, the digital noise in question is minimized. On the other hand, when the signal range is much narrower than the full scale range of the ADC, significant amounts of digital noise can be added to the digitized signal. In terms of the present work, these considerations are taken into account in the context of signal processing methods, Chapter 4, and instrumentation, Chapter 5, including when an external data acquisition system is employed.

3.3 Mass calibration function

The final element of the measurement system considered in this chapter is the mass analyzer. The mass analyzer defines the relation between the ion mass-to-charge ratios and abundances, on one hand, and the observed frequencies of ion motion and ion signal intensities, on the other. Hence, an error with which this relation is known contributes, among others, to the overall accuracies of mass and abundance measurements. Of these two, the mass accuracy as one of the key analytical characteristics for qualitative analysis of samples.

In Paper III, we developed a method for improved mass calibration aimed at MS analysis of complex petroleomic samples on Orbitrap FTMS. Let us start with the origin of the measurement errors in question. For a given mass analyzer, the mass measurement principle may be generally written via the mapping F from the mass-to-charge ratios m/z and other relevant physical quantities p_i , $1 \leq i \leq m$ (such as ion abundances, ion trajectories, etc.), to the corresponding ion motion frequencies f :

$$f = F(m/z, p_1, \dots, p_m) \quad (3.39)$$

To obtain the mass-to-charge ratios, Eq. 3.39 may be solved for m/z , thus providing the mass calibration function M_0 : $m/z = M_0(f, p_1, \dots, p_m)$. The following theoretical forms of the mass calibration function were previously derived for the model case of a single ion moving in: (i) a uniform magnetic field and an electric field with quadratic potential of a Penning trap; [69, 61] and (ii) an electric field with quadro-logarithmic potential of an orbitrap: [8]

$$M_{\text{theor}}(f) = \frac{C_1}{f} + \frac{C_2}{f^2} \quad (3.40)$$

$$M_{\text{theor}}(f) = \frac{C_2}{f^2} \quad (3.41)$$

where C_1 and C_2 are constants defined by magnetic and electric fields, respectively, for a given mass analyzer. However, due to various unaccounted effects, as discussed in Chapter 2, the fields may be perturbed such that the theoretical mass calibration functions, Eqs. 3.40 and 3.41,

do not provide sufficient mass accuracy even for routine applications. As such, conventional approaches for mass calibration in FTMS introduce additional empirical features so that the standard mass calibration functions utilized in FT-ICR MS and Orbitrap FTMS may be written via the following, or similar, *semi-theoretical* forms, respectively:

$$M_{\text{theor}}(f) = \frac{C_1(Q)}{f} + \frac{C_2(Q)}{f^2} \quad (3.42)$$

$$M_{\text{theor}}(f) = \frac{C_2(Q)}{f^2} + \frac{C_3(Q)}{f^3} \quad (3.43)$$

where Q represents the total charge of all ions in the trap.

The default way to obtain an estimate of the mass calibration function M_0 is to evaluate the semi-theoretical mass calibration functions, Eqs. 3.42 and 3.43, on the basis of calibrants analyzed in separate measurements ("external" calibration). In this case, the systematic errors in the observed mass-to-charge ratios, $(m/z)_{\text{obs}}$, are governed by the accuracy of the obtained estimate $M_{\text{def}}(f)$ relative to the actual mass calibration function M_0 :

$$(m/z)_{\text{obs}} = M_{\text{def}}(f) \quad (3.44)$$

With implementation of the automatic gain control (AGC)[70, 71] system to control the ion population in the mass analyzer, this calibration approach allows one to achieve mass accuracies on the order of 1 ppm in routine applications.[72, 73] Nevertheless, such performance is insufficient for analysis of complex samples such as crude oil fractions.

Hence, to reduce the systematic errors in mass measurements, an estimate of the mass calibration function with improved accuracy is required. For that, the semi-theoretical mass calibration functions, Eq. 3.42 and 3.43, may be evaluated on the basis of calibrants present within a given mass spectrum ("internal" calibration). However, to obtain more accurate estimates of the mass calibration function, further theoretical developments for the residual effects are required. For instance, among alternative mass calibration functions for FT-ICR MS,[74] the calibration functions incorporating ion abundances aim to provide the mass calibration function Eqs. 3.42 with a second-order correction for ion-ion interactions.[75, 76, 77, 78] These studies resulted in a number of calibration approaches implemented and evaluated for FT-ICR MS-based applications.[79, 80, 81, 82] In Orbitrap FTMS, importance of the residual effects has also been realized;[83] however, further studies are required since the current theories are not sufficiently comprehensive.

Hence, in the context of mass analysis of petroleomic samples, recalibration is performed on the basis of internal calibrants obtained from up to several prominent homologous series in the mass spectrum. However, for analysis of complex petroleomic samples by Orbitrap FTMS, high performance recalibration methods that routinely deliver suitable mass accuracies were not available. Thus, to enable comprehensive analysis of petroleum samples of with Orbitrap

FTMS, we developed a suitable method for mass spectra recalibration aimed to provide mass accuracy levels at hundreds ppb.

3.3.1 Recalibration method

Since no comprehensive theory for residual errors of the default mass calibration in Orbitrap FTMS is available, the developed method was designed to empirically find an estimate M_{est} of the mass calibration function M_0 . The estimation is performed on the basis of mass-to-charge ratios and abundances of internal calibrants. In this method we consider only monoisotopic species as internal calibrants in a mass spectrum since m/z values and abundances measured for non-monoisotopic species may be additionally shifted due to fine structure interference, as discussed in previously in this chapter.

It is worth noting that if the number of employed calibrants is insufficient to maximize the accuracy of M_{est} , then the potential of experimental data is not entirely released. Therefore, to estimate the mass calibration function with maximized accuracy, we implement an iterative approach of assignment followed by recalibration. In each iterative step, (i) assignment of elemental compositions is performed anew within the limits of available mass accuracy for the mass spectrum obtained in the previous iteration (in the case of the first iteration, the assignment is performed for the initial mass spectrum), (ii) then, an estimate of the mass calibration function is obtained via evaluation of the mass error function using all of the assigned monoisotopic peaks as internal calibrants, (iii) and finally, the initial mass spectrum is recalibrated using the obtained estimate of the mass calibration function.

In doing so, in each iterative step, the mass calibration function's estimate is obtained with increased accuracy, which in turn provides an increased number of calibrants, as peak assignment is limited by mass accuracy constraints on each iteration. Finally, the accuracy of the estimate reaches its limit since random errors become prevailing in the mass errors of the remaining monoisotopic peaks. Thus, after the final iteration, the accuracy of the estimate of the mass calibration function is maximized. The recalibrated mass spectrum is then available for a full-scale MS analysis consisting of assignment of elemental compositions. The method was implemented using the functionality of the pyFTMS software framework developed in-house.

3.3.2 Performance evaluation

Performance of the method was evaluated in a number of samples of crude oil fractions. For instance, in the analysis of the resin fraction from a Nigerian crude oil, 2804 elemental compositions were identified,[68] the assigned analytes belong to 10 heteroatom classes (N, NO, N₂, NO₂, OS, NS, N₂O, O, O₂, HC). Their monoisotopic peaks span a range of ion abundances of 2520:1 and the m/z range of $200.1 < m/z < 911.8$. The total number of peaks within this m/z range in the mass spectrum amounted to 5424.

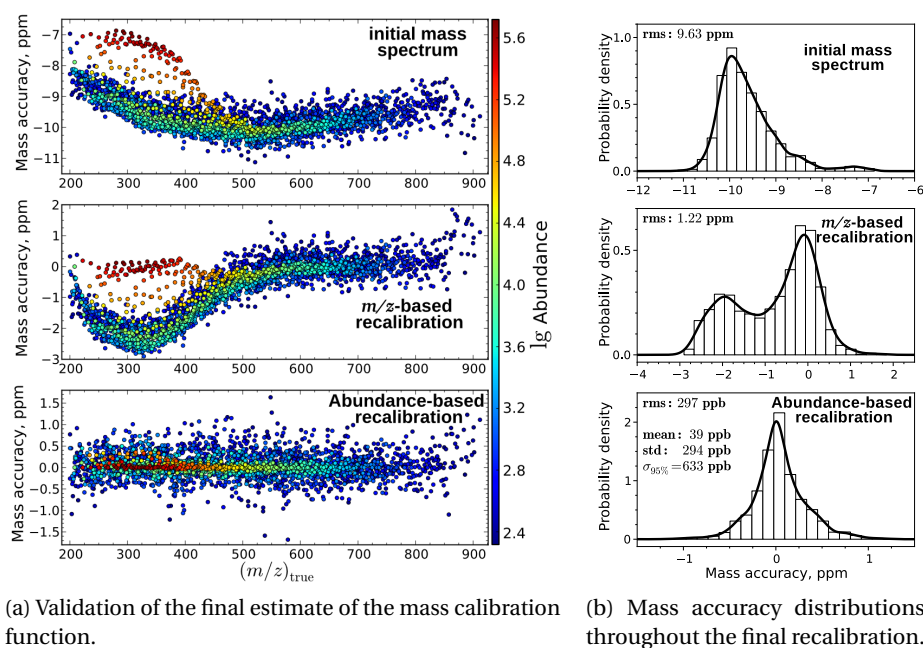


Figure 3.12: Recalibration method in the analysis of resin crude oil fraction on Orbitrap Elite FTMS. Figure adapted from Paper III.

Importantly, once the elemental compositions are assigned, we are able to track the location of each individual peak back to all stages and iterations of the recalibration. Thus, to assess the method performance, we trace how the mass accuracy for all identified monoisotopic peaks is changed as a function of true m/z and experimental ion abundances at three key stages of the recalibration: the initial mass spectrum (with default mass calibration); the mass spectrum after m/z -based recalibration; and the mass spectrum after abundance-based recalibration, as illustrated in Figure 3.12a. Figure 3.12a, top panel, shows mass accuracies for the initial mass spectrum. In this particular case, there was a significant, 7-11 ppm, mass error. Also, a noticeable difference in mass accuracy spread can be observed for peaks of various abundances. Next, the result of the intermediate m/z -based recalibration is shown in Figure 3.12a, middle panel. Overall, the mass accuracy is improved, while the abundance-dependent spread of approx. 3 ppm in mass accuracy remains since this intermediate step is based on high-abundance analyte peaks only. Finally, the abundance-based recalibration corrects the abundance-dependent spread in mass accuracies, Figure 3.12a.

Figure 3.12b compares the mass accuracy distributions corresponding to the three recalibration stages outlined above. Figure 3.12b, top panel, shows the mass accuracy distribution for the initial mass spectrum, with the root-mean-square (RMS) mass accuracy of 9.63 ppm, the mean of -9.61 ppm, and the standard deviation of 0.625 ppm. Notably, the distribution has a small sidelobe, which is due to high-abundance monoisotopic peaks, as follows from Figure 3.12a, top panel. The intermediate m/z -based recalibration improved the RMS mass accuracy to 1.22 ppm, Figure 3.12b, middle panel. However, an additional maximum in the

distribution appeared, which is due to the abundance correction not being taken into account, resulting in low-abundance peaks having significant deviations in mass-to-charge ratios in regions of the mass spectrum where m/z -only calibration was performed on high-abundance peaks, mainly in the 200-500 m/z range, as follows from Figure 3.12a, middle. It is likely that the middle abundance analytes represent the calibration law that could be reasonable in other, less demanding experiments, without taking into account individual ion abundances; the high abundance peaks in this case must be the result of increased electric forces acting on these ions and thus shifting them from the curve of middle-abundance peaks. Finally, the abundance-based recalibration provided the mass accuracy distribution, Figure 3.12b, bottom panel, with the RMS mass accuracy of 0.297 ppm, the mean of 0.039 ppm, the standard deviation of 0.294 ppm, and with 95% of the monoisotopic peaks being within the 0.633 ppm range away from the mean. Thus, the recalibration shows efficient elimination of both m/z - and abundance-dependent systematic shifts in mass accuracies for the tested set of all monoisotopic peaks assigned. The residual errors are limited partially by the limited scope of the parametrization employed in the method and largely by the SNR of experimental data. Hence, the method maximizes the range of abundances of assignable species for a given signal-to-noise ratio of experimental data. The method routinely provides RMS mass accuracies at the level of hundreds ppb, typically approx. 300 ppb, for analysis of petroleomic samples of relatively high complexity, e.g., at least up to and including maltene fractions, with the high-field orbitrap mass analyzer.

As an analytical case for future studies in FTMS, it would be useful to test whether the random-errors indicator developed in earlier years of FTMS will be consistent or not with the residual, random-like distribution of mass errors.[84]

4 Signal processing

Based on the considerations developed in Chapter 2, the current chapter presents advanced methodologies of signal processing aimed mainly at improved resolution performance for FTMS. Such characteristics as mass and abundance measurement accuracies are also involved since the two groups introduced by the uncertainty principle of measurements are closely inter-related. In the context of increased resolution performance, this chapter is focused on signal processing only, while the developments in instrumentation are discussed in Chapter 5.

As follows from Chapter 2, provided that different ion packets in a mass analyzer are sufficiently coherent during the data acquisition event so that the difference of total phases accumulated by ion packets corresponding to two close m/z values of interest is greater than the developed phase spread of those ion packets, it is the uncertainty principle of the standard signal processing that defines the resolution performance of the mass spectrometer. As such, this chapter describes a number of implemented signal processing methods with improved resolution performance, as well as reports their use in applications.

In Section 4.1, our consideration is limited specifically to FT-based signal processing methods. Here, the implications of phase intersection points of ions trapped in a mass analyzer are derived for magnitude- and absorption mode FT and generalized to the case of analytes with a regular distribution of peaks, hence resulting in a signal processing method providing two-fold resolution performance compared to absorption-mode FT. Next, super-resolution methods (SR) for signal processing are considered. In Section 4.2, implementation of the least-squares fitting (LSF) method for signal processing in FTMS is described and employed for applications in instrument development and for improved quantitative analysis in proteomics. In turn, Section 4.3 is focused on the filter-diagonalization method (FDM), where the uncertainty principle of FDM is characterized along with discussion of other super-resolution methods. Additionally, the concept of signal-to-noise (SNR)-limited resolution is introduced. Advantages for analytical applications are considered.

4.1 FT and phase intersection points

Advanced signal processing has recently gained a particular attention in FTMS community. Fourier transform (FT) with absorption-mode spectral representation is one of the striking examples 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. 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.[85, 86, 87] Recently, the absorption mode has received a particular attention in FT-ICR MS method development and applications.[88, 89, 90, 91, 92]

In Orbitrap FTMS, the absorption mode is advantageous as well and has recently been implemented as a principal part of the algorithm known as enhanced FT (eFT).[93, 94] 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 is relatively easy in comparison to FT-ICR MS.[95]

Here we derive how the use of information on ion phases results in the corresponding quantitative form of the uncertainty principle of measurements of ion frequencies using the FT method for three different types of spectral representation, viz, magnitude mode, absorption mode, and the new FT based method with double phase correction. Specifically, we compare the uncertainty principle for situations when the phase information is not taken into account at all (magnitude mode FT), a single phase intersection point of analyzed ions is employed (absorption mode FT), and when the case of equidistant phase intersection points takes place for selected frequency windows with regularly distributed analytes in a broadband mass spectrum. Based on the latter, we develop a method of double phase correction, which, for a selected frequency window, provides a two-fold gain in resolution performance compared to absorption-mode FT. This method is tailored specifically to the case of regular m/z distributions of chemical species including doublets as a particular case, e.g. quantitative proteomics with TMT reagents.

4.1.1 No information on ion phases

As shown previously in Chapter 3, the magnitude of a discrete spectrum $|S_m|$ has an inherent non-linearity, which is the interference of close spectral components located at $m = \pm m_k$, $k = \overline{1, K}$. This effect is significant when the detection period is insufficient to provide excessive baseline resolution of the interfering spectral peaks in the $|S_m|$, Chapter 3. Here we focus on the situation when the detection time is large enough such that the spectral interference can

be neglected. In this case, the $|S_m|$ allows for linearization as follows:

$$|S_m| \simeq \sum_{k=1}^K \left[\frac{A_k}{2} |W_{m-m_k}| + \frac{A_k}{2} |W_{m+m_k}| \right], \quad m_k = f_k N / f_s, \quad m = \overline{0, N-1} \quad (4.1)$$

where W_m is given by Eq. 3.6. Notice that the magnitude-mode FT does not employ the phase intersection points of the analyzed ions. We thus denote the obtained magnitude spectrum as $S_m^{(0)}$:

$$S_m^{(0)} \simeq \sum_{k=1}^K \left[\frac{A_k}{2} W_{m-m_k}^{(0)} + \frac{A_k}{2} W_{m+m_k}^{(0)} \right], \quad m_k = f_k N / f_s, \quad m = \overline{0, N-1} \quad (4.2)$$

where $W_m^{(0)}$ represents the envelope of an ion signal:

$$W_m^{(0)} = \left| \frac{\sin(\pi m)}{\sin(\pi m / N)} \right|, \quad m \in \mathcal{Z} \quad (4.3)$$

Thus, we obtained the uncertainty principle for frequency measurements when magnitude-mode FT is applied, Eq. 4.3. Next, we will derive the corresponding uncertainty principles for absorption mode FT, and based on these results the method for double phase correction will be developed with accordingly weakened uncertainty principle.

4.1.2 Single phase intersection point

The single phase intersection point of analyzed ions is related to absorption mode spectral representation. Those FT mass analyzers that implement ion excitation by means of ion injection, e.g. the orbitrap, have an inherent feature that the initial phases of trapped ions are a linear function of their frequencies, neglecting higher-order deviations due to instrumental limitations.[96] Equivalently, due to ion excitation by ion injection, there exists a phase intersection point of the ions, i.e., a time point at which the total phases of ions intersect. For instance, Figure 4.1 illustrates the phase intersection point revealed in the analysis of ultramark mixture on Q Exactive Orbitrap FTMS.

To continue the analysis of the resolution performance of FT, here we derive the uncertainty principle of FT with absorption-mode spectral representation. Let us denote the phase intersection in question by φ_0 . Let τ_1 denote the difference between the time when the first point of the transient signal is sampled and the time at which the phase intersection point is located: $\varphi_0 + 2\pi f_k \tau_1 = \varphi_k$, where φ_k is the initial phase of the k^{th} sinusoidal component. The corresponding difference in samples of the discrete transient signal s_n is $v = \tau_1 f_s$. For the purpose of current work, we consider in detail the case of linear phase functions and integer v . If $\tau_1 < 0$, to carry out the linear phase correction, it is sufficient to skip $|v|$ points in the beginning of the discrete signal s_n , whereas v zeros should be added to the beginning of the signal s_n if $\tau_1 > 0$ (sample-skipping and zero-padding, respectively). Implemented this way, linear phase correction leads to the following discrete Fourier spectrum of the zero-padded or

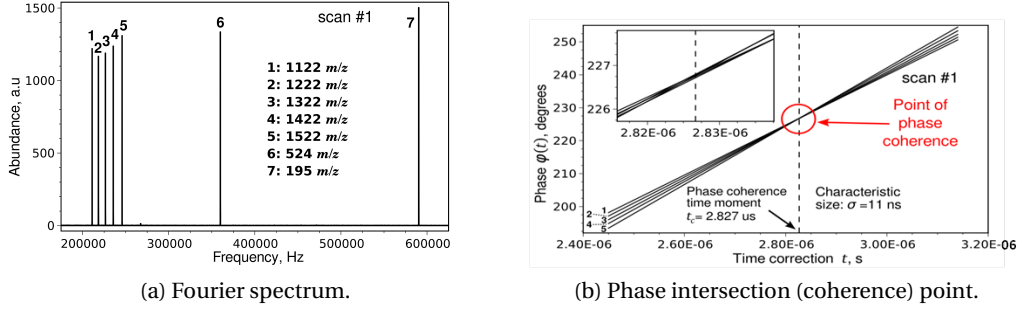


Figure 4.1: The phase coherence point for ultramark calibrants analyzed on Q Exactive Orbitrap FTMS. Using the implemented DAQ system, the transient signals were acquired at 100 MHz sampling frequency to make sure the influence of the analog (anti-aliasing) filter on the phase intersection point is negligible. The initial phases of ions were calculated using LSF processing of the 100 MHz transients. The transients were then FIR filtered and decimated down to the sampling frequency of 2 MHz. The phase intersection point was found using the least-squares solution of a system of linear phase equations written for each of the calibrants, taking into account linear phase shift due to FIR filtering. The constant (uncorrected) phase shift of approximately 227 degrees corresponds to the actual phase of zero in the point of phase intersection.

sample-skipped discrete transient:

$$S_m = \sum_{k=1}^K \left[\frac{A_k}{2} e^{j\varphi_0} W_{m-m_k} + \frac{A_k}{2} e^{-j\varphi_0} W_{m+m_k} \right], \quad (4.4)$$

$$m_k = f_k(N + \nu) / f_s, \quad m = \overline{0, N + \nu - 1}$$

where W_m is effectively the DFT of the rectangular window shifted by ν points in the positive direction of the time axis (or by $|\nu|$ in the negative direction) and calculated for the signal whose length was extended by ν (or reduced by $|\nu|$) points:

$$W_m = \frac{\sin(\pi m)}{\sin[\pi m / (N + \nu)]} \frac{e^{-j\pi m}}{e^{-j\pi m / (N + \nu)}}, \quad m \in \mathcal{I} \quad (4.5)$$

Now, all ions in the spectrum S_m , Eqs. 4.4 and 4.5, have the same initial phase φ_0 , which is the phase intersection point under consideration. Next, after the phase correction function is constructed for phase φ_0 :

$$\Phi_m = \varphi_0(H_m - H_{m+[(N+\nu)/2]+1}) - \varphi_0(H_{m+[(N+\nu)/2]} - H_{m+N+\nu}), \quad m = \overline{0, N + \nu - 1} \quad (4.6)$$

the phase corrected spectrum, which we denote as $S_m^{(1)}$ to point out the use of a single phase

intersection point, can be obtained as follows:

$$S_m^{(1)} = \Re(S_m e^{-j\Phi_m}) = \sum_{k=1}^K \left[\frac{A_k}{2} W_{m-m_k}^{(1)} + \frac{A_k}{2} W_{m+m_k}^{(1)} \right], \quad (4.7)$$

$$m_k = f_k(N + \nu) / f_s, \quad m = \overline{0, N + \nu - 1}$$

where $W_m^{(1)}$ represents the envelope of an ion signal and is the uncertainty principle for frequency measurements using the absorption-mode FT:

$$W_m^{(1)} = \frac{\sin(2\pi m)}{2 \sin[\pi m / (N + \nu)]} \cos[\pi m / (N + \nu)], \quad m \in \mathcal{Z} \quad (4.8)$$

The step with zero-padding or sample-skipping Eqs. 4.4 and 4.5 is relevant only in the context of linear phase functions and integer numbers ν . Alternatively, if ν significantly deviates from an integer number, the phase correction should be implemented immediately in the form of Eqs. 4.6, 4.7, and 4.8 with an appropriately chosen phase function. Zero-padding or sample-skipping are not employed in this case, i.e. $\tau_1 = 0$ and $\nu = 0$. The same procedure is required if the phase function has significant higher-order deviations, either due to some instrumental limitations (electrostatic traps with ion injection) or by design (FT-ICR MS), when the phase intersection point in question does not even exist, but can be obtained through the phase correction. Notably, in the situations when the phases of ions are not linear functions of ion frequencies, the circular replication property of the DFT, Chapter 3, leads to spectral artifacts in the absorption-mode expressions, Eqs. 4.7 and 4.8, explaining the effects of baseline roll or peak shape asymmetry reported previously for implementations of absorption-mode FT in FT-ICR MS.[91]

4.1.3 Multiple phase intersection points

Regular m/z distributions of chemical species, such as ^{13}C isotopic distributions of proteins or doublets in quantitative proteomics, $\Delta m/z = \text{const}$, cause the corresponding ion signals to be located in the Fourier spectrum equidistantly, $\Delta f \simeq \text{const}$. In turn, the equidistant frequency peaks $\Delta f \simeq \text{const}$ lead to equidistant points of phase intersection $\Delta t = 1/\Delta f = \text{const}$ for a frequency window containing these analytes. Hence, generalization of the single phase intersection with absorption-mode FT to the case of equidistant phase intersection points allows achieving 2-fold increase in resolution performance. Firstly, the phase correction similar to that of the absorption-mode FT is carried out. This procedure repeats the derivations from the previous subsection, Eqs. 4.4–4.8. Here we assume that the phase function of ions is linear. Moreover, the number ν is considered an integer. To take advantage of the equidistant phase intersection points, experimental deviations from these conditions must be sufficiently small. Hence, the high synchronization precision of the DAQ system described in Chapter 2 is of particular importance here.

Chapter 4. Signal processing

The obtained spectrum Eqs. 4.7 and 4.8 is subjected to the inverse DFT to obtain the corresponding discrete transient signal in which all components have zero initial phases:

$$s_n = \sum_{k=1}^K A_k \cos 2\pi f_k t_s n, \quad n = \overline{0, N + \nu - 1} \quad (4.9)$$

The transient signal is extended via even-symmetry reflection about the point $n = 0$:

$$s_n = \sum_{k=1}^K A_k \cos(2\pi f_k t_s n + \varphi_k), \quad n = \overline{0, 2(N + \nu) - 2}, \quad (4.10)$$

where the initial phases are given by the following expression:

$$\varphi_k = -2\pi f_k t_s (N + \nu - 1) \quad (4.11)$$

Next, the second phase correction is performed. This step requires the distribution's m/z spacing, $\Delta m/z$, which is assumed to be either known in advance or derivable when the distribution is slightly resolved in the absorption-mode or from different charge states otherwise. Using the value $\Delta m/z$ and the frequency-to- m/z calibration function of a particular mass analyzer, the corresponding frequency spacing, Δf , can be found. Using the frequency spacing Δf , the second phase intersection point is determined, (τ_2, φ) , where by τ_2 we denote the time difference between the first point of the transient signal and the time at which the second phase intersection point in question is located: $\varphi + 2\pi f_k \tau_2 = \varphi_k$. The corresponding difference in samples of the discrete transient signal s_n is $\mu = \tau_2 f_s$.

Finally, the linear phase correction is repeated to reveal the second phase intersection point:

$$\begin{aligned} S_m &= \sum_{k=1}^K \left[\frac{A_k}{2} e^{j\varphi} W_{m-m_k} + \frac{A_k}{2} e^{-j\varphi} W_{m+m_k} \right], \\ m_k &= f_k (2N + 2\nu + \mu - 1) / f_s, \quad m = \overline{0, 2(N + \nu) + \mu - 2} \end{aligned} \quad (4.12)$$

where W_m is given by:

$$W_m = \frac{\sin(\pi m)}{\sin[\pi m / (2N + 2\nu + \mu - 1)]} \frac{e^{-j\pi m}}{e^{-j\pi m / (2N + 2\nu + \mu - 1)}}, \quad m \in \mathcal{Z} \quad (4.13)$$

for which then the phase correction function is constructed:

$$\begin{aligned} \Phi_m &= \varphi(H_m - H_{m + [(2N + 2\nu + \mu - 1)/2] + 1}) - \varphi(H_{m + [(2N + 2\nu + \mu - 1)/2]} - H_{m + 2(N + \nu) + \mu}), \\ m &= \overline{0, 2(N + \nu) + \mu - 2} \end{aligned} \quad (4.14)$$

to obtain the double phase corrected spectrum, which is denoted as $S_m^{(2)}$ to point out the use of the property of equidistant phase intersection points:

$$S_m^{(2)} = \Re(S_m e^{-j\Phi_m}) = \sum_{k=1}^K \left[\frac{A_k}{2} W_{m-m_k}^{(2)} + \frac{A_k}{2} W_{m+m_k}^{(2)} \right], \quad (4.15)$$

$$m_k = f_k(2N + 2\nu + \mu - 1) / f_s, \quad m = \overline{0, 2(N + \nu) + \mu - 2}$$

where W_m represents the envelope of an ion signal and is the uncertainty principle for frequency measurements using the double phase correction FT:

$$W_m^{(2)} = \frac{\sin(2\pi m)}{2 \sin[\pi m / (2N + 2\nu + \mu - 1)]} \cos[\pi m / (2N + 2\nu + \mu - 1)], \quad m \in \mathcal{Z} \quad (4.16)$$

In conclusion, it is interesting to note that the presented methodology allows also the next step of generalization for the case when the multiple phase intersection points exist not only for an isolated frequency window of equidistant analytes, but also for a broadband frequency range containing many analytes with various m/z . Although this case is actually realized on the hardware level in this work (as an ICR working in the wide-band signal regime, Chapter 5), the close relation between the hardware level and the signal processing aspects is a remarkable consistency check.

4.1.4 Resolution performance comparison

The obtained peakshapes define the uncertainty principles for frequency measurements for each of the FT-based methods considered above. To compare the obtained uncertainty principles, Eqs. 4.3, 4.8, 4.16, we first reduce the obtained equations to the same frequency variable and then apply L'Hôpital's rule:

$$W^{(0)}(f) \simeq N \left| \frac{\sin(\pi f T)}{\pi f T} \right| \quad (4.17)$$

$$W^{(1)}(f) \simeq (N + \nu) \frac{\sin[2\pi f(T + \tau_1)]}{2\pi f(T + \tau_1)} \quad (4.18)$$

$$W^{(2)}(f) \simeq (2N + 2\nu + \mu - 1) \frac{\sin[4\pi f(T + \tau_1 + \tau_2/2)]}{4\pi f(T + \tau_1 + \tau_2/2)} \quad (4.19)$$

Eqs. 4.18 and 4.19 show the subsequent increase of the resolving power relative to the magnitude-

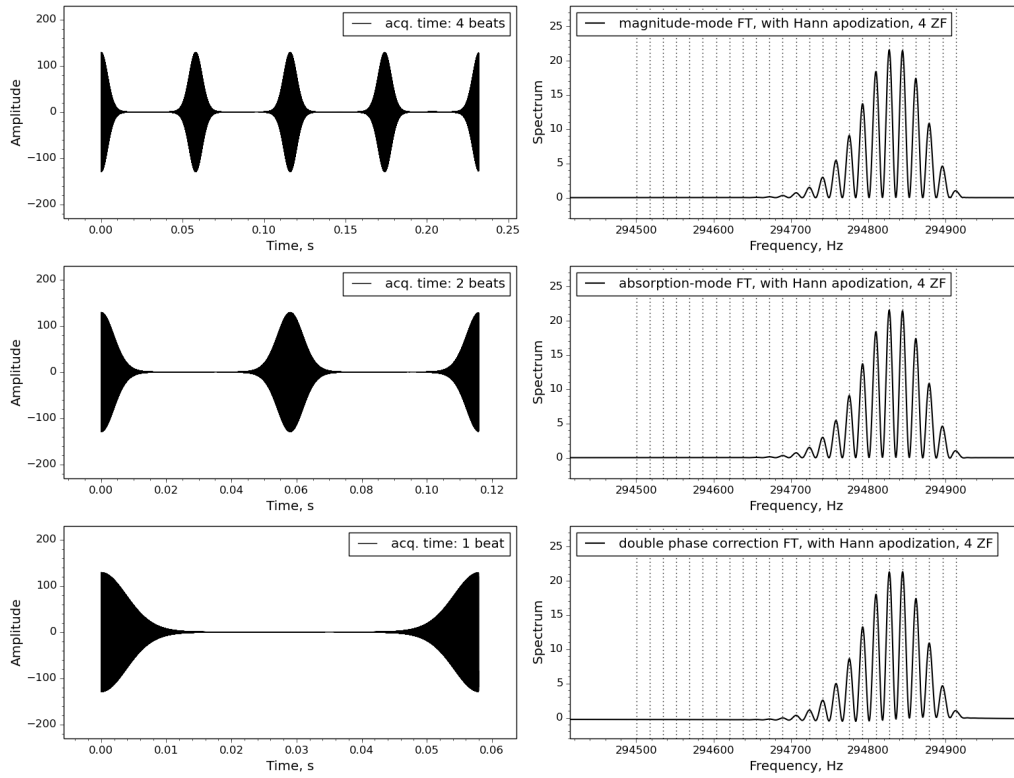


Figure 4.2: Illustration of baseline resolution and required detection period in analysis of a protein: magnitude-mode FT (top panels), absorption-mode FT (middle panels), and double phase correction FT (bottom panels). The initial transient signal was obtained numerically via de-noising of an experimental transient acquired on Q Exactive FTMS. For consistency, the Hann-type apodization windows were used for all the modes. Four zero-fillings were performed for improved visualization. For visual convenience, all transients were approximately aligned by their points of phase intersection. Normalization of the spectra for the number of points in the transient signals was additionally made.

mode FT, Eq. 4.17. Equivalently, comparisons can be made using the detection periods required to achieve baseline resolution of peaks:

$$T_{b/l}^{(0)} = 4/\Delta f \quad (4.20)$$

$$T_{b/l}^{(1)} = 2/\Delta f - \tau_1 \quad (4.21)$$

$$T_{b/l}^{(2)} = 1/\Delta f - \tau_1 - \tau_2/2 \quad (4.22)$$

$$(4.23)$$

Figure 4.2 compares performance of magnitude-mode FT, absorption-mode FT, and double phase correction FT methods. Effectively, the double phase correction FT method converts the

4.1. FT and phase intersection points

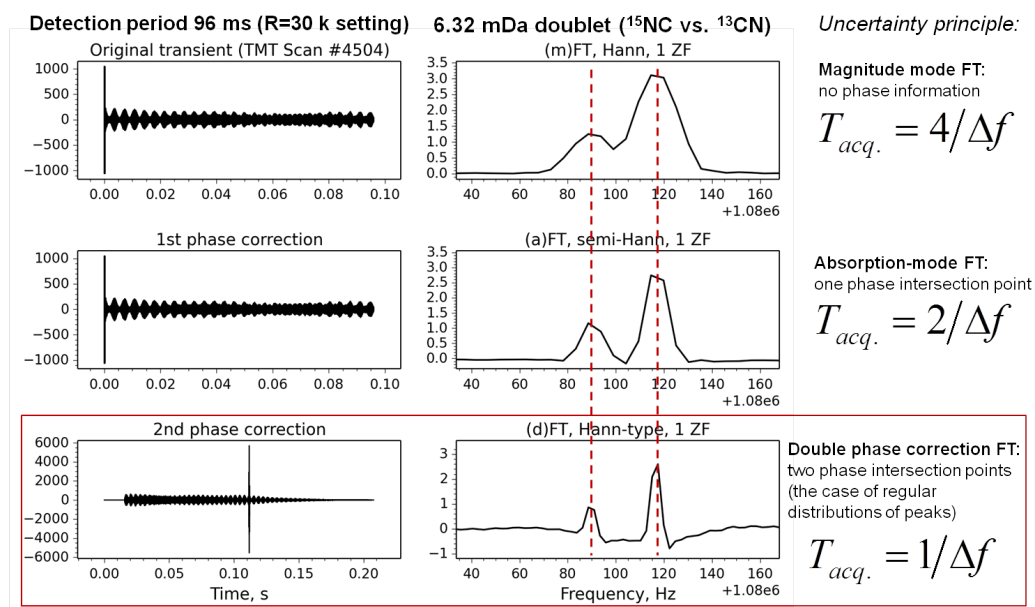


Figure 4.3: Illustration and performance comparison of magnitude-mode, absorption-mode, and double phase correction FT methods applied to a 6.32 mDa doublet of reporter ions in a transient signal acquired in the LC-MS analysis of 10-plex labeled yeast digest on Orbitrap Elite FTMS.

additional information on frequency spacing between the analytes into increased resolution performance, viz. doubled relative to that of absorption-mode FT. Since the knowledge of the m/z or frequency spacing is needed for the method to be applicable, its scope of applications includes quantitative proteomics.

Figure 4.3 illustrates application of the double phase correction method, along with magnitude mode FT and absorption mode FT, to a 6.32 mDa doublet of reporter ions in a transient signal acquired for LC-MS analysis of 10-plex labeled yeast digest on Orbitrap Elite FTMS, *vide infra*. The uncertainty principle changes in accordance with the obtained Eqs. 4.20-4.22. No post-processing baseline correction was applied so that the baseline roll in the double phase correction FT spectrum can be noted in accordance with the developed above theory. Indeed, the baseline roll resulted an a processing artifact when an integer number of points is used for (double) phase correction while a close but non-integer number of points would be required for this procedure to be exact. To correct for the baseline role in routine applications, corresponding methods from absorption-mode FT-ICR MS can be employed.[91]

Finally, for a shortened transient from the previous analysis, Figure 4.4 shows the advantage of the double phase correction FT method when the detection time is small so that both magnitude-mode FT and absorption-mode FT methods are unable to resolve the ions.

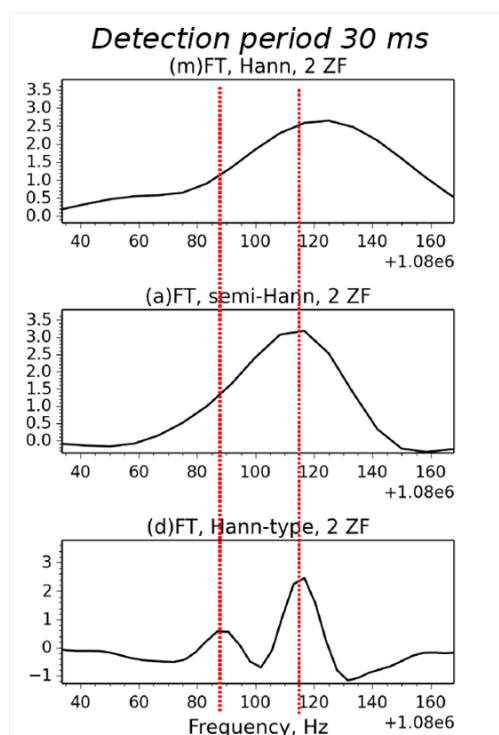


Figure 4.4: Advantage of the double phase correction FT method for short transients approximately equivalent to 1 isotopic beat of analytes in a frequency window of interest. The original, 96 ms-long transient was obtained in the analysis of 10-plex labeled yeast digest on Orbitrap Elite FTMS.

4.2 Least-squares fitting

Paper IV describes the implementation and characterization of the LSF method for analysis of transient signals in FTMS. In FTMS, non-FT signal processing methods 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), [97, 98, 99] linear prediction method (LPM), [100, 101, 102] and other methods. A number of questions, such as determination of ion phases for the double phase correction FT method, applications for instrument fine-tuning, and specific analytical applications such as quantitative proteomics, all require a high-performance LSF method for signal processing.

Although advantages of LSF-type methods have been named in patent applications [103, 104] as well as the LPM algorithm, an LSF-type method, has shown utility for frequency chasing of simulated FT-ICR MS transients [105], application of LSF methods for processing of experimental time-domain signals in FTMS was not previously reported in the peer-reviewed literature. Indeed, in the MS field the applications of LSF-type methods 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 data. However, the peak-shape fitting in the frequency

domain may imply reduced performance due to non-linear spectral interference effects, Chapter 3, and manual restriction of the fitted frequency range. Alternatively, fitting the raw data, viz. the transient signal, is justified as follows from the transient signal model, Chapter 2, represented as a sum of sinusoidal signals. Indeed, it is the theoretical form of the transient signal that enables measurement of frequencies, abundances, and phases of analyzed ions in FTMS. For instance, it is exactly due to this theoretical form that the ion's m/z is obtained from a corresponding peak maximum in a Fourier spectrum plotted in magnitude or absorption modes. Thus, the problem of signal analysis in FTMS is also consistent with the LSF method.

In the context of the extended physical model for ion motion described qualitatively in Chapter 2, 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 is not taken into account when an experimental transient is fitted using the LSF method. While in early FTMS instruments the signal damping effect could be described with an exponential function presumably resulting from ion-molecule collisions due to insufficient vacuum conditions,[106] in modern FTMS the signal damping effect is likely due to inharmonicity of the electric field in the mass analyzer. 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[107]. 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, Chapter 2. Nevertheless, regardless of the exact damping mechanism, the theoretical form with sinusoidal components is reasonable in the current work as the time scale of interest is sufficiently short such that ion decoherence does not develop significantly during signal acquisition.

4.2.1 Method implementation

In accordance with its spectral composition, Chapter 2, the transient signal, as a sequence of instant voltages s_n , Figure 4.5 (where $n = 0, \dots, N - 1$, and N is the total number of sampled points: $N = f_s T$, with f_s being the sampling frequency and T the detection time) was parameterized by the fitting function F_n defined as a sum of K sinusoidal components:

$$F_n = \sum_{k=1}^K A_k \sin(2\pi f_k n t_s + \varphi_k), n = 0, \dots, N - 1, \quad (4.24)$$

where t_s is the sample time, $t_s = 1/f_s$. The parameters to determine 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. The function subjected to minimization was the χ^2 function defined as:

$$\chi^2(A_1, \dots, A_K, f_1, \dots, f_K, \varphi_1, \dots, \varphi_K) = \sum_{n=0}^{N-1} (F_n - s_n)^2 \quad (4.25)$$

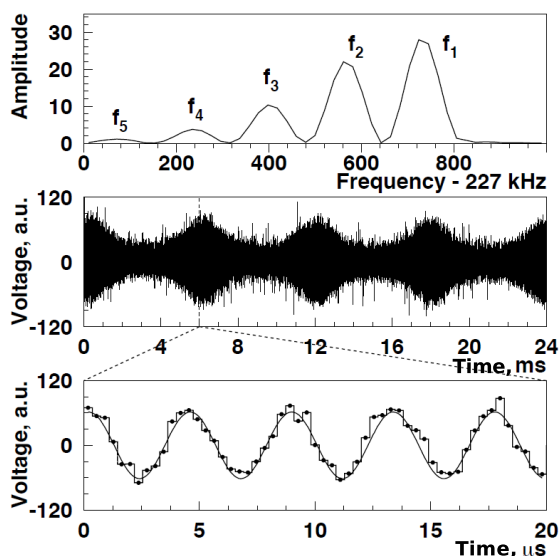


Figure 4.5: The principle of least-squares fitting (LSF) of transient signals in FTMS. Top panel shows the magnitude-mode Fourier spectrum of a 24 ms-long experimental transient signal shown in the middle panel. Bottom panel shows an expanded view of the transient with sampled points and a curve corresponding 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. Figure adapted from Paper IV.

For minimization purpose, the MINUIT package was used.[60] The parameters f_k , A_k , and φ_k are varied to locate the minimum deviation of the function, F_n , from the digitized transient signal, s_n , $n = 0, \dots, N - 1$, in terms of the minimum value of χ^2 :

$$\chi^2 \xrightarrow{\substack{A_1, \dots, A_K \\ f_1, \dots, f_K \\ \varphi_1, \dots, \varphi_K}} \chi_{\min}^2 \quad (4.26)$$

As seed values are required by the minimization procedure, a list of seeds for amplitudes, frequencies, and phases is submitted to the method when it initializes. For the purposes of the current work, these parameters were obtained using FT processing of transients. The method provides the set of solutions f_k , A_k , and φ_k , as well as the minimum value of χ_{\min}^2 .

4.2.2 Performance evaluation

Firstly, we evaluate the implemented method in the analysis of single- and multiple-component experimental and simulated ion cyclotron resonance (ICR) and Orbitrap FTMS transient signals. The MS experiments were performed on the 10 T FT-ICR MS and the Orbitrap Elite

FTMS. Ions were produced with the nESI ion source. The total charge injected into the ICR cell or the orbitrap was controlled with the automatic gain control (AGC) function. The target value for the total charge was set to the default setting of 2×10^5 . For the purpose of the current work, 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 experimental 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.

In numerical experiments, 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 the generated 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 $(\chi_{\min}^2/N)^{1/2}$.

Results of LSF analysis of transients were compared to those of FT processing. The transients were Fourier 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.

Figure 4.6 shows the analysis of 1000 single-scan ICR transients and 1000 single-scan Orbitrap transients containing the singlet. For the set of ICR transients, the amplitude, frequency, and phase distributions obtained with LSF and FT methods (except for phases) are shown in Figure 4.6, top panel. The corresponding mean values and standard deviations are listed in Table 1, top section, Paper IV. 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.

Hence, to verify which method provided more accurate values of ion frequencies, a numerical experiment with a set of 1000 modeled FT-ICR MS transient signals was performed Figure 4.6, middle panel. The simulated transients were generated using the mean amplitude, frequency, and initial phase values, as well as the noise standard deviation obtained from the LSF

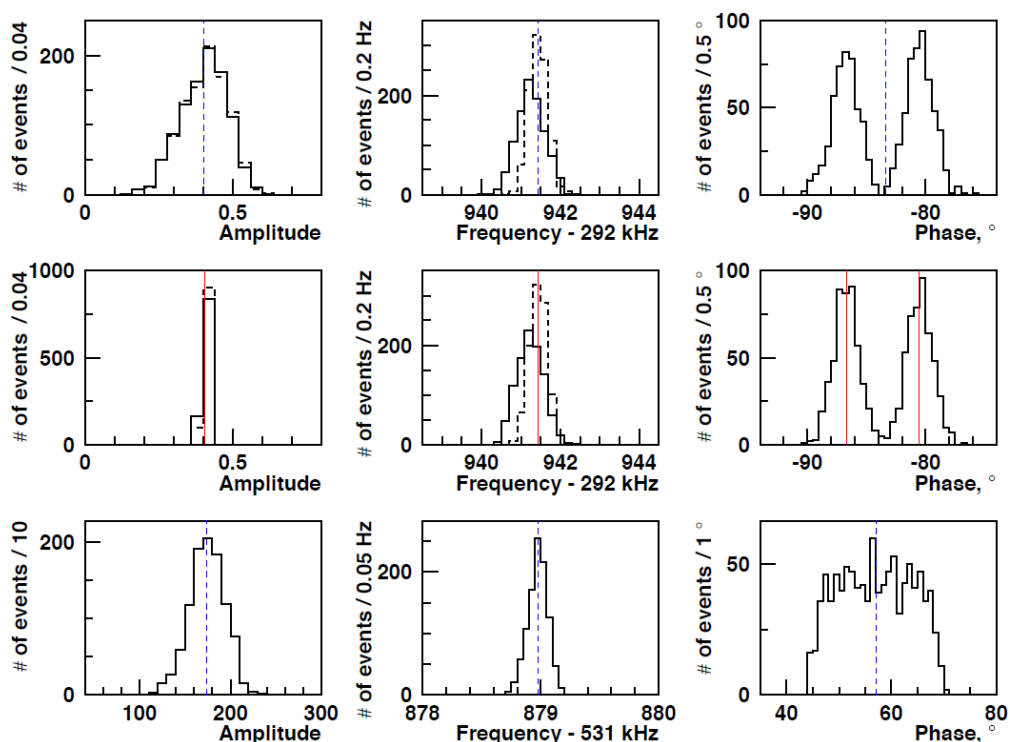


Figure 4.6: Results of LSF (solid lines) and FT (dashed lines) processing of single-component transients: (top panels) 1000 ICR transients; (middle panels) 1000 simulated transients with frequency $f = 292941.44$ Hz, amplitude $A = 0.405$, standard deviation of noise $\sigma = 0.65$, and two randomly assigned phase values $\varphi = -86.70$ and $\varphi = -80.50$; 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 (Paper IV). 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 (Paper IV). Figure adapted from Paper IV.

processing of the ICR transients (the values in question are shown with red lines in Figure 4.6, middle panel and Table 1 of Paper IV).

For both experimental and simulated sets of ICR transients, the same gain in frequency precision (ratio of standard deviations), approximately 1.6, Table 1 (Paper IV), of the LSF method relative to the FT signal processing was obtained. Hence, the increased standard deviation of the FT frequency distributions is 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, 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 previously for the ICR transients, Figure 4.6, top, we conclude that the LSF frequency distribution is unbiased, whereas FT processing produced shifted frequency values. As discussed in Section 3.1.3, the origin of this deviation is due to correlated errors as artifacts of the conventional signal processing workflow.

On the other hand, the amplitude distributions obtained for the modeled FT-ICR MS transients, Figure 4.6, middle panel, are similar for the both methods, as well as the two obtained previously for the ICR transients, Figure 4.6, 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 4.6, top panel, does indicate influence of physical effects (e.g., the scan-to-scan variation in the number of ions injected into the mass analyzer).

For the set of Orbitrap transients, the amplitude, frequency, and phase distributions obtained with LSF and FT methods are shown Figure 4.6, bottom panel; the corresponding mean values and standard deviations are given in Table 1, bottom section of Paper IV. The frequency distributions do not demonstrate a noticeable difference in their standard deviations. 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 4.6, 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.

As the next step, we analyze experimental transients containing the isotopologues from five isotopic fine-structure clusters of peptide substance P. 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 first performed in order to test the LSF's applicability to transients where some components would be unresolved by the LSF method.

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, Table S1 (left columns) of Paper IV. 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

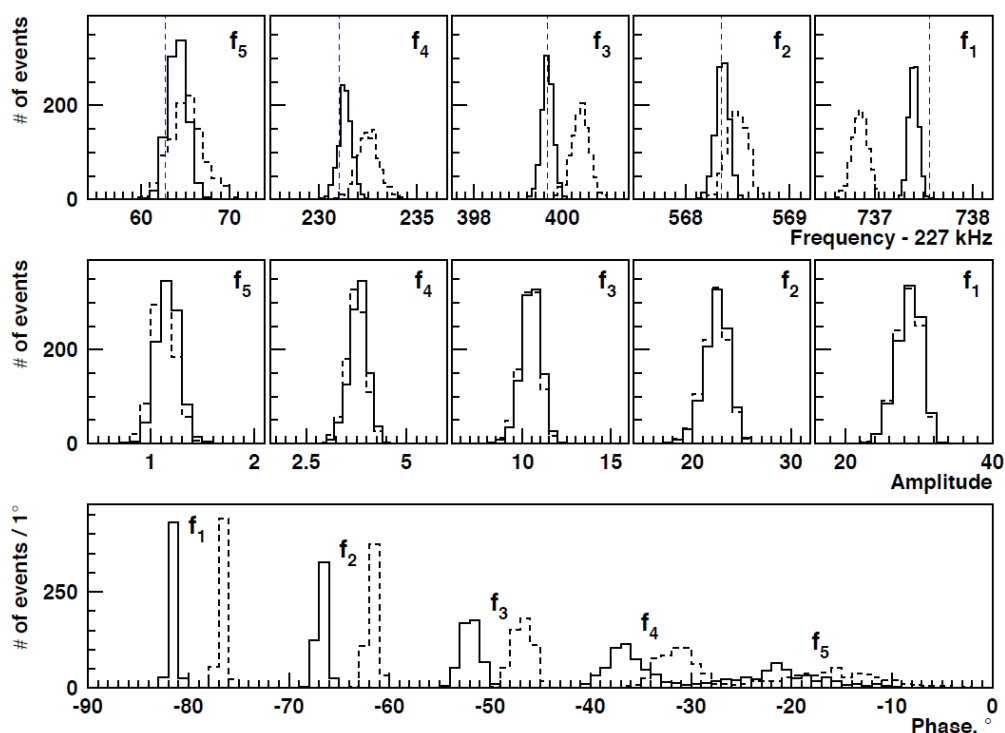


Figure 4.7: Results of LSF and FT processing of ICR multiple-component transients. Top and middle panels show frequency and amplitude distributions obtained with LSF (solid lines) and FT (dashed lines). Bottom panel shows phase distributions obtained with LSF; solid and dashed lines correspond to the two-mode split of the phase distributions. 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 4.5. 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 (Paper IV). Figure adapted from Paper IV.

values of frequencies and amplitudes are listed in Table S1 (center columns) of Paper IV. The frequencies and amplitudes obtained in the LSF analysis of the modeled transient confirm these points, Table S1 (right columns) and, hence, validate the LSF's applicability to the experimental transients with isotopic fine structure ions.

We thus proceed to the analysis of the experimental transients in question. Figure 4.7 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 of Paper IV 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 4.7 top, similar to the analysis of the ICR transients containing the singlet discussed above.

Additionally, systematic errors were observed in the FT results: the frequency distributions

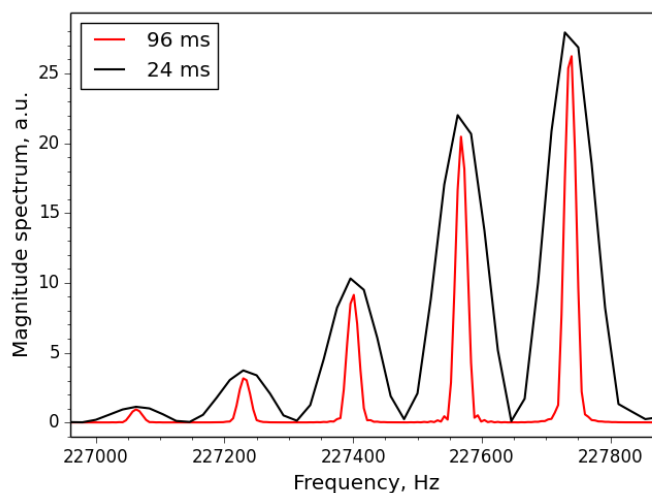


Figure 4.8: Magnitude mode FT spectra plotted for 24 ms and 96 ms transient signals. The 96 ms transient is the averaged set of 1000 transient signals of substance P analyzed on 10 T FT-ICR MS. The 24 ms transient is a truncated variant of the 96 ms transient.

obtained with the FT processing shift toward higher frequencies (e.g., peak f_3 at approx. 227400 Hz), or toward lower frequencies (e.g., peak f_1 at approx. 227737 Hz), as follows from Figure 4.7 top panel. These deviations are presumably due to the spectral artifacts of the conventional FT signal processing workflow, including limited accuracy of the peak maximum interpolation procedure and the nonlinear interference of spectral components discussed in Chapter 3.

Speaking of the second effect, the resolving power was just the minimum required to baseline-resolve the peaks in question, Figure 4.5, top panel. Therefore, the maximum of a given peak is somewhat influenced by the tails of other peak shapes. Indeed, as the acquisition time increases, a better accuracy of the peak interpolation procedure is achieved, and the spectral interference effect reduces because of the increased resolution of peaks. 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 4.7, top panel and Table 2 (rightmost column) of Paper IV. Relative to the frequencies obtained with LSF processing of 24-ms transients, the frequencies obtained 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, to be due to the spectral artifacts of the conventional FT signal processing. For reference, Figure 4.8 illustrates Fourier spectra obtained for 96 ms-long averaged transient signal of the set of 1000 ICR transients of substance P, and its truncated variant with 24 ms length.

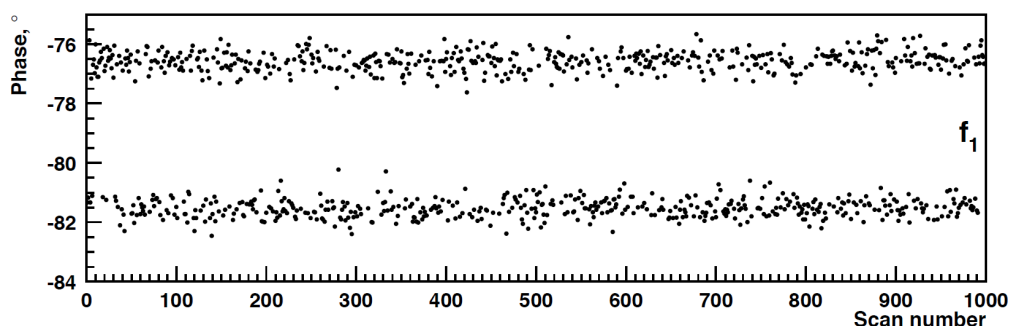


Figure 4.9: Phase jittering of the monoisotopic ion in the LSF analysis of the set of 1000 ICR transient signals acquired for the isolated isotopic distribution of a doubly charged peptide substance P. No correlation between the phase and the scan number was found. Figure adapted from Paper IV, Supporting Information.

4.2.3 Implications for development of FTMS

Importantly, in addition to amplitude and frequency values, the LSF method provides the initial phases of the sinusoidal components, shown as distributions in Figure 4.6, right panels. 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. Similarly, the phase distribution obtained for the set of multi-component ICR transients is also split between the two modes (solid and dashed lines), Figure 4.7.

The observed phase difference between the two modes corresponds to approx. $1/16 \mu\text{s}$, which translates into approx. 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 of the built in electronics. Therefore, the two-mode phase distribution is likely due to the finite precision of the electronics synchronization. Figure 4.9 points to a random nature of phase allocation to the first and second modes. Another striking feature of the phase distributions shown in Figure 4.7 is the high phase coherence, especially evident for the most abundant monoisotopic and A+1 isotopic peaks, as follows from their narrow distributions.

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 followed by FT processing, such as peak shape artifacts reported previously in absorption-mode FT NMR spectra,[108] or baseline roll in absorption mode FT-ICR MS.[91] From the phase correction equations derived in Section 4.1 it follows that the baseline roll is an intrinsic artifact of FT-ICR MS with absorption-mode spectral representation due to virtual nature of the phase intersection point, whereas the peak shape artifacts in FT NMR is a processing error resulting when an integer number of points is used for linear phase correction while a non-integer number of points is actually required for this procedure to be numerically exact, i.e. free of processing artifacts.

4.2.4 Advantages and limitations

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 expected 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, Eq. 4.24, as is, the LSF method can be used as a reference method for development and characterization of the filter diagonalization method (FDM) MS.

Convergence to the global minimum of χ^2 -function may be complicated when the number of fitting parameters is large. 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. Hence, in contrast to instrument development, the scope of analytical applications of the LSF method is mainly limited to specific classes of analytes with sufficiently low number of peaks and with their positions known in advance.

4.2.5 LSF-based workflow for quantitative proteomics

In the following description, we take advantage of the implemented LSF method to improve the analysis in LC-MS-based quantitative proteomics on Orbitrap FTMS. An example of such analysis is acquisition of transient signals at resolving power of $R=30'000$, which corresponds to $T=96$ ms acquisition time (with absorption-mode FT) in the case of Orbitrap Elite FTMS platform. However, in such applications it is desirable to increase the throughput, as well as to identify and quantify more peptides. Experimental time is spent among a number of stages including instrumentation aspects, such as ion transmission and trapping, and signal detection and processing. Here we address the second question, aiming to reduce the required detection time 4-fold, i.e. 24 ms for the considered example, or below, while preserving the abundance accuracies at analytically useful levels.

The experimental method was specifically designed and evaluated for quantitative proteomics with Tandem Mass Tags (TMT).[109] Figure 4.10 shows large-scale LC-Orbitrap Elite FTMS analysis was conducted for 10-plex TMT labeled yeast digest sample mixed at the ratios of 1:2:5:3:4:6:8:10:9:7. The inset illustrates analysis of a selected MS/MS scan. While absorption-mode FT (eFT algorithm) requires 96 ms detection period, analysis of 4-fold shorter, 24 ms, transients is routinely achievable with the LSF processing.

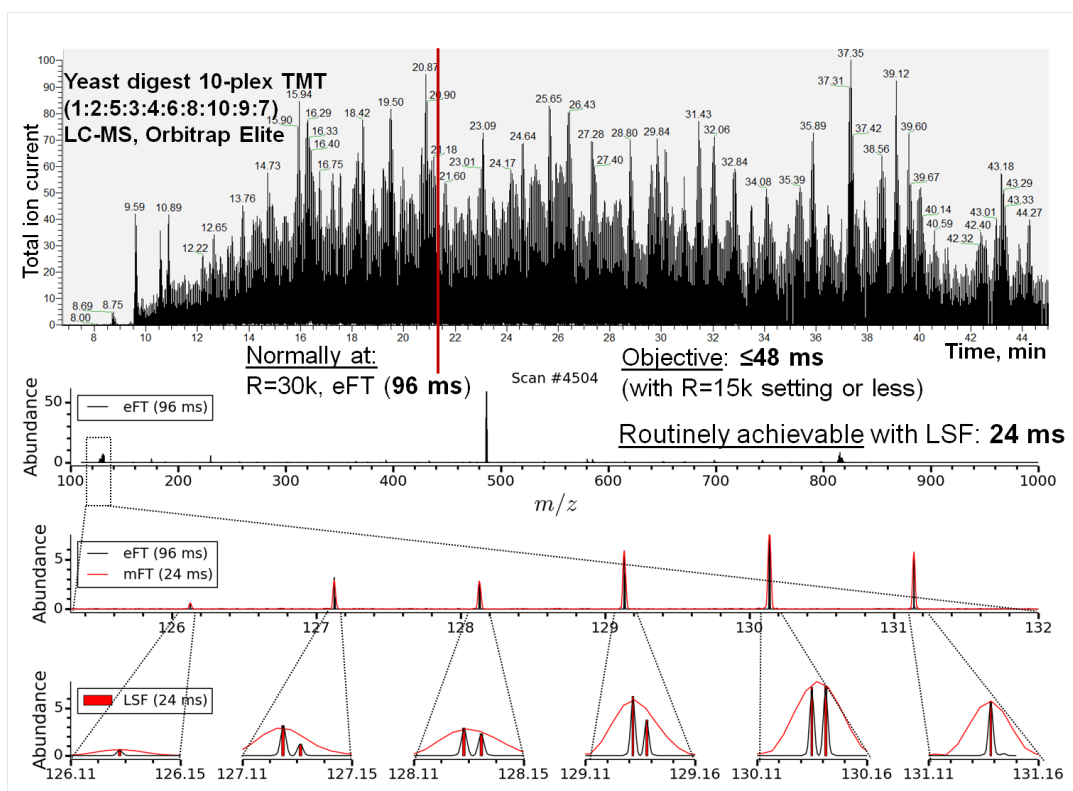
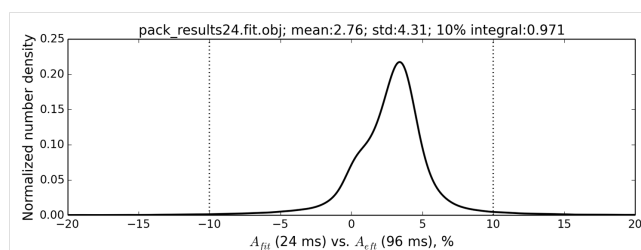


Figure 4.10: LC-MS analysis of yeast digest labeled with 10-plex TMT. Experiments were performed on Orbitrap Elite FTMS. Approximately 6000 transient signals were acquired and analyzed. With LSF processing, 4-fold reduction of required detection time per each transient was routinely achieved.

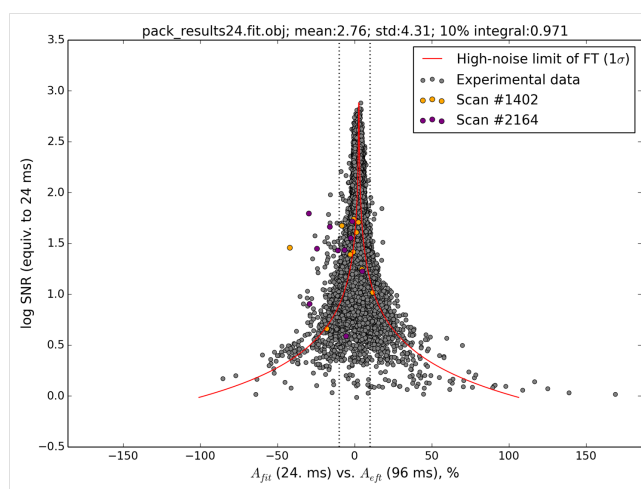
For all reporter ions in this dataset, Figure 4.11 shows results of statistical evaluation of the relative differences of ion abundances between those obtained with LSF processing of 24 ms transients and the eFT algorithm with 96 ms detection period. Namely, Figure 4.11a demonstrates that more than 97% of all reporter ions are within the 10% range of relative differences of ion abundances. The 10% range was selected as a reasonable threshold for the range of analytically useful abundances. It should be noted that the 97% value is a lower bound estimation (i.e. an underestimation) of the LSF results since the reference values obtained using the eFT processing also contain measurement errors.

In turn, Figure 4.11b evaluates the LSF method performance with respect to signal-to-noise ratio (SNR) values of the experimental data. Specifically, the vertical axis represents SNR values equivalent to FT analysis of 24 ms transient signals. The red lines illustrate one σ error of FT for 24 ms transient length at low SNR limit, based on the theoretical estimations described in Paper VI, Supporting Information:

$$\frac{A_{fit} - A_{eft}}{A_{eft}} \cdot 100\% = \pm 100\% \frac{1}{10^{\log SNR}} + \langle A_{eft} \rangle \quad (4.27)$$



(a) Normalized number density plotted for the full set of relative differences of ion abundances obtained using the two approaches. More than 97% of reporter ions are within 10% range of relative differences of ion abundances.



(b) Signal-to-noise equivalent to FT analysis of 24 ms transients vs. relative differences of ion abundances. The red line indicates theoretical one σ error of FT for low SNR limit.

Figure 4.11: Precision of the LSF method evaluated for the full range of signal-to-noise ratios of all TMT reporters from the set of approximately 6000 experimental transients. Abundances of reporter ions, obtained with LSF processing of 24 ms transient signals, are referenced to ion abundances obtained using the eFT algorithm and 96 ms detection period.

where the $\langle A_{eft} \rangle$ term is added to focus on random errors by correcting for a systematic shift. Notably, the LSF data is well within the high-noise FT limit, meaning that the LSF method demonstrates stability for all ion abundances of the analyzed set in the broad range, 3 orders of magnitude, of SNR values. On the other hand, at high SNR values the LSF method shows its precision performance by revealing two outlying points: one of them was due to a systematic error (spectral artifact) of the reference eFT data while the second one was due to interference from extra species occasionally present in the m/z range of interest (yet, a numerical correction is allowed).

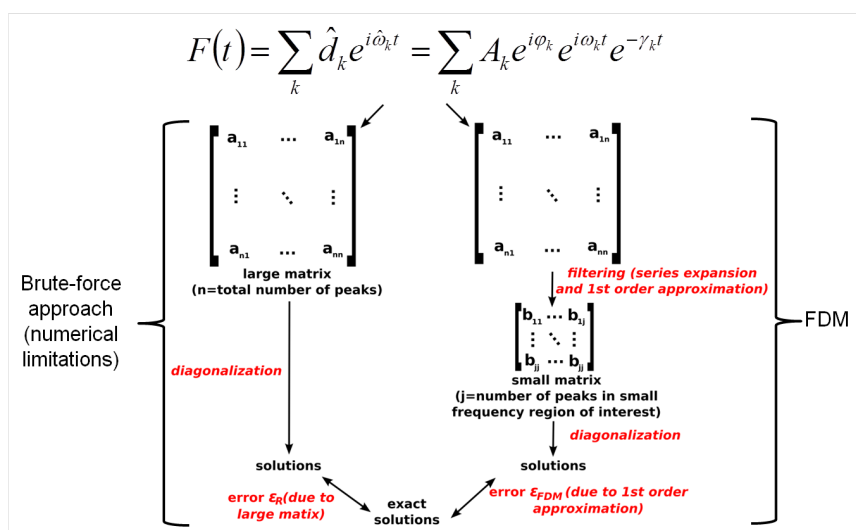


Figure 4.12: Comparison of FDM and a brute-force numerical method for analysis of a harmonic inverse problem.

4.3 Filter-diagonalization method

Frequencies and amplitudes of transient signal components can be obtained by either spectral or parameter estimation, as discussed previously in Section 4.2. The FT signal processing is arguably the most robust solution, albeit, being a spectral estimator, it has several drawbacks. These include measurement accuracies, as derived in Chapter 3 and experimentally revealed in Section 4.2. and rather limited resolution, as shown in Section 4.1 by considering the use of information on ion phases to weaken the quantitative form of the FT's uncertainty principle.

In earlier works, super-resolution (SR) signal processing methods have been developed to overcome the FT resolution limitations.[110] 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. Historically, the FDM was first introduced for quantum mechanics and further developed for nuclear magnetic resonance (NMR) spectroscopy.[111, 112, 113, 114, 115, 116, 117, 118] Previously, the FDM was also applied to trace frequency modulation effects in experimental transient signals in FT-ICR MS and to evaluate this method for super-resolution processing of simulated transients.[119, 120] In Paper V, we introduce FDM-based mass spectrometry (FDM MS) to accelerate high resolution FT-ICR MS data acquisition and demonstrate its performance on the experimental data obtained for common classes of molecules and macromolecules.

4.3.1 Methodology

The FDM-based signal processing was implemented as follows. The FDM calculations were based on the standard numerical scheme of FDM, Figure 4.12, while, being MS-conditioned,

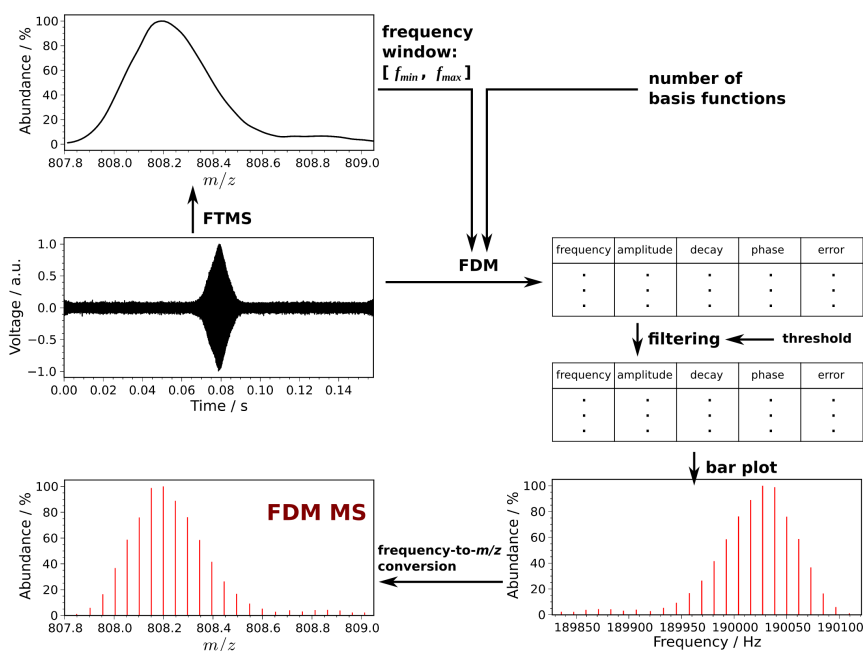
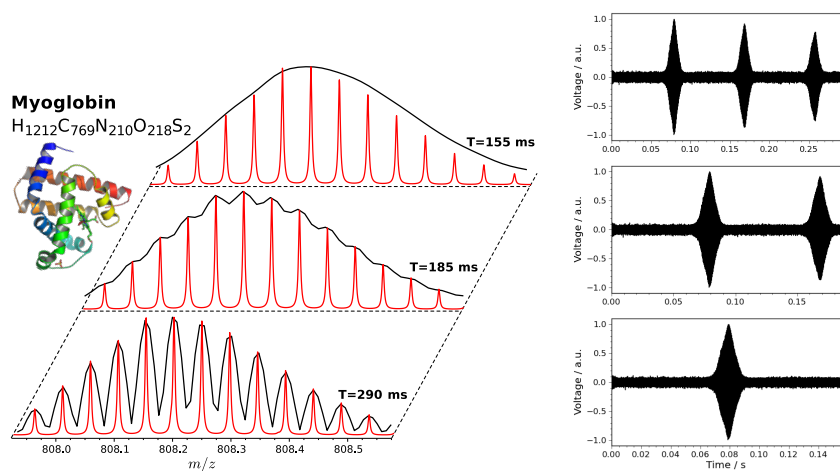


Figure 4.13: Main steps in the FDM-based mass spectrometry methodology. The FDM algorithm is described elsewhere.[111, 114, 115, 116] Other steps are realized in the custom-written pyFTMS package. Figure adapted from Paper V.

the methodology of the application of FDM in FTMS was developed, Figure 4.13. Specifically, the harmonic inversion problems, Figure 4.12, are solved for the selected frequency (mass-to-charge) window and chosen number of the basis functions, Figure 4.13. The frequency window selection can be performed without (data-independent approach) or with (data-dependent approach) use of the preliminary FTMS step performed to find the location of the frequency regions with peaks of interest. In the latter case, since FT processing does not necessarily resolve the peaks of interest, FT processing is used to reveal the corresponding frequency regions for further analysis. The transient signal is subjected to FDM calculations for the selected frequency window. The FDM calculations are carried out with a user-defined number of basis functions. The number of basis functions is selected as an upper-bound constraint for the method's numerical implementation. For the appropriate results, the number of basis functions should exceed the actual number of ion peaks to be resolved in a given m/z (frequency) window. The number of basis functions can be selected on the basis of the particular sample being analyzed, MS experimental conditions, or stability of the calculations performed for a set of different numbers of basis functions. In the calculations performed here, the number of basis functions was between 15 and 240 per a frequency window. In addition, the control calculations for the sets of different numbers of basis functions were carried out to verify the calculations stability and to confirm the obtained results. The FDM processing provides the solutions as FDM table that contains information on frequency, amplitude, phase, decay, and roughly estimated complex frequency error as a figure of merit for each solution. In the following step, the possible spurious solutions (false positives) are filtered out on the basis



(a) FDM vs magnitude-mode FT mass spectra for (top) 155 ms, (middle) 185 ms, and (bottom) 290 ms acquisition times. (b) Corresponding transient signals with length of (bottom) 155 ms, (middle) 185 ms, and (top) 290 ms.

Figure 4.14: FDM versus magnitude-mode FT for analysis of isolated charge state of myoglobin on 10 T FT-ICR MS as a function of data acquisition time (transient length, T). Figures are adapted from Paper V.

of the threshold values specified for user-defined acceptable peak parameters. If the applied threshold is not sufficiently strong, the spurious peaks can remain in the spectra and, *vice versa*, if the threshold is too strong, the real ion peaks can be lost. Therefore, method optimization should be performed for the particular experimental conditions and type of samples employed. Further, on the basis of the frequencies and amplitudes from the FDM table, the bar plot in the frequency scale is created and then subjected to the standard frequency-to- m/z conversion (mass calibration).[61]

In this work, considered applications for FT vs. FDM performance comparison included structural analysis of common classes of molecules and macromolecules. Comparison was made as a function of data acquisition time. For example, Figure 4.14a shows analysis of isolated charge state of protein myoglobin. The corresponding ICR transient signals are shown in Figure 4.14b. A custom library developed in this thesis for FDM processing includes the following features. The maximum number of basis function as a constraint of the algorithm is allowed to be user-defined. To keep appropriate precision of internal number representation, user-defined internal numerical precision of the algorithm is implemented instead of the original setting. As an optimization, a parallel algorithm for the calculations was realized. Probability density distributions calculated for abundance and frequency (m/z) values of analyzed ions were implemented in place of raw bar-plot data representation.

4.3.2 Super-resolution mode in FTMS

In this section, we characterize resolution performance of FDM processing relative to that of the standard FT processing. Similarly to the upper-bound resolving power limit in FT processing, Chapter 2, there exists a lower-bound of detection period, $T_{\text{FT}}^{\text{min}}$, required for the FT-based signal processing to baseline resolve peaks of interest. Specifically, as follows from the analysis in Chapter 2, the time-vs.-frequency uncertainty principle of the FT signal processing results in the following condition of baseline resolution for an interval of frequencies, $[f_1; f_2]$:

$$T_{\text{FT}}^{\text{min}} = C_{\text{FT}} / \Delta f_{\text{min}} \quad (4.28)$$

as illustrated in Figure 4.15a. The Δf_{min} is the minimum frequency difference for the components to be resolved in the interval in question. Recall that C_{FT} is defined by the apodization function and type of spectral representation and $C_{\text{FT}} = 4$ for the magnitude mode and von Hann apodization window, Chapter 2. As previously, the lower bound is approached when the ion packet coherence is sufficiently high. However, if two coherent ion packets get separated in terms of their total accumulated phases in the phase space, but the resolution performance of the FT signal processing is insufficient and hides the actual ion separation.

In turn, the time-vs.-frequency uncertainty principle of FDM for the frequency interval under consideration, $[f_1; f_2]$, defines the lower bound of detection period required to resolve N components in the frequency interval in question:[111]

$$T_{\text{FDM}}^{\text{min}} = C_{\text{FDM}} / \langle \Delta f \rangle \quad (4.29)$$

where C_{FDM} is a constant defined by the method implementation, e.g., $C_{\text{FDM}} = 2$ for the standard implementation, and $\langle \Delta f \rangle$ is the inverse value of the local peak density $\rho(f)$ averaged over the interval $[f_1; f_2]$:

$$\langle \Delta f \rangle = \left[\frac{1}{f_2 - f_1} \int_{f_1}^{f_2} \rho(f) df \right]^{-1} = \frac{f_2 - f_1}{N} \quad (4.30)$$

Since $f_{\text{min}} \leq \langle \Delta f \rangle$, $T_{\text{FDM}}^{\text{min}} < T_{\text{FT}}^{\text{min}}$ and thus one may speak about the super-resolution mode that corresponds to the region of time T and frequency difference Δf , where $T_{\text{FDM}}^{\text{min}} \leq T < T_{\text{FT}}^{\text{min}}$ and $f_{\text{min}} < f \leq \langle \Delta f \rangle$, which is located below the FT lower-bound $T_{\text{FT}}^{\text{min}}$, as schematically represented in Figure 4.15a.

Provided that the scale of frequency details to be resolved is consistent with the ion packet coherence, superior resolution performance of FDM compared to FT-based signal processing can be achieved. Speaking of SR methods, it should be noted that whereas the common definitions of resolution and resolving power given for peak shape mass spectra are not directly applicable to FDM mass spectra, the uncertainty principle of FDM provides a convenient measure to quantify its resolution performance in terms of the detection time, Figure 4.15a.

4.3.3 SNR limited resolution

With noise present in analyzed transient signals, the uncertainty principle of FDM, Eq. 4.29, can be realized as a lower-bound estimate of the minimum required detection period, $T_{\text{FDM}}^{\text{min}}$. Given finite SNR of measurements, it is reasonable to take into account dispersion of experimental m/z and ion abundance values while characterizing the resolution performance of a signal processing method. Indeed, recall that due to experimental errors, a quantity to be measured is represented with a certain distribution. Specifically, the frequency obtained in a single measurement deviates from the exact value, meaning that in repeated measurements the measured frequency f is a random value that can be characterized by a probability density distribution, $p(f)$. The standard deviation of f is given by the variance of f as $\sqrt{\text{var } f}$. The standard deviation defines the width of the distribution $p(f)$, represents the influence of the random errors, and characterizes the measurement precision. By definition common in the measurement theory, the signal-to-noise ratio of the measured value f is:

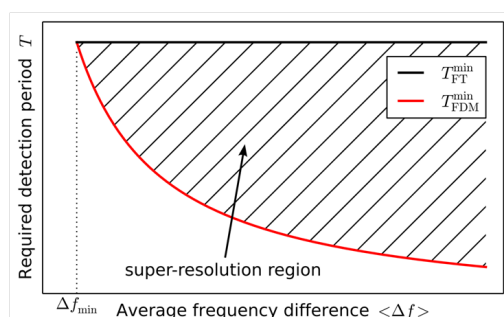
$$\text{SNR} = \langle f \rangle / \sqrt{\text{var } f} \quad (4.31)$$

where $\langle f \rangle$ is the mean of the distribution $p(f)$. Thus, a finite precision of measurements leads to a finite SNR of the measured result $\langle f \rangle$.

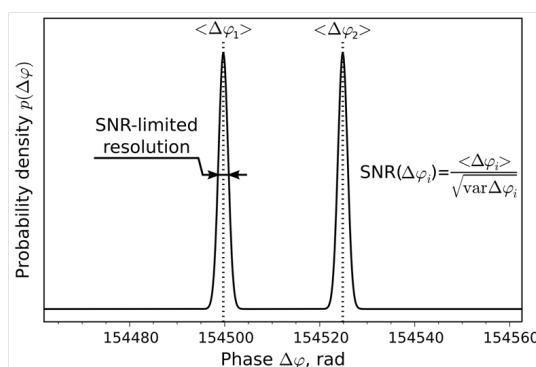
For example, when FDM is applied to FTMS transients, the SNR value in question is governed by the instrumental random error, which is due to variation of the measured frequencies in consecutive MS scans, and the signal processing random error, which is due to the noise component of the transient. Consider two distributions $p_1(f)$ and $p_2(f)$ obtained in frequency measurements for two ion packets with frequencies f_1 and f_2 , Figure 4.15b. The concept of SNR-limited resolution allows applications of the terms of resolution and resolving power to a single peak, viz. $p_1(f)$ or $p_2(f)$ (FWHM definition), or to the pair of peaks $p_1(f)$ and $p_2(f)$ (with peak valley and abundances specified). For instance, one may speak qualitatively about the achieved resolution, e.g., baseline resolution, as well as quantify the resolution and resolving power as discussed above for the peak shape mass spectra. Thus, the SNR-based criterion allows the regular definitions of resolution and resolving power to be applied to FDM data, whereas otherwise, as FDM is a parameter estimator, they are not applicable to FDM bar plot mass spectra. This consideration is likewise applicable to other parameter estimators, e.g., LSE. For the sake of consistency, if an SR method is characterized in terms of the SNR-limited resolution, a similar logic should also be applied to the FT signal processing in those cases.

4.3.4 Spectral irregularity

Figure 4.16a compares FDM and magnitude-mode FT spectra obtained for a transient signal acquired in the analysis of a monomer-dimer mixture of protein superoxide dismutase on Orbitrap Elite FTMS. Due to regular distribution of the ion peaks of this mixture, the minimum frequency spacing equals the average frequency spacing. These values are equal to 3.77 Hz in this experiment. From the uncertainty principle of FDM, Eq. 4.29, it follows that the resolution



(a) Resolution performance of FT and FDM methods in terms of the detection period required to resolve analyte ions in a given frequency interval. The resolution advantage of FDM enables the super-resolution mode in FTMS for the indicated region of average frequency differences and detection periods.



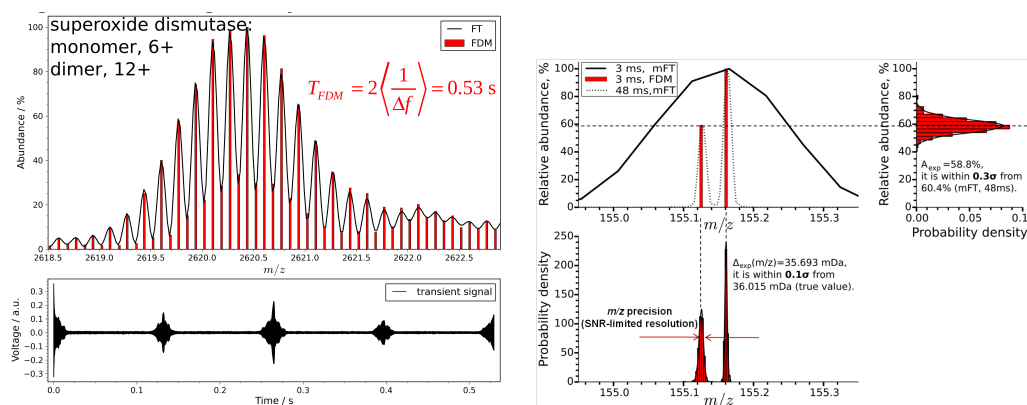
(b) Signal-to-noise ratio (SNR)-limited resolution illustrated for two ions with different m/z .

Figure 4.15: Quantifying resolution performance of FDM.

performance of FDM is governed by irregularity of spectral components with respect to their frequencies. The magnitude-mode FT and FDM results are in agreement with the uncertainty principles Eq. 4.28 and Eq. 4.29: to baseline resolution magnitude-mode FT required the acquisition time of $T = 4/3.77 \text{ Hz} = 1.06 \text{ s}$ and the required time of FDM is $T = 2/3.77 = 0.53 \text{ s}$. Hence, in accordance with the theory, only a 2-fold gain in resolution performance of FDM vs. magnitude-mode FT was achieved due to regular distribution of the peaks. This is also consistent with the analysis myoglobin on FT-ICR MS, Figure 4.14, described previously.

On the other hand, from Eq. 4.28 and Eq. 4.29 it follows that a high gain can be achieved for highly irregular distribution of peaks. Figure 4.16b shows FDM and magnitude-mode FT m/z values and ion abundances obtained in the analysis of isotopically enriched amino acids L-Lysine- $^{13}\text{C}_6$, $^{15}\text{N}_2$ and L-Lysine- $^2\text{H}_8$ with mass difference of 36 mDa. In accordance with the uncertainty principles, a significant gain in resolution performance of FDM vs magnitude-mode FT, more than 10-fold, was achieved in this analysis.

Importantly, the 10-fold gain also points out that the actual size of ion packets in these



(a) FDM in the analysis of a monomer-dimer mixture of protein superoxide dismutase on Orbitrap Elite FTMS. Resolution gain is minimal due to regular distribution of peaks.

(b) Significant resolution advantage of FDM in the analysis of isobaric mixture of isotopically enriched amino acids on Orbitrap Elite FTMS. The mass differences between all pairs of peaks of these analytes are about 36 mDa.

Figure 4.16: Effect of spectral irregularity on resolution gain of FDM in proteomic applications.

experiments was sufficiently small, i.e. the ion packets were sufficiently coherent so that their size was at least 10-fold less than 4π (1/5 of full circle). Indeed, 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 should not exceed the distance between them at the end of ion detection. As shown for isotopic fine structure analysis of peptides using long transients acquired with a standard ICR cell, Paper VI, large size of ion packets developed in the phase space may ultimately limit the overall resolution performance of the measurement scheme with the FDM as a signal processing element, hence reducing the size of SR-region on the plot Figure 4.15a. Similarly, the effect of SNR considered previously, Figure 4.15b, would restrict the achievable resolution performance, Figure 4.15a. Therefore, the aimed resolution gain when using the FDM must be consistent with the value of finite precision of FDM measurements resulted from finite SNR, as shown for the amino acids analysis, Figure 4.16b.

5 Instrumentation

Focused on the final part of the measurement setup, this chapter describes developments in instrumentation of the present work. Based on the preliminary developments of the Chapter 2, the current chapter takes advantage of the uncertainty principle of measurements with respect to resolution of ion packets with different m/z and mass measurement accuracy. Specifically, new ICR cells of two design classes are implemented to achieve increased analytical performance via developments in ion motion conditions as well as via developments in signal detection.

In Section 5.1, a multi-electrode ion cyclotron resonance (ICR) cell, referred to as the "4X cell", is implemented for signal generation at the quadruple frequency multiple and characterized on the 10 FT-ICR MS. As discussed in the previous chapters, ion packets with different m/z oftentimes get separated in a mass analyzer to a much higher degree than it appears as a resolution level in a mass spectrum when the conventional measurement scheme is employed. As such, the 4X cell with the signal generation at the quadruple frequency multiple provides a 4-fold increase in resolution performance compared to a serial ICR cell.

In turn, Section 5.2 describes an ICR cell, referred to as the "NADEL cell", with narrow aperture detection electrodes positioned radially inward the cell's axis. The NADEL cell provides three distinct regimes: the regime of a standard ICR cell with its performance characteristics comparable to those of the state-of-the-art ICR cells, a regime for increased resolution performance via generation of wide band transient signals, and a regime for increased mass measurement accuracy with ion motion tailored to signal generation at "unperturbed" ion cyclotron frequency instead of reduced cyclotron frequency.

5.1 4X cell

Paper VII presents implementation of the 4X cell aimed at increased throughput for biomolecular analysis in FT-ICR MS. In recent decades, a common approach to improve the throughput in FT-ICR MS was to increase the values of generated fundamental frequency of ion motion

in the ICR cell via increased magnetic fields.[18, 121, 122] However, there are technological limitations for increasing magnetic fields. For instance, currently the state-of-the-art FT-ICR MS instruments have 10–15 T magnetic fields, and a number of projects to build 21 T FT-ICR MS are in progress.[123] On the other hand, recent developments of ICR cells with improved harmonization of the electric field extend the value of achievable resolving powers via longer lifetimes, about 10-100 fold, of transient signals.[124, 125, 126, 127, 128, 107, 129, 130] However, these approaches do not increase the resolving power per given detection period and hence do not improve the throughput of FT-ICR MS.

Apart from increased fundamental frequency and improved harmonization, developments of the measurement principle of FTMS, including methods for signal processing and signal detection, are vital for achieving high throughput of high-resolution FTMS. Developments in signal detection are represented by the frequency multiple method that improves the resolution performance of FTMS.[131, 132, 133] Although this method for signal detection has been known for almost 30 years, it has been poorly explored and has not been implemented in the commercial FT-ICR MS instruments. A shortage in the historic viewpoint was because the performance of FT-ICR MS at frequency multiples was estimated using the model of highly damped transients. As discussed previously in the thesis, Chapter 4, transient damping can be neglected in the case of modern FTMS instruments when the acquisition time is sufficiently short. In general, it is assumed now that shorter acquisition times will translate into more interrogated peptides and more proteins being identified, as well as improvements in the reproducibility of proteomics experiments.[134] Therefore, implementation of frequency multiple detection of undamped time-domain signals at high acquisition rate could be considered as a complementary approach to the increase in magnetic field strength in FT-ICR MS.

5.1.1 Design of the 4X cell

The design of the 4X cell is based on the platform of a standard open-ended cylindrical Ultra ICR cell (Thermo Scientific) and was installed on the 10 T FT-ICR MS in place of the Ultra cell. Figure 5.1 shows a schematic diagram and electric wiring for the 4X ICR cell. The overall dimensions, including inner diameter of approx. 56 mm and length of excitation and detection electrodes of 70 mm, correspond to those of Ultra ICR cell. The original four electrodes for ion excitation and detection were replaced with 16 equal-size electrodes. The metal grid employed for ion excitation in the Ultra cell was removed. The configuration of the excitation and detection electrodes of the 4X cell reproduces design suggested earlier.[133] One electrode set consisting of eight electrodes (four electrodes for excitation 1 and four electrodes for excitation 2) provided a standard RF dipolar ion excitation; the other electrode set consisting of eight electrodes (four electrodes for detection 1 and four electrodes for 2) was used for induced ion current detection, Figure 5.1. Ion confinement was performed using trapping rings similar to the Ultra cell design. The offset DC potentials applied to the excitation and detection electrodes of the 4X cell were adjusted to increase the signal-to-noise ratio (SNR)

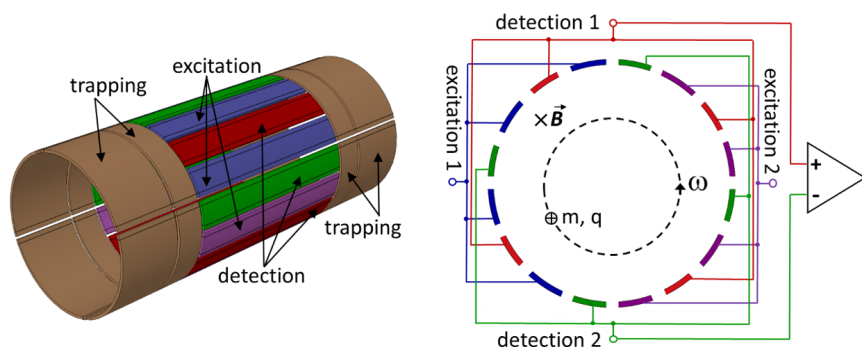


Figure 5.1: Schematics of the 4X ICR cell aligned along magnetic field lines: (left) perspective schematic and (right) cross section with a wiring diagram. RF excitation is provided through two pairs of four electrodes each. Detection is organized through other two pairs of four electrodes each. Figure adapted from Paper VII.

and reduce magnetron sidebands.

5.1.2 Resolution performance

Figure 5.2 shows the broadband frequency spectra for the ultramark calibration mixture (bottom panel) and one of its components at m/z 1422 after isolation in the linear ion trap before injection into the 4X cell (top panel). They demonstrate that the resolving power at quadruple ($4f$) frequency multiple exceeds that at single (f), double ($2f$), and triple ($3f$) frequency multiples for the 4X cell under consideration. In both cases, the four-fold increase in the peak resolution has been observed at the quadruple frequency compared to that at the single frequency at the same detection periods. Specifically, the frequency spectra for both the calibration mixture and the isolated ions at m/z 1422 were measured for the detection periods of 384 ms and 786 ms, respectively.

In both cases the time-domain signals were limited by the detection period time, T , hence, for the standard detection scheme i.e. with measurements at the fundamental ion frequency, the resolving power limit could be obtained using Eq. 2.15, where $C_{\text{FWHM}} = 0.83$, which corresponds to magnitude-mode spectral representation with no signal apodization employed:

$$R_{\text{FWHM}}^{\text{max}} = \frac{fT}{0.83} \quad (5.1)$$

For example, the estimated resolving power for the peak at the fundamental frequency (single frequency multiple) of the ultramark component (fluorinated phosphazine) at m/z 1422

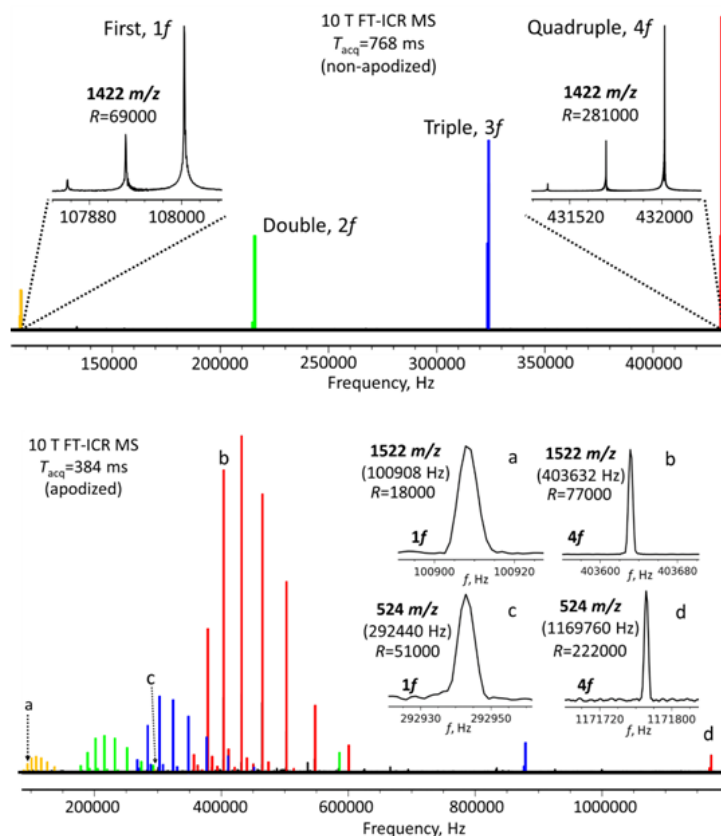


Figure 5.2: 4X cell FT-ICR MS of a calibration mixture containing MRFA and ultramark compounds recorded in a 10 T magnetic field. (Top) non-apodized frequency spectrum of isolated singly protonated ultramark ions at m/z 1422 (acquisition time 768 ms; isolation width 5 m/z ; trapping potential 0.1 V; excitation signal length $T_{\text{exc}} = 20$ ms and excitation voltage $V_{\text{exc}} = 100$ V) and (bottom) apodized broadband frequency spectrum (acquisition time 384 ms; trapping potential 0.1 V; excitation signal length $T_{\text{exc}} = 60$ ms and excitation voltage $V_{\text{exc}} = 100$ V). Figure adapted from Paper VII.

shown in Figure 5.2, top panel, will be 68,000 which correlates with the experimentally obtained value of 69,000. In turn, the measured resolving power is 281,000 for peak at the quadruple frequency multiple, Figure 5.2. Effectively, the peak width in the magnitude-mode Fourier spectra obtained for the undamped 768 ms transients at m/z 1422 is equivalent to 40 T magnetic field strength according to Eq. 5.1 and the cyclotron frequency expression $\omega_c = qB/m$.

Thus, this data supports the thesis that the ions trapped in the cell can separate in the phase space much better than it appears in a mass spectrum obtained with conventional measurement scheme, Chapter 2. To reveal the ultimate limitation imposed by the size of ion packet, the following measurements were performed. Figure 5.3 shows the dependence of resolving power on the detection time for isolated singly protonated ultramark component

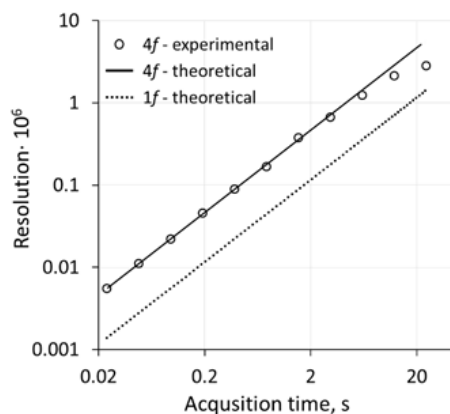


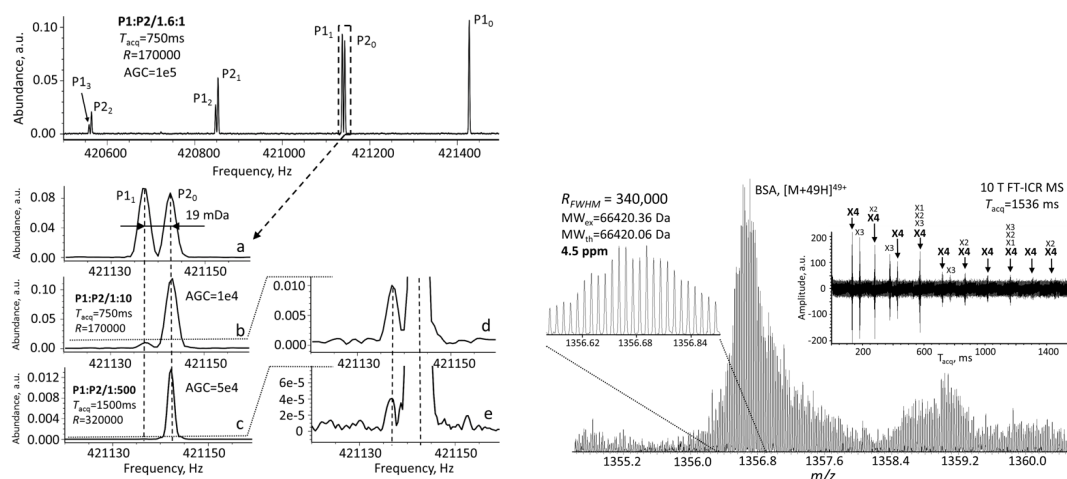
Figure 5.3: Resolution performance of the 4X cell evaluated in the analysis of singly protonated ultramark component at m/z 1322. Solid and dashed lines show FT resolution limit for the quadruple and first frequency multiples, correspondingly. Figure adapted from Paper VII.

at m/z 1322. This plot demonstrates that the peak resolution at the quadruple frequency $4f$ aspired to the resolution corresponding to the fundamental frequency f for substantially long, up to 6 s and more, detection periods. In turn, deviation from the straight line as the detection time increases points out the effect in question, i.e. the uncertainty principle limit due to finite size of the ion packet starts to manifest itself. The limiting level, also known as the aberration limit, is examined in further text for the NADEL cell.

5.1.3 Evaluation for applications

Figure 5.4a shows quadruple frequency spectra for a mixture of two singly protonated model peptides, P1 (GYQYLLEPGDFR, monoisotopic mass 1458.6127 Da) and its de-amidated version, peptide P2 (GYEYLLEPGDFR, monoisotopic mass 1459.5975 Da). The difference between peptides is due to de-amidation that results in the overlapping of a monoisotopic peak of peptide P2 and ^{13}C isotopic peak of peptide P1 with 19 mDa mass difference between them. The resolving power in excess of 300,000 is required to baseline-resolve these peptides at m/z 1459 (corresponds to reduced cyclotron frequency of approx. 105.25 kHz in a 10 T magnetic field). This resolving power was achieved using 4X cell with 1.536 s transients and with abundance ratio between the two peptides of up to 500 as shown in Figure 5.4a c, e.

Notably, at the lower peptide concentration ratios up to 1:10 the isotopic patterns were clearly resolved even for the shorter acquisition time of 0.768 s (Figure 5.4a 3 a, b). The other notable observation was that the averaging of only 10 single scans was sufficient to observe the low abundance peaks even for relatively low number of charges loaded into the cell with the AGC settings of $1\text{E}4$ to $5\text{E}4$. In addition, the signal-to-noise ratio (SNR) of 600 was achieved for these two peptides at 1459 m/z with AGC= $5\text{E}4$ for a sum of 10 scans using 4X cell. This SNR value is comparable with the SNR of approx. 700 obtained at 1422 m/z using the standard



(a) Isobaric peptide mass measurements and spectral dynamic range evaluation at quadruple frequency with 4X cell. The spectra were obtained for as low as 10 summed transients. (b) Protein mass measurements with 16-electrode cell in 10 T FT-ICR MS. Analysis of isolated BSA49+ cluster at quadruple frequency.

Figure 5.4: Advantages of the 4X cell for FTMS applications. Figure adapted from Paper VII.

(Ultra) ICR cell under the same experimental conditions.

Figure 5.4b shows a mass spectrum of the isolated 49+ charge state bovine serum albumin (66 kDa) at m/z 1356. The average resolving power across all baseline resolved ^{13}C isotopic pattern peaks in the magnitude-mode apodized mass spectrum was 340,000 (left inset) for a single scan with acquisition time of only 1536 ms. The typical isotopic beats in a transient signal show the difference between measurements at first and quadruple frequencies, see right inset. As expected, the difference is two beats at the first frequency against 10 beats recorded at the quadruple frequency.

5.2 NADEL cell

Today the most common ICR cell configuration is a cylindrical cell with a pair of 90 degrees wide excitation electrodes and a pair of similar 90 degrees wide detection electrodes, all positioned radially at a similar distance from the ICR cell axis. [4, 18]. The drawbacks of this configuration include a limitation on the resolution performance, substantial perturbation of excitation radio-frequency (RF) field by detection electrodes, and a low degree of design flexibility. In Paper VIII, we describe a concept of ICR cells with narrow aperture (flat) detection electrodes (NADEL) addressing these limitations of ICR cells in FT-ICR MS.

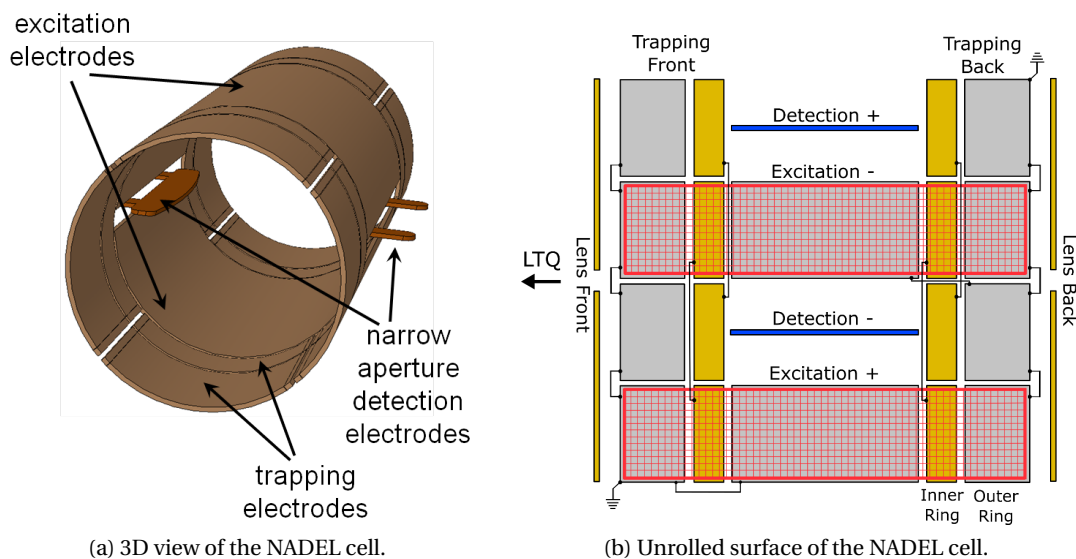


Figure 5.5: Schematic representation of the narrow aperture detection electrodes (NADEL) ICR cell. The NADEL cell contains two pairs, outer and inner, of trapping ring electrodes, excitation grids, conventional 90 degrees excitation and narrow aperture (flat) detection electrodes with a thickness of 2 mm and curvature $R = 125$ mm. Figure adapted from Paper VIII.

5.2.1 Design of the NADEL cell

The design of the NADEL cell was developed based on the standard open-ended cylindrical ICR cell (Ultra Cell, Thermo Scientific). In the NADEL cell, Figure 5.5, the standard 90 degrees detection electrodes were substituted with narrow aperture (flat) detection electrodes with a thickness of 1-3 mm. The new detection electrodes were extended closer to the center of the cell in the on-axis plane. The NADEL cell as well as the standard Ultra (90 degrees detection electrodes) were mounted, one at a time, onto the 10 T FT-ICR MS. Ion trapping in the cell was achieved with two pairs of trapping ring electrodes. The outer trapping rings were grounded throughout the complete experimental sequence. A potential of 3 V was applied to the inner trapping rings to confine transferred ions inside the ICR cell during ion trapping and relaxation events. The same level of trapping potential was kept during ion excitation event, whereas it was user-defined in the range of 0–10 V during the ion detection event. Approximately 4.6-fold higher potential was applied to the sections of inner rings covered with grid. The shape of the detection electrodes in NADEL ICR cell was designed to form trapping potential distribution closer to quadratic. Based on SIMION calculations, the NADEL electrodes were shaped with a curvature radius of 125 mm:[135]

$$r(z) = r_0 + R \left(1 - \sqrt{1 - \frac{z^2}{R^2}} \right), \quad (5.2)$$

where z is an axial coordinate, r_0 is the distance from the detector electrode edge to the ICR ion trap center (axis) when $z=0$, R is the curvature radius.

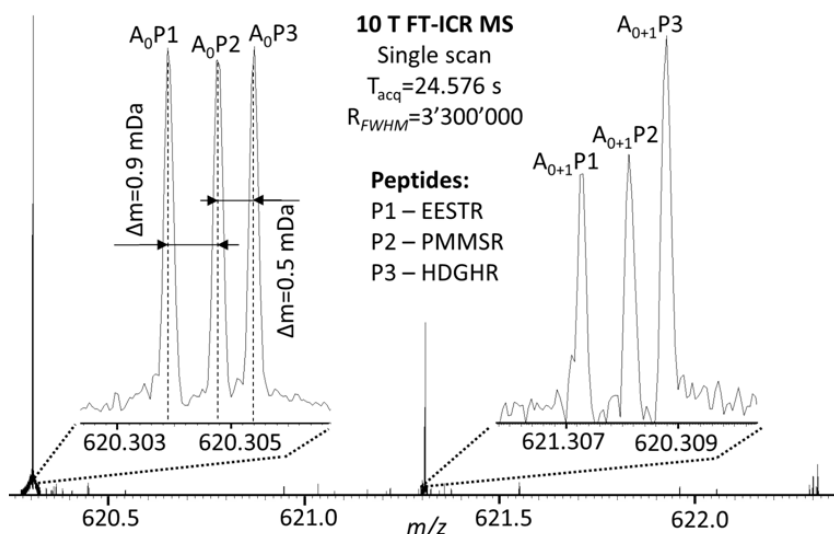


Figure 5.6: Isobaric peptide mixture measurements with a NADEL cell in 10 T FT-ICR MS. The resolving power 3'300'000 was achieved for a single scan with acquisition time of 25.576 s and number of charges (AGC) of 3E4 for isolated ions of triplet of singly protonated peptides P1, P2 and P3. Figure adapted from Paper VIII.

Ion excitation with NADEL ICR cell was technically performed in the same way as with Ultra ICR cell. Ion detection was performed using the commercial ion detection configuration of Ultra ICR cell. The transients of variable length in the range of 92 ms – 25 s were recorded at 1-5 MHz sampling frequency (narrowband ion detection) using the advanced software interface of the built-in data acquisition system (Thermo Scientific). For broadband ion detection (up to 50-100 MHz sampling frequency) a custom DAQ was employed, Chapter 2. A given number, 10–100, of transients were averaged to obtain the final transients, which were Hann-apodized and zero-filled once before fast Fourier transformation (FFT) to yield Fourier spectra which were further calibrated to mass spectra using standard data analysis software (Xcalibur, Thermo Scientific) or the framework pyFTMS developed in-house, Chapter 2. The offset DC potentials in the range of from –100 to 100 mV were applied to the NADEL and Ultra ICR cell detection and excitation electrodes independently relative to each other to efficiently vary the position of the ion cloud prior to ion excitation and during ion detection events. The influence of offset potentials on ion motion was monitored with diverse quality attributes of mass spectra, such as peak shape and magnetron sidebands. Overall, it was possible to optimize the offset potentials to effectively decrease the magnetron sideband intensities below 1 percent relative to the corresponding reduced cyclotron frequency peak for each experiment.

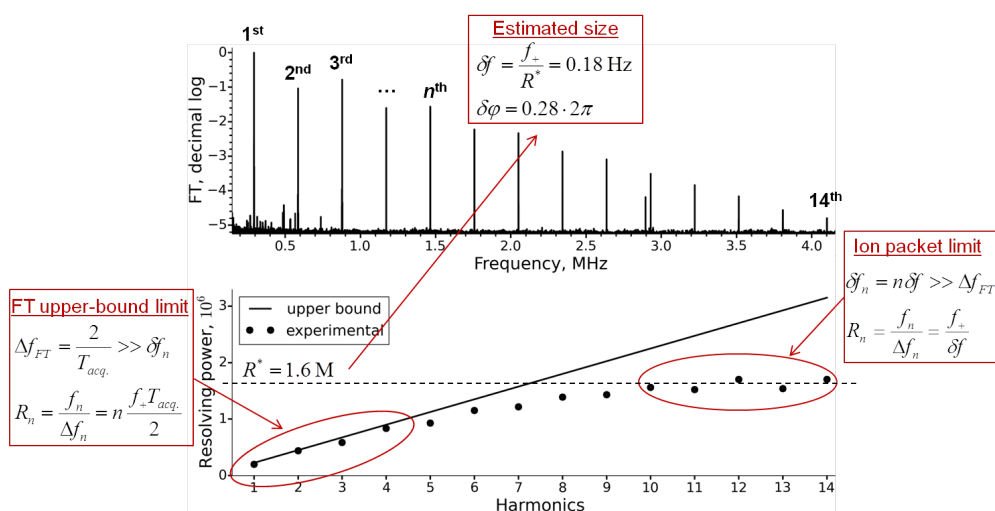


Figure 5.7: Dependence of resolving power on harmonics order in a Fourier spectrum of a transient signal acquired in the analysis of monoisotopic ions of MRFA peptide with the NADEL cell.

5.2.2 Conventional regime

The resolving power performance of the NADEL ICR cell operating in the conventional regime was evaluated for analysis of peptide and protein samples with maximum acquisition time of 25.576 s for the employed hardware configuration. Acquisition of longer transients was not possible due to the technical limitations imposed by the on-board electronics of the FT-ICR MS computer memory, whereas the lifetime of excited and coherent ion motion apparently surpasses this temporal limitation, for example see Figure S7 of Paper VIII. For instance, Figure 5.6 shows a mass spectrum of three isobaric singly protonated peptides P1, (EESTR, 619.29255 Da), P2 (PMMSR, 619.29342 Da), and P3 (HDGHR, 619.29389 Da), all positioned at m/z 620, with mass differences of 0.9 mDa and 0.5 mDa between the pairs of peptides. Notably, mass of 0.5 mDa is close to the mass of an electron. Expanded segments of the mass spectrum show baseline-resolved monoisotopic and ^{13}C peaks for the three peptides of interest with the resolving power achieving 3'300'000 for a single scan. To avoid the peak coalescence, a relatively low number of ions was employed (AGC setting of 3E4). The particular importance of the result reported is in the long time of ion cloud coherence after excitation, which allows recording of 25 s and longer transients. The reported separation of this peptide triplet can be compared to the notable example of a peptide doublet separation (mass difference close to the mass of an electron) by Marshall and co-workers on a 9.4 T FT-ICR MS.

Evaluation tests of this cell show that the combined benefits of trapping and excitation field configuration in the NADEL cell significantly improve performance of standard ICR cells in routine analysis in regard to resolving power performance and approach conditions of the most sophisticated ICR cells with complex approaches to electric field harmonization, either statically or dynamically.

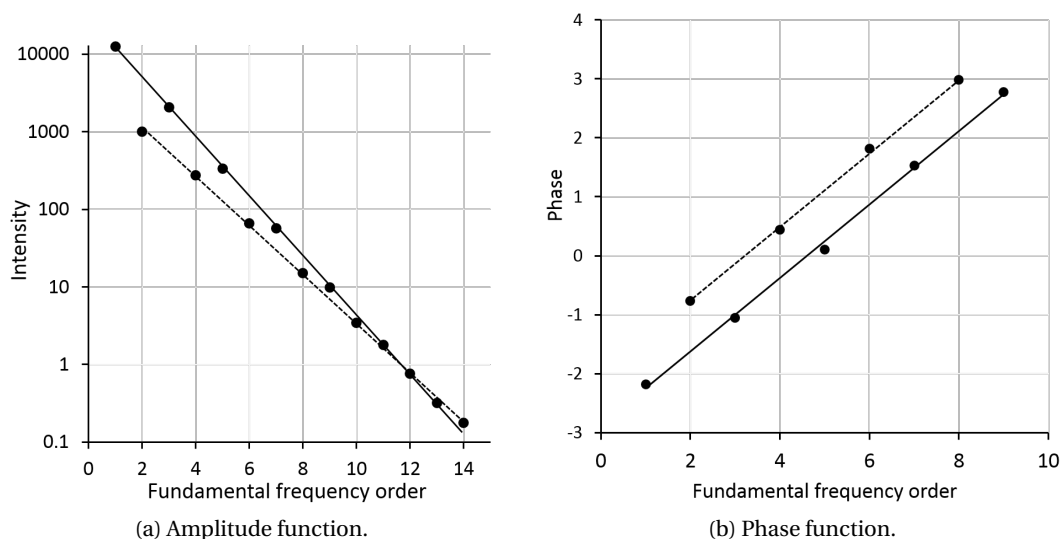
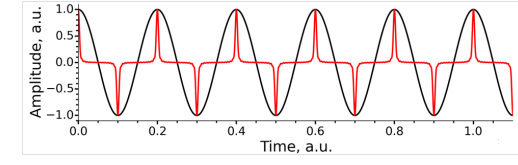


Figure 5.8: Amplitude $A(n)$ and phase $\varphi(n)$ functions of a transient signal of a monoisotopic ion of MRFA peptide. The dependencies represent amplitude and phase calibration of the NADEL cell's response function.

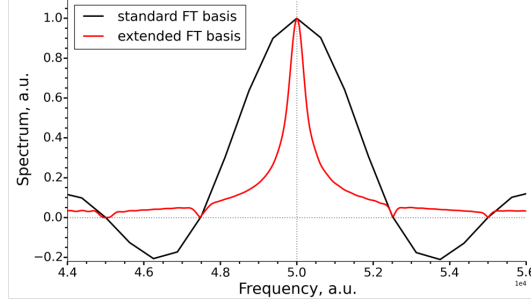
5.2.3 Increased resolution performance

Here, based on the fundamental aspects from Chapter 2, we develop an integrated approach towards increased resolution performance using specific design of a mass analyzer and tailored signal processing. Specifically, the second regime of the NADEL cell consists in accelerating ions to a high radius, leading, in terms of spectral analysis, to rich harmonic content of resulting transient signals. Hence, this regime is distinct by increased *spatial* resolution of ion packets by the detection electrodes with narrow aperture compared to the conventional 90 degrees detection electrodes. Importantly, increased spatial resolution would naturally lead to increased resolution of ion frequencies, when an appropriate method for signal processing is applied. And, as a limiting case, the performance for resolution of such a measurement scheme is restricted by the ion packet sizes in the phase space, as discussed in Chapter 2 in the context of uncertainty principle limits.

For example, when a transient signal acquired in this regime is processed with the conventional FT signal processing, Figure 5.7, the spectrum reveals a number of harmonics of the fundamental ion frequency. Dependence of the resolving power value on the harmonics order shows two limits. Data points for low orders lie on the limiting slope due to the uncertainty principle of the FT signal processing, while the points for high orders approach the limit due to the finite size of the ion packet. Hence, experimental confirmation of the both limits makes perfect sense from the point of view of considerations developed in Chapter 2. Specifically, as follows from the point of intersection of the horizontal level and FT limiting line, for the considered example as much as 7-fold gain in resolution performance can be possible.



(a) Simulated sinusoidal and sharp transient signals.



(b) Fourier spectra obtained with FT processing of the sinusoidal signal and xFT processing of the sharp signal.

Figure 5.9: Numerical evaluation of the FT method with extended basis of functions (simulated transients).

Here we develop a prototype method, based solely on Fourier transform, of signal processing to make this hidden resolution available. As follows from Chapter 2, in order to determine experimental quantities of interest, such as frequencies or amplitudes, the generalized measurement correlator takes a measured signal and compares it with an expected form of the signal to obtain the parameters of interest. Similarity between the expected form and the actual experimental signal governs the resulting limitations such as resolution and other characteristics. Therefore, to convert the increased spatial resolution into increased frequency resolution, the information on ion packets motion, encoded in the sharp signals, can be obtained using an appropriate basis of functions that well describe a single ion's transient signal, i.e. the response function of the measurement system. Hence, to describe the response function, we represent the signal from a single ion, $f(t)$ using amplitude, $A(n)$, and phase characteristics, $\varphi(n)$, i.e. the values of amplitudes and phases of spectral components plotted as functions of the frequency order n , e.g. as exemplified for the NADEL cell in Figure 5.8:

$$s(k) = \sum_{n=1}^N A(n) \cos(2\pi n f_1 t_s k + \varphi(n)) \quad (5.3)$$

where f_1 is the fundamental frequency of the ion, t_s is the sample time, and k in the index of the discrete signal $s(k)$.

Hence, the prototype method to process such transients, henceforth referred to as the xFT method, can be developed using Fourier transform with accordingly extended basis of functions compared to the basis of the standard FT. To meet Eq. 5.3, the extended basis must be the response function, with probably omitted phase of the fundamental frequency, $\varphi(1)$, but yet

preserved phase differences between the subsequent harmonics. Neglecting vs. employing the value of $\varphi(1)$ is equivalent to magnitude vs. absorption mode representations of the standard FT method, Chapter 3.

Figure 5.9 compares results obtained with the regular FT signal processing of a signal of the conventional theoretical form vs. spectra obtained with the xFT method applied to a signal of the modified form. FT processing was performed in a standard way, i.e. as correlation between the sinusoidal signal and the standard basis. The xFT processing was performed as correlation between the sharp signal and the extended basis describing such transients. The obtained spectra confirm gain in resolution performances of the xFT method vs. FT. The gain in resolution performance is defined by the uncertainty principle of this method and depends, specifically, on the amplitude characteristic.

5.2.4 Accurate mass measurements regime

In addition to the standard regime with reduced cyclotron frequency ω_+ employed for ion m/z measurements, the NADEL cell may operate in a regime such that a measured transient signal is composed of sinusoidal components with frequencies close to unperturbed cyclotron frequencies ω_c of ions, as follows from a weak dependence of the measured frequency on the trapping voltage V , compared with the ω_+ mode, Figure 5.10. Since the ω_c regime is achieved via optimizing the electrical potential on the trapping electrodes as well as the offset potentials applied to the detection and excitation electrodes, generation of the ω_c frequency is likely to be due to complex trajectory of ion motion, achieved due to the tuned potentials and specific geometry of the cell's including the shape of the detection electrodes.

Figure 5.11 compares the ω_c regime and the ω_+ regimes in the analysis of bovine ubiquitin in the broadband m/z range. With acquisition time of 1.536, all ubiquitin peaks in various charge states, ranging from 6+ to 14+, are baseline resolved. To verify lack of frequency shifts within the entire frequency range upon variation of V , data was collected for $V = 2$ V and $V = 4$ V. Figure 5.11, bottom left shows the isotopologue distribution for the 10+ charge state for the two V values from the corresponding broadband frequency spectra. No perceptible frequency shift was observed, as is expected for "unperturbed" cyclotron frequencies. Other charge state distributions within the corresponding broadband frequency spectra showed the same trend. In contrast, the frequency of all the constituent peaks of spectra acquired under similar experimental conditions, but in the ω_+ regime, was systematically decreased with increasing V value, as is expected for reduced cyclotron frequencies. Figure 5.11.

Petroleomic samples were chosen as exemplary complex mixtures for mass accuracy comparison of the two regimes. Figure 5.12 shows a positive ESI broadband mass spectrum of the maltene fraction of S. American crude oil. The sample was analyzed in the two, ω_c and ω_+ , detection regimes. 29 monoisotopic peaks of the N heteroatom class were employed for internal mass calibration using the standard two-parametric formula, $m/z = A/f + B/f^2$. [61, 62] Comparison of residual mass errors calculated for calibrants after internal

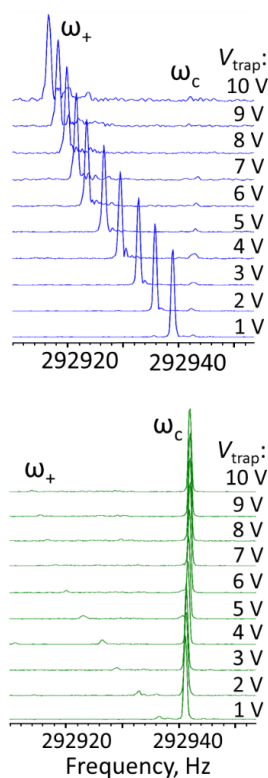


Figure 5.10: Dependence of the reduced cyclotron frequency ω_+ and the "unperturbed" cyclotron frequency ω_c on the trapping potential V , applied during detection event, for the optimized regimes: the ω_+ regime (blue lines, top) and the ω_c regime (green lines, bottom). The measurements were performed for monoisotopic ions of peptide MRFA ($524 m/z$).

calibration shows that the ω_c provides more accurate m/z values than the ω_+ . Large-scale evaluation of the two regimes using 2793 monoisotopic peaks of identified heteroatom classes shows that, indeed, the ω_c regime provides better mass measurement accuracy with the root-mean-square (RMS) error of 80 ppb, while the results of the ω_+ with the RMS error of 300 ppb are no worse than performance of state-of-the-art FT-ICR MS instruments for analysis of petroleomic samples.

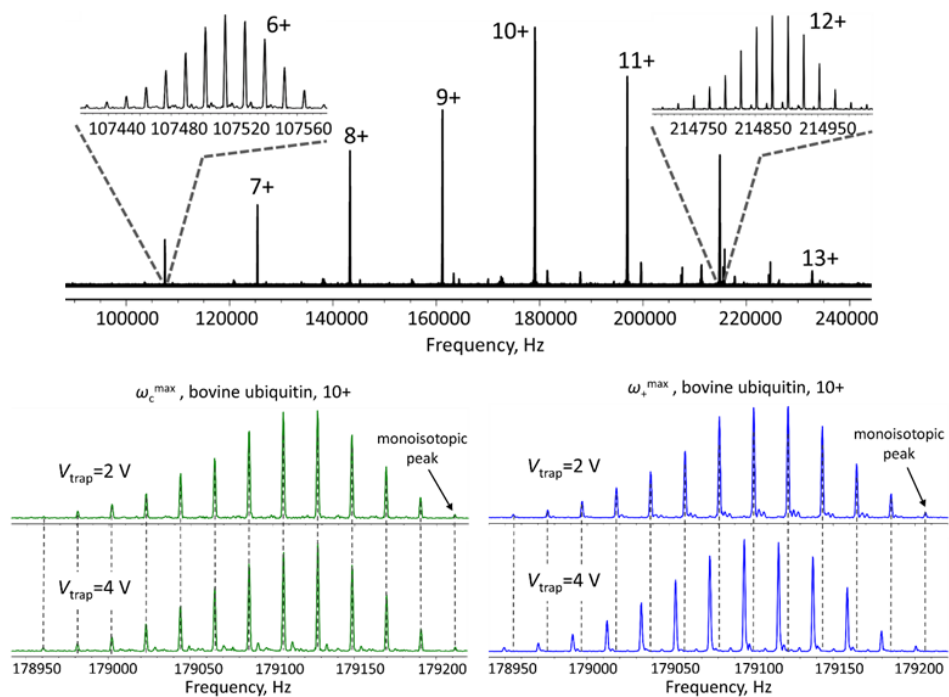


Figure 5.11: Top: The broadband frequency spectrum of bovine ubiquitin for the regime where unperturbed cyclotron frequency at maximum. The insets demonstrate baseline resolved isotopic distribution of the 6+ and 12+ charge states. Bottom: zoom-ins of 10+ charge state of bovine ubiquitin from the broadband frequency spectra for both regimes.

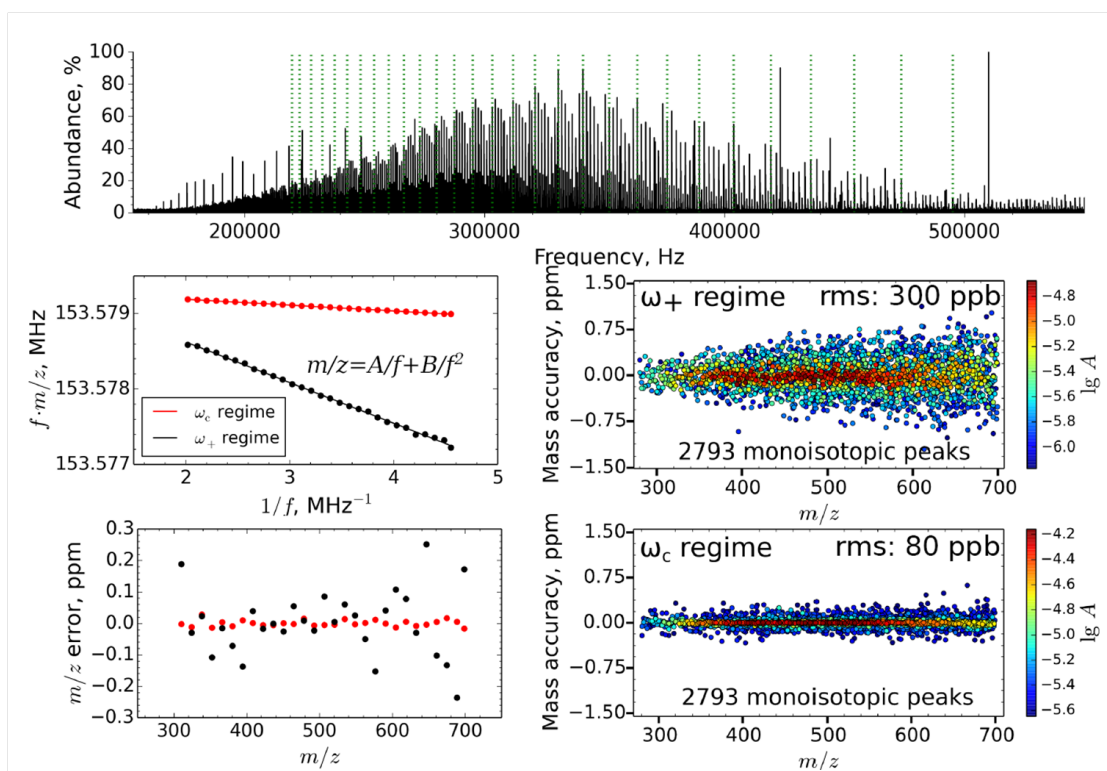


Figure 5.12: Comparison of the ω_c and ω_+ regimes in the analysis of a crude oil fraction on 10 T FT-ICR MS equipped with a NADEL cell. Top panel: Example of broadband mass spectrum. Internal mass calibration was performed using 29 ions (monoisotopic peaks) corresponding to the N heteroatom class of the sample (green lines). Left panels: Internal calibration for each of the regimes and corresponding residual mass errors. Right panels: Evaluation of mass accuracies for 2793 monoisotopic peaks of identified heteroatom classes of the sample.

Conclusions

Fulfilled goals

The present research has been dedicated to the fundamental principle behind the methodology of FTMS. The uncertainty principle in FTMS asserts a limit to the precision with which complementary physical quantities, such as detection period and scale of frequency details to resolve, can be measured. Specifically, two corollaries of the uncertainty principle have been considered, namely: resolution performance at finite detection times and performance for accurate measurements at finite signal-to-noise ratios. Based on these considerations, we have developed and successfully implemented novel schemes for measurements in FTMS aimed to advance FTMS-based molecular analysis in currently challenging applications. The implemented approaches include methods for data analysis, signal processing methods and designs of mass analyzers. It has been shown that typical values of analytical characteristics such as measurement accuracy and performance for frequency (mass) resolution are significantly limited in state-of-the-art FTMS mainly due to various parts of the default measurement scheme, whereas the ultimate analytical characteristics are governed by the physical limit defined by ion motion in the mass analyzer. The developed and implemented methodologies have been shown to enable a number of challenging analytical studies to be carried out in chemistry and biochemistry. The key developments of the present thesis are:

1. The measurement principle of FTMS has been formalized for improved understanding of the fundamental limit of analytical performance in FTMS, imposed by ion motion, Chapter 2. This in part includes Paper I in which the theory for the transient signal in FT-ICR MS is revisited to analytically describe the effect of amplitude-phase modulation of the ion trajectory.
2. The methods for improved data analysis in FTMS aimed at increased measurement accuracy of ion m/z ratios and abundances as well as increased range of analytically useful ion abundances, Chapter 3. In particular, the mass recalibration method was developed to enable comprehensive analysis of complex mixtures, such as petroleomic samples, on Orbitrap FTMS (Paper III). This method, along with the method for mass spectra thresholding (Appendix in Paper II), and the method for petroleomic data visualization[136] form the Orbitrap FTMS platform for analysis of petroleomic samples.[68]

Conclusions

3. The double phase correction method has been developed to double the resolution performance in the case of equidistant distribution of analyte peaks in the m/z (and frequency) domain, including doublets, with *a priori* known m/z spacing, Chapter 3.
4. The least-squares fitting method (LSF) for analysis of transient signals in FTMS has been implemented (Paper IV), Chapter 4, aimed specifically at improved measurement accuracies and resolution performance, and tailored to applications in instrument development and specific analytical problems such as quantitative proteomics.
5. Implementation of a high-throughput quantitative proteomics workflow with Orbitrap FTMS and LSF or double phase correction method has been evaluated in large-scale LC-MS studies, Chapter 4. Specifically, with the double phase correction method, a two-fold gain in resolution performance is provided compared to the absorption mode FT (eFT algorithm), while a 4-fold gain is routinely achievable with the LSF method.
6. Filter-diagonalization method has been implemented in FTMS (Paper V). With this method, the gain in resolution performance depends on spectral irregularity of frequency peaks for analyzed ions, Chapter 4. The target applications are qualitative molecular analysis with highly irregular distribution of analyte peaks, including analysis of isotopic fine structure analysis of peptides (Paper VI) or isobaric samples.
7. The 4X ICR cell has been developed to implement ion detection at quadruple frequency multiple for increased throughput in qualitative molecular analysis in FT-ICR MS (Paper VII). The NADEL cell has been developed to implement signal generation at "unperturbed" cyclotron frequency for accurate mass measurements, and generation of sharp transient signals for increased resolution performance in FT-ICR MS (Paper VIII).

Notably, to achieve the goal of the present research, a number of scientific fields including ion physics, mathematical analysis, theory of random data, radio-electronics, computer science, cheminformatics, and analytical chemistry were required to be involved, hence qualifying the present research to be indeed an interdisciplinary work. It is also worth mentioning that an open source code repository comprising a number of methods for data analysis and signal processing implemented during this work is made available in the public domain. Finally, the design concept of the data acquisition system developed in this work should be useful for future generations of FTMS instruments.

Future perspectives

The final objectives of these developments include areas of health, preventive medicine, clinical applications, as well as comprehensive personal medicine. Indeed, in terms of the life sciences applications, it can be expected that with the achieved analytical performance, further progress will be feasible in characterization of IgGs, leading to better drug discovery, and therefore health. In terms of personal medicine, multiplex protein quantification is a must

in modern analysis in the sense of reducing the analysis time required per a single proteome. The future perspective of this work is immediate use of these developments for applications. We consider the achieved results to have potential for being commercially adopted. For instance, at the moment we are about to employ the LSF method specifically for large-scale applications in quantitative proteomics. Likewise, in the context of the developed ICR cells, they certainly have potential for applications that are currently requiring high resolution and high measurement accuracy at short acquisition times. Currently, there is strong interest among the MS companies in regard to these designs, as they indeed represent a paradigm shift in FTMS instrumentation and methodology.

Concluding remarks

History knows a number of breakthroughs in the analytical performance of FTMS, with each of those being of a different nature. First, a significant breakthrough of the past days of FTMS was due to developments in vacuum technology, while before that the analytical performance of FTMS was limited primarily due to insufficient vacuum conditions in the mass analyzer. Next, since then and until present days, the limiting factor has been considered the magnitude of electromagnetic fields employed for confinement of ions in the mass analyzer, thus leading to continuously-repeating little increases in values of employed magnetic fields in FT-ICR MS (e.g. from 0.3 T of the first FT-ICR MS to 21 T of today's projects), and electric fields in Orbitrap FTMS.

It is also interesting to note that since the classical experiments of Comisarow and Marshall, growing requirements for analytical performance of FTMS have been supported by simultaneous developments in the hardware architectures for data acquisition. This tendency remains for the scope of the present thesis as well: combination of the new architecture for data acquisition systems and the acquisition mode using data streaming between data acquisition devices certainly makes for future generations of FTMS instruments. Moreover, given that each following increase of electromagnetic fields is getting more and more complicated, in light of the present research the next breakthrough in FTMS may likely be due to improved understanding of the underlying uncertainty principle in FTMS, resulting in rational design of measurement schemes aimed at improved analytical performance, as presented in this work. For instance, already today experimental designs of mass analyzers with harmonized trapping fields unintentionally hint that the limiting factor in the future will be ion decoherence effects of regular or even high-performance oriented ion traps, which in turn is in agreement with the results of this thesis.

Finally, once such a transition is made in FTMS of tomorrow, it may then be expected that the following breakthrough leading to next generations of FTMS instruments will be due to addressing the limitations in physics of ion confinement in future mass analyzers. Our most recent results path possible ways towards such studies by achieving improved ion coherence via developments in ion quadrupolar excitation.

A Papers

A.1 Paper I



Sidebands in Fourier transform ion cyclotron resonance mass spectra

Saša M. Miladinović^{a,1}, Anton N. Kozhinov^{a,1}, Oleg Yu. Tsybin^b, Yury O. Tsybin^{a,*}

^a Biomolecular Mass Spectrometry Laboratory, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

^b Ion Physics Laboratory, State Polytechnical University, 195251 Saint-Petersburg, Russia

ARTICLE INFO

Article history:

Received 1 May 2012

Received in revised form 2 August 2012

Accepted 6 August 2012

Available online 14 August 2012

Dedicated to Prof. Dr. Eugene Nikolaev.

Keywords:

Fourier transform mass spectrometry

Ion cyclotron resonance

Magnetron ion motion

Cyclotron ion motion

Transient signal

Sidebands

ABSTRACT

Sidebands in mass spectra are an intrinsic feature of Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS). Appearance of the sidebands there is detrimental for the analytical performance, especially in case of complex mixtures analyzed at high resolution. Yet, the sidebands have a practical potential as well. Specifically, they can be applied for fine tuning of ICR cells and were previously employed to improve mass measurement accuracy for small molecules and atoms in fundamental physics experiments. Moreover, experimental characteristics of sidebands allow evaluating the theory of the ICR signal, which provides the metrological basis in FT-ICR MS. Here, we revisit the sidebands phenomenon in the conventional FT-ICR MS, specifically applied to macromolecules. We extend the previous reports on sidebands by examining the appearance of sidebands as functions of ICR cell trapping potentials, resolution, and number of charges for the first three harmonics of the reduced cyclotron frequency. Next, we develop an analytical model of sidebands that contributes to the existing theory of the ICR signal by showing the origin of sidebands to be the result of the broadband amplitude-phase modulation occurring in the ICR signal. Finally, we evaluate the theory of the ICR signal on the basis of the obtained experimental sidebands. Further progress in the theory of the ICR signal shall outline the way for further improvements in FT-ICR MS performance.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) provides an outstanding analytical performance for accurate and high-resolution mass measurements in physics, chemistry, and life sciences [1–6]. FT-ICR MS is based on the principle that the cyclotron frequency ω_c of ion motion in the uniform magnetic field B provides the ion's mass-to-charge ratio, m/q : $\omega_c = qB/m$ [7]. In the model case of homogeneous magnetic field and quadrupolar trapping potential [7–10], the ion in the ICR cell undergoes three types of fundamental motions, viz., cyclotron motion, magnetron motion, and axial motion. Such motion is described with a system of linear equations whose eigenvalues define the reduced frequency of ion cyclotron motion (ω_+), the frequency of ion magnetron motion (ω_-), and the frequency of ion axial oscillations (ω_z). Despite such linearity of the ion motion equations, the ICR cell acts as a system with a (slightly) non-linear response, which leads to a transient signal whose frequency, and therefore mass, spectrum may contain other peaks in addition to the expected

ones, which correspond to the frequencies of the fundamental motions. Specifically, the peaks corresponding to the harmonics and interharmonics of the fundamental frequencies may appear in the frequency (mass) spectrum.

Therefore, the key principle of FT-ICR measurements is based on the ICR cell's operation, which is characterized by the cell's response function, i.e., the ICR transient signal. The first theoretical model of the ICR transient signal was proposed by Comisarow [11]. In that model, only the ion cyclotron motion was considered. Moreover, due to two-dimensional infinitely extended detection plates, the ICR cell, in fact, behaved as a linear system. Thus, the model was capable of describing only the cyclotron frequency peak in the spectrum, or, actually, the ω_+ peak, if one further takes into account the trapping field. Later, another simple model was proposed by Allemann and co-workers, who observed the sidebands $\omega = \omega_+ \pm \omega_-$ of the ω_+ peak in experimental ICR spectra [12]. To describe the sidebands, the model assumed an amplitude modulation of the signal with frequency ω_+ by the signal with frequency ω_- .

Later, more detailed approaches were developed. Nikolaev and Gorshkov considered an infinitely elongated ICR cell but with finite cross-sectional dimensions [13]. Their model, in fact, described the non-linear response of the ICR cell. Thus, the odd order harmonics $(2n+1)\omega_+$ of the reduced cyclotron frequency were predicted. Finally, Marshall and co-workers developed a theory for the ICR transient signal in case of three-dimensional finite ICR cells

* Corresponding author at: EPFL ISIC LSMB, BCH 4307, 1015 Lausanne, Switzerland. Tel.: +41 21 693 97 51; fax: +41 21 693 98 95.

E-mail address: yury.tsybin@epfl.ch (Y.O. Tsybin).

¹ These authors contributed equally.

[14]. This comprehensive theory predicted, in particular, (i) odd order harmonics $(2n+1)\omega_+$ of the reduced cyclotron frequency ω_+ , (ii) odd order harmonics $(2n+1)\omega_-$ of the magnetron frequency ω_- , and (iii) odd order interharmonics $(2n+1)\omega_+ \pm 2k\omega_-$ and $2n\omega_+ \pm (2k+1)\omega_-$ of the frequencies ω_+ and ω_- . Hereafter, referring to the even/odd parity of an interharmonic, we mean its total order, e.g., $2n+2k+1$, for the above-mentioned interharmonics. Moreover, the presence of odd order harmonics $(2n+1)\omega_+$ and interharmonics $\omega_+ \pm 2\omega_-$ (i.e., evenly spaced pair of magnetron sidebands of the reduced cyclotron frequency) was confirmed later with SIMION modeling [15].

The phenomenon of sidebands (interharmonics) is well known in FT-ICR MS-based particle physics. Sidebands received a particular attention there to provide precise mass measurements [8,16,17]. Specifically, use of sidebands allowed accurate determination of masses by measurement of the cyclotron frequency ω_c as the frequency of a sideband peak $\omega = \omega_+ + \omega_-$ in a mass spectrum, instead of its deduction from the experimental reduced cyclotron frequency ω_+ . Contrastingly, in the conventional FT-ICR MS of macromolecules, the ω_+ peak is often made to contain the main part of spectral energy compared to the other peaks, especially to the sidebands, and, thus, measurements of the ion's mass-to-charge ratio are solely based on the ω_+ peak [14,18–21]. However, the question of sidebands there has a practical potential as well, namely, either (i) to apply the sidebands for fine tuning of real ICR cells and, potentially, in higher-precision mass measurements, or (ii) to remove them in case they reduce experimental performance.

Here, we revisit the question of spectral composition of the ICR transient signal. To the best of our knowledge, experimental characterization and potential use of sidebands in the conventional FT-ICR MS of macromolecules have not been extensively studied. We report and examine the characteristics of sidebands present around not only the first, but also higher order harmonics of the reduced cyclotron frequency in the experimental spectra. Next, we develop further the analytical model of the ICR transient signal and show the origin of sidebands to be the result of the broadband amplitude-phase modulation occurring in the ICR transient signal. Finally, we evaluate the theory of the ICR transient signal on the basis of experimental results.

2. Methods

2.1. Sample preparation

The calibration mixture with caffeine, MRFA peptide, and Ultra-mark 1621 in methanol/water (50:50, v/v) solvent containing 1% of acetic acid was provided by the vendor (Thermo Scientific, Bremen, Germany). Peptide substance P was purchased from Sigma–Aldrich (Buchs, Switzerland) and used without further purification. Substance P was dissolved in water/acetonitrile (50:50, v/v) mixture containing 0.5% of formic acid to the final concentration of approximately 1 μ M.

2.2. Mass spectrometry

The experiments were performed on a hybrid linear ion trap Fourier transform ion cyclotron resonance mass spectrometer (LTQ FT-ICR MS, Thermo Scientific, Bremen, Germany) equipped with a superconducting magnet (Oxford Nanoscience, Oxon, UK), electrospray ionization source, and operated by the standard data acquisition and instrument control software (XCalibur, Thermo Scientific) [22]. The experiments were conducted either at 11.5 T or 10 T magnetic fields, see the values below specified for the experimental data. The isolation of the ions in question was achieved in the linear ion trap using the isolation window of 5 m/z . Automatic

gain control (AGC) was employed to control the number of charges injected into the ICR cell. After isolation, ions were transferred to the ICR cell using a set of RF-only multipole ion guides.

The open-ended cylindrical ICR cell (Ultra Cell, Thermo Scientific) employed in the measurements contains two pairs of trapping ring electrodes, namely, outer trapping rings and inner trapping rings [2]. The outer trapping rings were grounded [2]. The potential of the two grid-free sections of each inner trapping ring was manipulated in the range of 0.0–2.0 V using the instrument control software, whereas the potential of the other two sections (covered with grid) was automatically set by the instrument control software to about 5 times the value of the selected inner trapping rings potential. During the ion trapping event, the grid-free electrodes were fixed to a potential of 3 V (consequently, potential of the grid-covered electrodes would rise to about 15 V), whereas immediately after frequency-sweep ion excitation, the trapping potentials were set back to the original values [22]. The offset potentials on the ICR cell detection and excitation electrodes were preliminary tuned to reduce the magnetron sidebands to the vendor-specified standard operation conditions.

Standard dipolar differential detection was employed to obtain ICR transient signals. The transient signals were recorded for a period of approximately 6 s in MIDAS data format [23]. Further, a number of single scans was summed to obtain the final transient signal, which was apodized with the Hann window, zero-filled once, and fast Fourier transformed using MIDAS data analysis software [23]. The reported frequency values were the averages of at least 10 measurements with the standard deviation of less than 9×10^{-5} Hz. The plots were created using the Origin Pro 8G software (OriginLab Corp., Northampton, MA).

3. Results and discussion

3.1. Observation of harmonics and sidebands

A broadband FT-ICR mass spectrum of a monoisotopic ion of singly protonated MRFA peptide (nominal m/z 524) is shown in Fig. 1. The spectrum contains peaks at $n\omega_+$, $n = 1, 2, 3$, etc., corresponding to the n th harmonics of the reduced cyclotron frequency ω_+ , and their magnetron sidebands at $\omega_{nk} = n\omega_+ \pm k\omega_-$, $k = 1, 2, 3$, etc. (see Section 3.3 for details on sidebands identification). The high order harmonics and the sidebands have low intensities compared to that of the ω_+ peak. Moreover, intensities of the sidebands ω_{nk} , relative to that of the corresponding harmonics peaks $n\omega_+$, increase with the order n of the harmonics. The high resolution FT mass spectrum (resolving power, $RP > 100,000$ at $m/z = 400$) reveals the magnetron sidebands around the first harmonic as symmetric pairs of peaks shifted equally above and below the ω_+ peak by frequency difference, Δf , equal to 1.6 Hz for the employed experimental parameters. The frequency ranges around the second and third harmonics show $2\omega_+$ and $3\omega_+$ peaks surrounded by their sidebands. The intensities distribution around the second harmonic is apparently asymmetrical in contrast to the first and third harmonics. Importantly, the sideband peak spacings Δf are the same for all harmonics.

3.2. Harmonics and sidebands display at different resolution

Depending on the ICR cell offset potentials and excitation amplitude, the sidebands' intensities are ca. 5–50% of the ω_+ peak intensity. The sidebands are observed when sufficient resolution is achieved. In principle, their intensities may be reduced to less than 5% of intensity of the first harmonic in the mass spectra using appropriate offset potentials, excitation amplitude, and excitation delays. Upon variation of resolution of the mass spectrum,

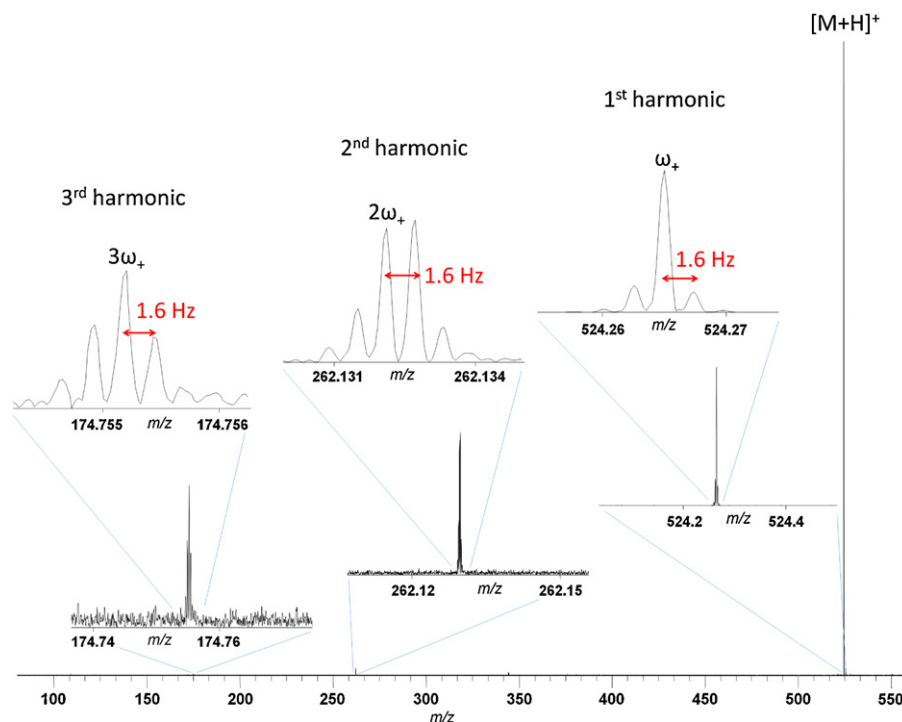


Fig. 1. Harmonics and sidebands in FT-ICR mass spectrum of monoisotopic ion of singly protonated MRFA peptide. Insets show the expanded regions around the 1st, 2nd, and 3rd harmonics of the reduced cyclotron frequency. The magnetic field was 11.5 T, the trapping potential was 0.8 V, and ion excitation amplitude was $105 V_{p-p}$ (35% of the maximum available value). The number of charges accumulated in the linear ion trap for subsequent injection into the ICR trap was set to 10^6 (AGC value).

intensity variation of the sidebands was observed (Fig. 2). Even in the case of nearly symmetric sidebands, the intensity variation was asymmetric, which is presumably due to the spectral interference in magnitude mode FTMS [24]. Fig. 2A represents 10 T FT-ICR mass spectrum of singly protonated MRFA peptide recorded with a trapping potential, U_t , of 0.8 V. Processing shorter parts of the same transient created the lower resolution mass spectra. At lower resolving power, RP, of 170,000, sidebands start to merge with the main peak producing the “tailing” effect and at RP=40,000 sidebands cannot be resolved anymore. The 10 T FT-ICR mass spectrum of the same sample recorded with $U_t = 1.6$ V is shown in Fig. 2B. As expected, the $\Delta f = 4.6$ Hz in these spectra was larger than for the spectra in Fig. 2A ($\Delta f = 2.0$ Hz) because of the difference in the values of trapping potential. Due to the larger Δf , the “tailing” and merging of the signals occurred at lower RP compared to frequency spectra recorded with lower U_t . Note, for a quadrupolar trapping potential, which is achievable, for example, in the Penning (hyperbolic) trap, a linear dependence of the frequency spacing, Δf , on the applied trapping potential, U_t , is expected theoretically [7–9]:

$$2\pi\Delta f = \omega_- = \frac{\omega_z^2}{2\omega_c} = \alpha \frac{U_t}{B}$$

where α is a constant related to geometry of the ICR cell. Thus, the deviation from this linear law indicates that in the used ICR cell the trapping field contains a certain perturbation of the ideal trapping field with quadrupolar potential, which is consistent with the cylindrical shape of the ICR cell.

In a complementary experiment, the Δf increased with the decrease in the magnetic field, e.g., in $B_1 = 11.5$ T magnetic field Δf was ~ 1.6 Hz for $U_t = 0.8$ V (Fig. 1), whereas for $B_2 = 10$ T magnetic field the Δf value was ~ 2.0 Hz (Fig. 2A). Here, the inverse dependence of the frequency spacing, Δf , on the magnetic field, B , is expected. Again, the deviation from this law is the subject of the

3.3. Influence of trapping potentials on harmonics and sidebands

The influence of the trapping potential on the first harmonic of the reduced cyclotron frequency and its sidebands was further examined with singly protonated MRFA peptide and doubly protonated peptide substance P (nominal m/z 674) (Fig. 3). The mass spectra were obtained in the 10 T magnetic field environment with the excitation amplitude of $240 V_{p-p}$ (80% of the maximum available value of the instrument control software); the number of charges before ion transfer to the ICR cell was approximately 10^6 (AGC value). For both peptides, the ω_+ peak shifted linearly to lower frequencies with an increase of U_t , whereas the frequency differences Δf showed linear increase. For both MRFA and substance P, the trend lines that connect series of the same peaks with different U_t intercept each other nearly at the same point (Fig. 3). Specifically, the cross point of the three lines for MRFA experiment was at $U_t = 0.03$ V and for substance P at $U_t = 0.04$ V. At this point, the total electric potential produced by both the trapping field and the ion space charge field is expected to be close to zero, and, thus, the frequency ω_+ is close to the cyclotron frequency ω_c .

3.4. Influence of number of charges on harmonics and sidebands

Influence of the number of charges on behavior of the first harmonic ω_+ and its sidebands is shown in Fig. 4 as trend lines of the trapping voltage U_t versus the frequencies of the ω_+ , $\omega_+ + \omega_-$, and $\omega_+ - \omega_-$ peaks for the MRFA peptide (Fig. 4A) and substance P (Fig. 4B) measured with different ion numbers (the total charge Q was controlled by AGC settings). The frequencies of the peaks are shifted to the lower values with an increase in the number of ions, whereas the Δf increases. Importantly, the frequency of the higher frequency sideband ω_R is nearly constant with the trapping potential change, and decreases when the total charge Q of

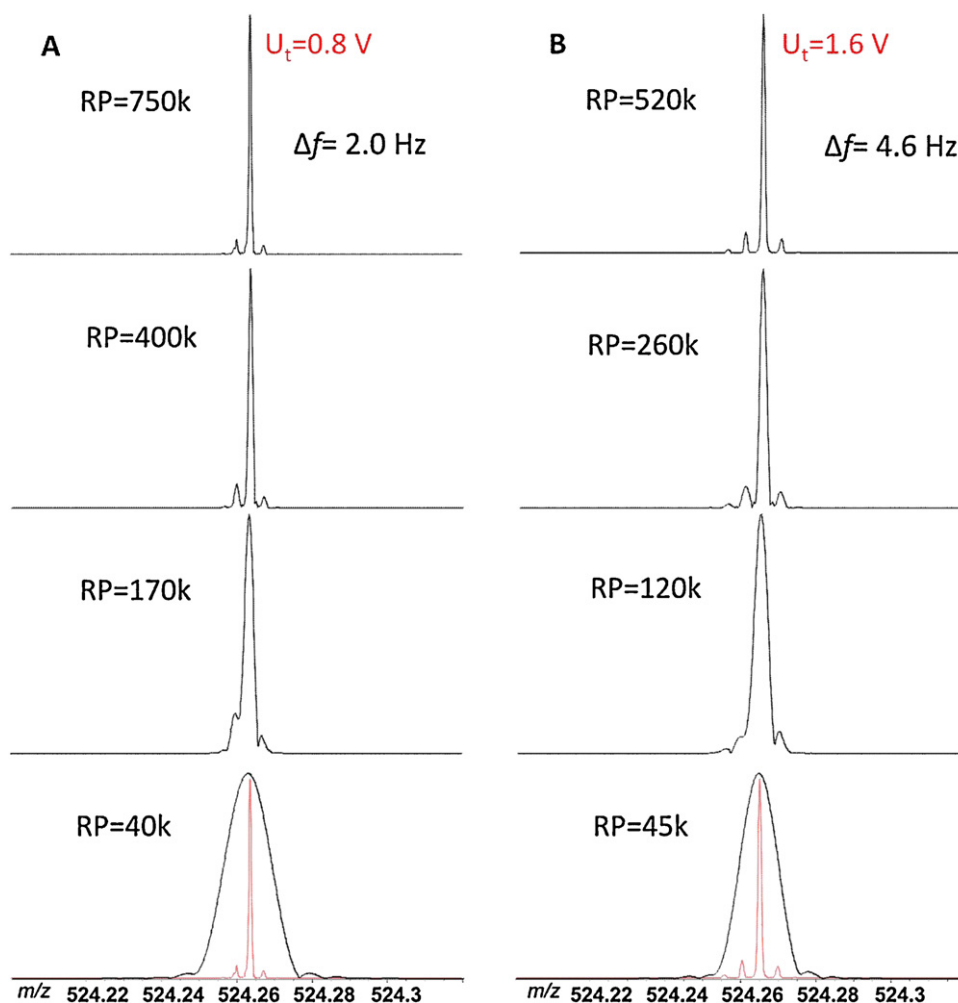


Fig. 2. Display of the ω_+ peak and its sidebands in FT-ICR mass spectra of monoisotopic ion of singly protonated MRFA peptide recorded at different resolution settings. The magnetic field was 10 T. The trapping potential was 0.8 V (panel A) and 1.6 V (panel B). In bottom panels high resolution mass spectra (in red) were plotted against the low-resolution mass spectra. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

ions increases, $\omega_R = \omega_c - \delta\omega(Q)$, $0 < \delta\omega \ll \omega_c$. The sidebands' Δf values are the same for MRFA peptide and substance P, as well as eight Ultramark calibrants in the 1022–1722 m/z range (data for Ultramark not shown).

3.5. Sidebands origin as a result of broadband amplitude-phase modulation

To obtain the ICR frequency spectrum, the ion motion equation may be solved to get the ion trajectory, and then, for the determined trajectory, the corresponding transient signal may be obtained and further expanded into Fourier series to get the ICR frequency (mass) spectrum. For the first part of this outline, we consider the approximation of quadrupolar trapping potential to have the ion trajectory in a simple and making sense form. Though such approach is justified due to a number of other approximations below, we warn further studies to consider the ion trajectory in exact trapping potential if necessary.

In cylindrical coordinates $\{\rho, \varphi, z\}$, for the magnetic field $\vec{B} = |0, 0, B_z|$ and trapping potential $\Phi(\vec{\rho}, z) = V[d + a(2z^2 - \vec{\rho}^2)]$, where $a > 0$ and $d > 0$ are dimensional constants and V is the potential on the trapping electrodes, and $\vec{\rho} = |\rho, \varphi, 0|$ is the projection of the radius-vector of the ion to the x - y plane, the motion of the ion

with mass m and charge q is described by the following system of equations:

$$\begin{cases} m\ddot{z} = -q\nabla_z\Phi(\vec{\rho}, z) \\ m\ddot{\vec{\rho}} = -q\nabla_{\vec{\rho}}\Phi(\vec{\rho}, z) + q\dot{\vec{\rho}} \times \vec{B} \end{cases} \quad (1)$$

In Cartesian coordinates $\{x = \rho \cos \varphi, y = \rho \sin \varphi\}$, the solution of this system is [8–10]:

$$\begin{aligned} x(t) &= \rho_+ \cos(\phi_+ - \omega_+ t) + \rho_- \cos(\phi_- - \omega_- t) \\ y(t) &= \rho_+ \sin(\phi_+ - \omega_+ t) + \rho_- \sin(\phi_- - \omega_- t) \\ z(t) &= z_0 \cos(\omega_z t + \phi_z) \end{aligned} \quad (2)$$

where ρ_+ and ρ_- are the radii of cyclotron and magnetron motions, ϕ_+ and ϕ_- are the initial phases of cyclotron and magnetron motions, z_0 and ϕ_z are the amplitude and initial phase of axial oscillations, and t is the time. These six parameters are constants related to the three coordinates and the three velocity components of the ion at a given moment of time, for example, $t=0$, and are defined by the excitation event. In Eq. (2), we have appropriately chosen the signs of these six parameters to conform their physical meanings with the proper sense of a positive ion's motion in the selected reference frame.

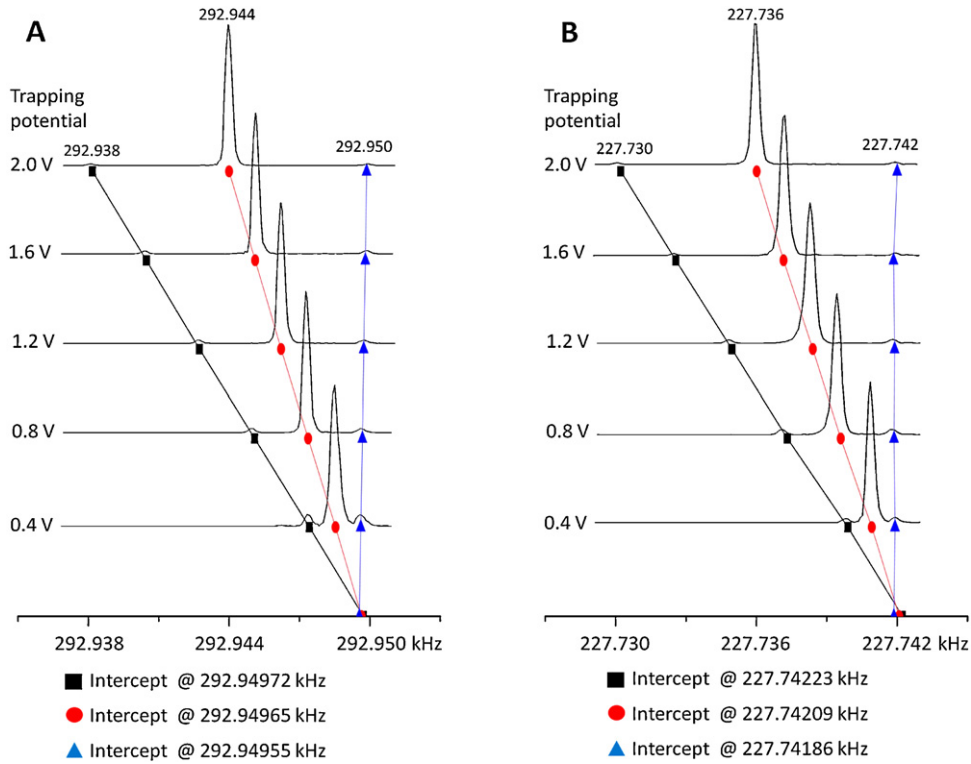


Fig. 3. The 1st harmonic of the reduced cyclotron frequency and its sidebands as a function of the trapping potential in FT-ICR frequency spectrum of monoisotopic ion of singly protonated MRFA peptide (panel A) and doubly protonated peptide substance P (panel B). The magnetic field was 10 T. The excitation amplitude was 240 V_{p-p} (80% of the maximum available value). The number of charges accumulated in the linear ion trap for subsequent injection into the ICR trap was set to 10⁶ (AGC value).

Ion motion in the x - y plane is independent of that along the z -axis. Thus, the planar motion can be considered separately. We have found it convenient to develop the model for the ion trajectory in cylindrical coordinates. A complex radius-vector may be introduced to describe the projection of the ion trajectory on the x - y plane:

$$\hat{\rho}(t) = x(t) + jy(t) \quad (3)$$

According to Eq. (2), the complex radius-vector described in Eq. (3) is:

$$\hat{\rho}(t) = \rho_+ e^{j(\phi_+ - \omega_+ t)} + \rho_- e^{j(\phi_- - \omega_- t)} \quad (4)$$

Additionally, the radius-vector may be expressed in cylindrical coordinates as:

$$\hat{\rho}(t) = \rho(t) e^{j\varphi(t)} \quad (5)$$

where $\rho(t)$ and $\varphi(t)$ are the ion's radial and azimuthal coordinates respectively. Based on Eq. (5), the radial and azimuthal coordinates can be written as follows:

$$\rho(t) = |\hat{\rho}(t)| = \sqrt{\hat{\rho}^*(t)\hat{\rho}(t)} \quad (6)$$

where $\hat{\rho}^*(t)$ is the complex conjugate of $\hat{\rho}(t)$

$$\varphi(t) = \arg \hat{\rho}(t) = \frac{1}{j} \ln \frac{\hat{\rho}(t)}{\rho(t)} \quad (7)$$

Now, based on Eqs. (6) and (4), the radial coordinate is:

$$\rho(t) = \sqrt{\rho_+^2 + \rho_-^2 + 2\rho_+\rho_- \cos(\Delta\omega t - \Delta\phi)} \quad (8)$$

where $\Delta\omega = \omega_+ - \omega_-$ and $\Delta\phi = \phi_+ - \phi_-$. The frequency $\Delta\omega$ is known as the parametric frequency of ion motion [25,26]. From Eqs. (7), (8) and (4), the azimuthal coordinate is:

$$\varphi(t) = \frac{1}{j} \ln \left\{ \frac{\rho_+ e^{j(\phi_+ - \omega_+ t)} + \rho_- e^{j(\phi_- - \omega_- t)}}{\sqrt{\rho_+^2 + \rho_-^2 + 2\rho_+\rho_- \cos(\Delta\omega t - \Delta\phi)}} \right\} \quad (9)$$

Thus, we have obtained the ion trajectory $\{\rho(t), \varphi(t)\}$, Eqs. (8) and (9), in cylindrical coordinates.

Post-excitation ion trajectories are characterized by a $\rho_+ \gg \rho_-$ condition. Therefore, the trajectory described by Eqs. (8) and (9) can be further linearized. Here, we neglect the second and higher order terms of the small parameter $\varepsilon = \rho_-/\rho_+ \ll 1$, see Eqs. (A1)–(A6):

$$\rho(t) \cong \rho_+ [1 + \varepsilon \cos(\Delta\omega t - \Delta\phi)] \quad (10)$$

$$\varphi(t) \cong \varphi_+ - \omega_+ t + \varepsilon \sin(\Delta\omega t - \Delta\phi) \quad (11)$$

In this way, we have found the linearized ion trajectory $\{\rho(t), \varphi(t)\}$, Eqs. (10) and (11). The obtained trajectory is a circular orbit $\{\rho = \rho_+, \varphi(t) = \phi_+ - \omega_+ t\}$ perturbed by the curve $\{\Delta\rho(t) = \rho_+ \varepsilon \cos(\Delta\omega t - \Delta\phi), \Delta\varphi(t) = \varepsilon \sin(\Delta\omega t - \Delta\phi)\}$ (Fig. 5). Thus, the ion trajectory is a modulated circular orbit. Indeed, the ion trajectory shows azimuthal deviations $\Delta\varphi(t) = \varepsilon \sin(\Delta\omega t - \Delta\phi)$ around the $\phi_+ - \omega_+ t$ value as well as radial deviations $\Delta\rho(t) = \rho_+ \varepsilon \cos(\Delta\omega t - \Delta\phi)$ around the ρ_+ value oscillating with the parametric frequency $\Delta\omega$ in both cases.

Now, we may proceed to obtaining the ICR transient signal. If $\Delta Q(\rho, \varphi, z)$ is the difference of charges induced on the detection plates by the ion at point $\vec{\rho} = |\rho \varphi z|$ in the ICR cell and with charge q , then the ICR transient signal is given by the following equation [27], disregarding a decaying exponential term in the beginning of the signal and an arbitrary phase spectrum of the signal:

$$\Delta V(t) = \frac{q}{C} \frac{\Delta Q(\rho, \varphi, z)}{q} \quad (12)$$

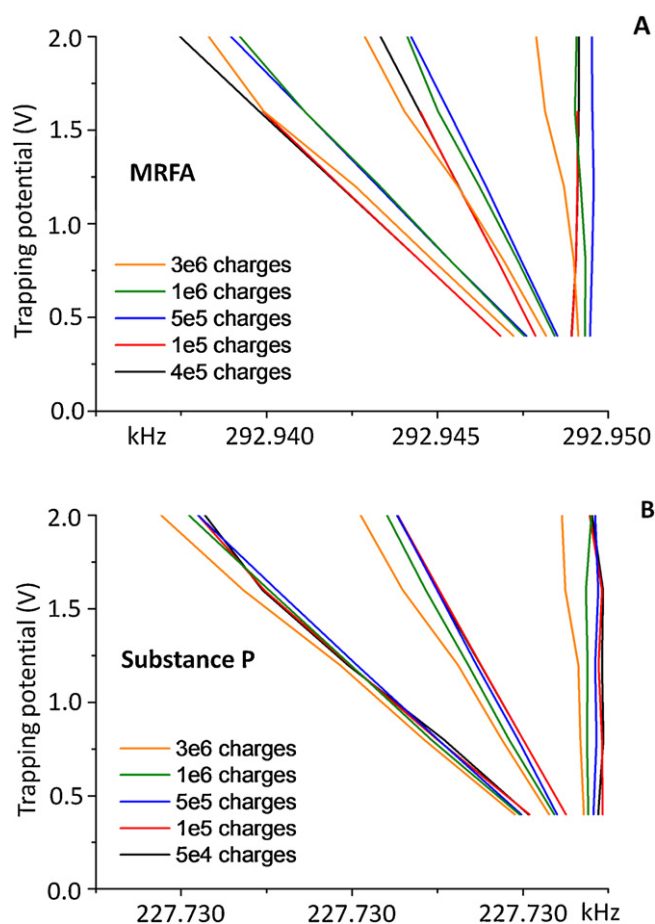


Fig. 4. The 1st harmonic of the reduced cyclotron frequency and its sidebands as a function of the number of charges and trapping potentials in FT-ICR frequency spectrum of monoisotopic ion of singly protonated MRFA peptide (panel A) and doubly protonated peptide substance P (panel B).

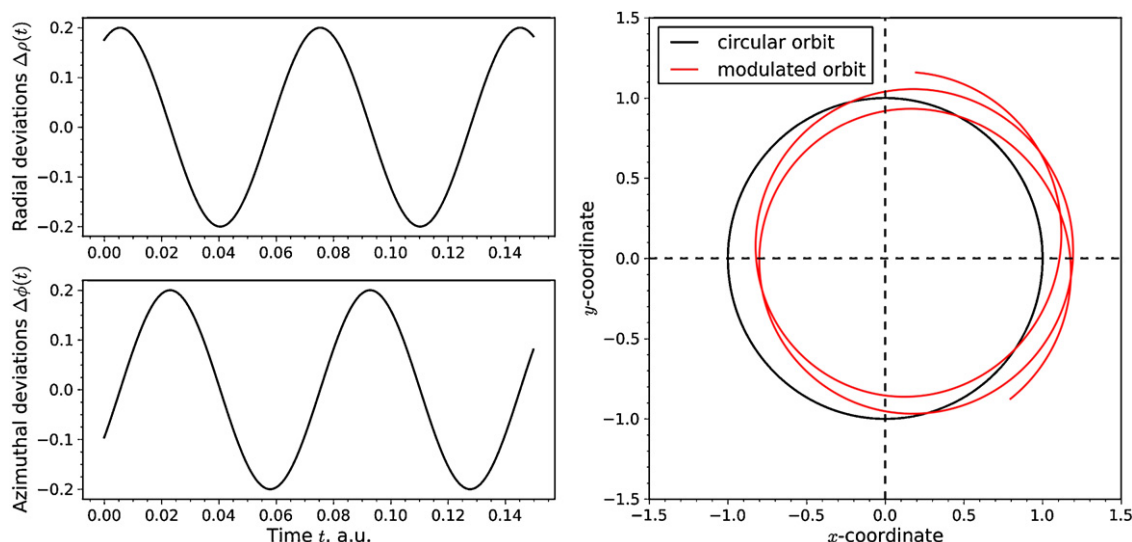


Fig. 5. Numerical simulation of the linearized ion trajectory described by Eqs. (10) and (11). Top left panel: radial deviations around the ρ_+ value; bottom left panel: azimuthal deviations around the $\phi_+ - \omega_+ t$ value; right panel: the ion trajectory (red) as a result of the radial-azimuthal modulation of the circular orbit (black). The following parameters were used: $\omega_+ = 100$ a.u., $\Delta\omega = 90$ a.u., $\varepsilon = 0.2$, $\phi_+ = 1.5$, $\Delta\phi = 0.5$, $\rho_+ = 1.0$ a.u., $\rho_+/a = 0.8$. Duration of the calculated ion motion was $T = 0.15$ a.u. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

where C is a dimensional constant (effective capacitance). An analytical expression for the relative charge difference is not available for the open ICR cell with trapping ring electrodes. However, the relative charge difference is known for a closed cylindrical ICR cell and can be written as [14]:

$$\frac{\Delta Q(\rho, \varphi, z)}{q} = \frac{-16}{\pi^2} \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} \frac{\sin[(2m+1)(\pi/4)] \cos[(2m+1)\varphi]}{2m+1} \times \frac{\sin[(2k+1)(\pi z/c)] I_{2m+1}[(2k+1)(\pi\rho/c)]}{2k+1 I_{2m+1}[(2k+1)(\pi a/c)]} \quad (13)$$

where $I_m(k)$ are the modified Bessel functions of the first kind, provided that the trapping electrodes are located at $z=0$ and $z=c$, whereas the detection electrodes are located at $\rho=a$, $-\pi/4 < \varphi < \pi/4$, and $3\pi/4 < \varphi < 5\pi/4$. We perform a shift of z -axis to match its $z=0$ point with the center of the cell: $z \rightarrow z+c/2$, so that Eq. (13) takes the following form [28]:

$$\frac{\Delta Q(\rho, \varphi, z)}{q} = \frac{-16}{\pi^2} \sum_{m,k=0}^{\infty} \frac{\sin[(2m+1)(\pi/4)] \cos[(2m+1)\varphi]}{2m+1} \times \frac{\cos[(2k+1)(\pi z/c)] I_{2m+1}[(2k+1)(\pi\rho/c)]}{(-1)^k (2k+1) I_{2m+1}[(2k+1)(\pi a/c)]} \quad (14)$$

Next, we simplify Eq. (14) in the approximation of an ICR cell infinitely elongated along the z -axis with finite axial ion motion $c \rightarrow \infty$, $|z| \ll c$ [29]:

$$\frac{\Delta Q(\rho, \varphi)}{q} \cong \sum_{m=0}^{\infty} a_m \left(\frac{\rho}{\rho_+}\right)^{2m+1} \cos[(2m+1)\varphi] \quad (15)$$

where $a_m = (-4/\pi) \sin[(2m+1)(\pi/4)] / (2m+1) (\rho_+/a)^{2m+1}$. In the considered approximation, we expect the relative charge difference for the open ICR cell to be the same as that obtained above for the closed ICR cell. Now, by taking into account the relative charge difference Eq. (15) and ion trajectory Eqs. (10) and (11), the transient signal Eq. (12) can be written in a form of amplitude-phase modulation:

$$\Delta V(t) \cong \frac{q}{C} \sum_{m=0}^{\infty} a_m AM_m(t) PM_m(t) \quad (16)$$

$$AM_m(t) = \left(\frac{\rho}{\rho_+} \right)^{2m+1} = [1 + \varepsilon \cos(\Delta\omega t - \Delta\phi)]^{2m+1} \quad (17)$$

$$PM_m(t) = \cos[(2m+1)\varphi] \\ = \cos[(2m+1)(\phi_+ - \omega_+ t + \varepsilon \sin(\Delta\omega t - \Delta\phi))] \quad (18)$$

where $AM_m(t)$ and $PM_m(t)$ represent the amplitude modulation and the phase modulation terms of the m th order. Thus, the oscillating radial and azimuthal deviations of the ion trajectory lead to, respectively, the amplitude and the phase modulations occurring in the transient signal.

Finally, we aim to establish the relation between the amplitude-phase modulation and the appearance of sidebands in the spectrum. In the considered case of $\varepsilon \ll 1$, the modulation expressions Eqs. (17) and (18) can be approximated as follows, see Eqs. (A7)–(A16):

$$AM_m(t) \cong 1 + \varepsilon(2m+1)\cos(\Delta\omega t - \Delta\phi) \quad (19)$$

$$PM_m(t) \cong \cos[(2m+1)(\phi_+ - \omega_+ t)] + \frac{\varepsilon(2m+1)}{2} \\ \times \sum_{k=\pm 1} \text{sgn}(k) \cos[(2m+1)(\phi_+ - \omega_+ t) + k(\Delta\omega t - \Delta\phi)] \quad (20)$$

In the following text, we refer to the peaks only at positive frequencies due to complex symmetry of Fourier transform of a real-valued signal. With taking into account Eqs. (19) and (20), the Fourier series expansion of Eq. (16) provides peaks at $\omega = (2m+1)\omega_+$ and their parametric frequency sidebands $\omega_{PM} = \omega \pm \Delta\omega$ due to the phase modulation Eq. (20), whereas for each of these peaks, the amplitude modulation Eq. (19) provides the parametric frequency sidebands $\omega_{AM} = |\omega_{PM} \pm \Delta\omega|$. Thus, the full set of the peaks due to the amplitude-phase modulation is:

$$\omega_{APM} = \{(2m+1)\omega_+, (2m+1)\omega_+ \pm \Delta\omega, |(2m+1)\omega_+ \pm 2\Delta\omega|\} \quad (21)$$

In accordance with the parametric frequency definition $\Delta\omega = \omega_+ - \omega_-$, the set Eq. (21) can be further expanded:

$$\omega_{APM} = \left\{ \begin{array}{l} (2m+1)\omega_+, (2m+2)\omega_+ - \omega_-, 2m\omega_+ + \omega_-, \\ (2m+3)\omega_+ - 2\omega_-, |(2m-1)\omega_+ + 2\omega_-| \end{array} \right\} \quad (22)$$

The relative intensities of the peaks are defined by (i) the amplitudes a_m in the series Eq. (16), (ii) the modulation depths $\varepsilon(2m+1)$ in the modulation laws Eqs. (17) and (18), and (iii) the initial phases ϕ_+ and ϕ_- . Therefore, the intensities can be found accordingly. It is worth noting that certain spectral components with frequencies from the set Eq. (22) may, in fact, turn out to have zero intensity. The regions of the 1st, 2nd, and 3rd harmonics of the frequency ω_+ corresponding to Fig. 1 can be obtained from the set Eq. (22) for $m=0, 1, 2$. The following peaks are provided:

$$\begin{aligned} m=0: & \quad \omega_+, 2\omega_+ - \omega_-, \omega_-, 3\omega_+ - 2\omega_-, \omega_+ - 2\omega_- \\ m=1: & \quad 3\omega_+, 4\omega_+ - \omega_-, 2\omega_+ + \omega_-, 5\omega_+ - 2\omega_-, \omega_+ + 2\omega_- \\ m=2: & \quad 5\omega_+, 6\omega_+ - \omega_-, 4\omega_+ + \omega_-, 7\omega_+ - 2\omega_-, 3\omega_+ + 2\omega_- \end{aligned} \quad (23)$$

Regarding the regions of interest, the set Eq. (23) contains the following harmonics and interharmonics:

$$\omega_+, 3\omega_+, \omega_+ \pm 2\omega_-, 2\omega_+ \pm \omega_-, 3\omega_+ \pm 2\omega_- \quad (24)$$

This set of peaks, Eq. (24), is consistent with the theory of the transient signal [14]. The numerical simulation of the procedure conducted above is presented in Fig. 6 (top), which shows the magnitude mode Fourier spectrum of the signal Eq. (16) with

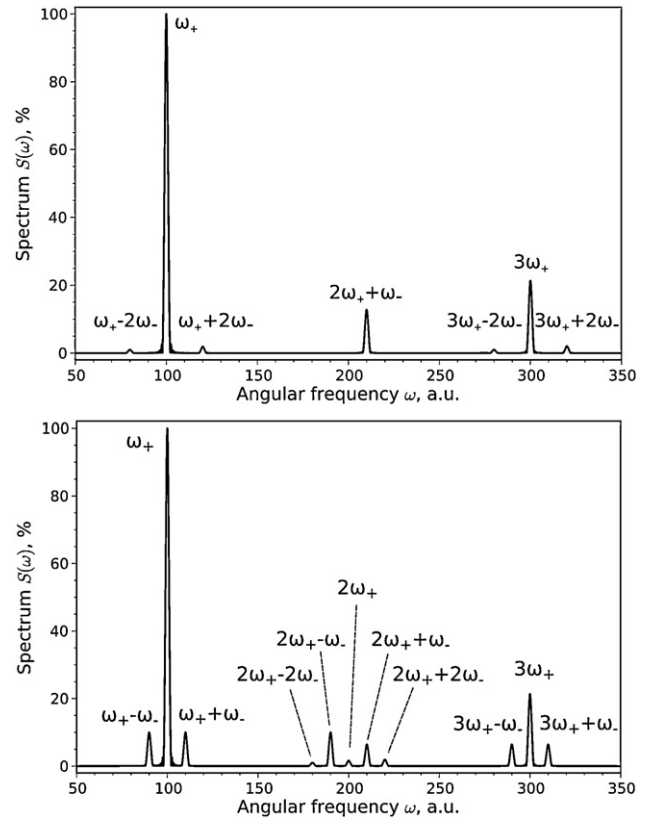


Fig. 6. Fourier spectra of the numerically simulated transient signals according to Eq. (16) with the phase modulation Eq. (20) and (top) with the amplitude modulation Eq. (19), and (bottom) with the amplitude modulation Eq. (19). Parameters of the models were the same as for Fig. 5. The length of the signal was 6 a.u. Hann apodization window, three zero-fillings, and magnitude mode spectral representation with normalization by the most intense peak were employed.

the amplitude modulation Eq. (19) and the phase modulation Eq. (20). The frequency spectrum in Fig. 6 (top) is consistent with the obtained above set of peaks Eq. (24).

Thus, the physical origin of the sidebands has been shown to be as follows: (i) finite cross-sectional dimensions of the ICR cell lead to its slightly non-linear response; (ii) if the ion magnetron motion is excited along with the cyclotron motion, the ion trajectory shows azimuthal and radial deviations oscillating with the parametric frequency around the ideal circular orbit; (iii) the oscillating azimuthal and radial deviations result in broadband amplitude and phase modulations, respectively, occurring in the ICR transient signal; and (iv) the amplitude-phase modulation produces interharmonics of the reduced cyclotron and magnetron frequencies and thus, specifically, the magnetron sidebands of the harmonics of the reduced cyclotron frequency.

Besides, in the limiting case of $\varepsilon = 0$ (circular trajectory), the transient signal, Eqs. (16)–(18), is consistent with the previous studies considered circular ion motion [13,14,29]:

$$\Delta V(t) = \frac{q}{C} \sum_{m=0}^{\infty} a_m \cos(2m+1)\omega_+ t \quad (25)$$

In light of our approach, the corollary Eq. (25) shows that if the magnetron motion is not excited, the amplitude-phase modulation does not take place so that the ICR cell does not provide the interharmonics (sidebands).

3.6. Experimental vs. theoretical approaches

For the conventional FT-ICR MS, the analytical and numerical studies predict presence of odd order harmonics and

interharmonics of the reduced cyclotron and magnetron frequencies [14,15]. The obtained set of peaks, ω_+ , $\omega_+ \pm 2\omega_-$, $2\omega_+ \pm \omega_-$, $3\omega_+$ and $3\omega_+ \pm 2\omega_-$, in the broadband amplitude-phase modulation approach developed above is consistent with the theory of the transient signal and the SIMION modeling [14,15], whereas the experimental data presented above only partly agrees with them. Specifically, the experimental results confirm the presence of odd order harmonics $(2n+1)\omega_+$, as well as odd order interharmonics $(2n+1)\omega_+ \pm 2k\omega_-$ and $2n\omega_+ \pm (2k+1)\omega_-$. But, in addition, the experimental data demonstrates the presence of even order harmonics $2n\omega_+$, as well as even order interharmonics $(2n+1)\omega_+ \pm (2k+1)\omega_-$ and $2n\omega_+ \pm 2k\omega_-$, all of which are not supported by the analytical and numerical approaches [14,15]. Moreover, the simultaneous presence of even order harmonics and even order interharmonics can be of the same origin. In addition, other experimental reports indicate, sometimes indirectly, that ICR frequency (mass) spectra might not follow the theory exactly. For example, Dunbar mentioned presence of the $2\omega_-$ peak along with the ω_- one, whereas one may expect the $3\omega_-$ peak instead of (or at least of higher intensity than) the $2\omega_-$ peak [30].

It is doubtless that the ICR signal theory is correct as given for the physical model employed in its development [14]. The partial discrepancy of the theory and experimental data rather means that the experimental case shows additional features, Fig. 1 versus Fig. 6 top, which are currently unaccounted for in the theoretical model [14,15]. The reason can be a failure of the electrostatics-based approach, Eqs. (12) and (13), to properly describe the ICR transient signal detection. Alternatively to the considered above theory, in agreement with the general practice of electrodynamics calculations, one may suppose that the charge ΔQ in Eq. (14) should be integrated along the ion's trajectory before its submission to the ICR signal's expression, Eq. (12). The underlying reason is that a detection plate as a part of the detection circuit integrates fast radial, as well as azimuthal, oscillations over the time of flight of the ion along the detection plate. Therefore, Eq. (19) of the developed above amplitude-phase modulation model shall be represented as:

$$AM_m(t) \cong 1 + \varepsilon(2m+1) \cos(\omega_- t - \Delta\phi) \quad (19')$$

We notice, Eq. (19') is a way to qualitatively represent the modulation behavior rather than the exact solution. The result of the numerical simulation of the ICR frequency spectrum for the amplitude-phase modulation according to Eqs. (16), (19') and (20) is shown in Fig. 6 (bottom). Overall, experimental Fig. 1 corresponds to Fig. 6 (bottom) better than to Fig. 6 (top). However, lower correlation between the theory and the experiment for the 2nd harmonic's region, compared to the 1st and 3rd ones, can be due to, presumably, the current form of Eq. (14) that should instead be derived for a time-dependent instantaneous frequency of the ion motion.

4. Conclusions

The developed analytical model of sidebands in FT-ICR mass spectra is not only consistent with the previous analytical and numerical studies on the ICR signal but also contributes to the existing theory by showing the origin of the sidebands. According to the developed model, we conclude that the origin of the sidebands is the broadband amplitude-phase modulation occurring in the ICR transient signal caused by the oscillating radial and azimuthal deviations of the ion trajectory relative to a circular orbit.

From the experimental results obtained, we conclude that the FT-ICR spectra appear more complex than predicted by the theory. Specifically, the experimental data only partly agrees with the theoretical results. Such partial discrepancy between the theory and the experiment warns further studies on the ICR transient signal. We introduced a modified amplitude modulation equation that

considers an integrated ion trajectory along the detection electrodes and received a better correlation between the experimental and theoretical results. The advantages and limitations of the latter approach are to be further evaluated.

From a practical point of view, the spectral composition of the ICR signal and, in particular, the question of sidebands have potential either to remove the sidebands in case they reduce experimental performance, or to apply them for fine tuning of ICR cells and, potentially, in higher-precision mass measurements. Moreover, the spectral composition of the ICR transient signal is the basis for the measurement procedure in FT-ICR MS. The limitation of a long transient duration required to resolve sidebands in FTMS can be potentially overcome by employing non-FT signal processing complementary to the FT, as recently demonstrated for the elucidation of isotopic fine structure information in peptide analysis [31,32].

Acknowledgments

We acknowledge Konstantin Zhurov, who provided valuable comments on the manuscript. We are also grateful for financial support through the Swiss National Science Foundation (SNF) project 200021-125147/1, the Joint Russia-Switzerland Research Program (grant agreement 128357), and the European Research Council (ERC Starting Grant 280271 to YOT).

Appendix A.

A.1. Linearized ion trajectory

The radial coordinate Eq. (8) can be rewritten in terms of the parameter $\varepsilon = \rho_-/\rho_+$ as follows:

$$\rho(t) = \rho_+ \sqrt{1 + \varepsilon^2 + 2\varepsilon \cos(\Delta\omega t - \Delta\phi)} \quad (A1)$$

Taylor expansion of Eq. (A1), $\varepsilon \ll 1$, provides the linearized radial coordinate:

$$\rho(t) = \rho_+ [1 + \varepsilon \cos(\Delta\omega t - \Delta\phi) + o(\varepsilon)] \quad (A2)$$

Here and below, $o(\varepsilon)/\varepsilon \rightarrow 0$ as $\varepsilon \rightarrow 0$ by definition. Further, the azimuthal coordinate Eq. (9) in terms of the parameter $\varepsilon = \rho_-/\rho_+$ is:

$$\varphi(t) = \frac{1}{j} \ln \left\{ \frac{e^{j(\phi_+ - \omega_+ t)} (1 + \varepsilon e^{j(\Delta\omega t - \Delta\phi)})}{\sqrt{1 + 2\varepsilon \cos(\Delta\omega t - \Delta\phi) + \varepsilon^2}} \right\} \quad (A3)$$

Its expansion into Taylor series followed by equivalent transformations provides the linearized azimuthal coordinate:

$$\varphi(t) = \frac{1}{j} \ln \{ e^{j(\phi_+ - \omega_+ t)} (1 + \varepsilon e^{j(\Delta\omega t - \Delta\phi)}) (1 - \varepsilon \cos(\Delta\omega t - \Delta\phi) + o(\varepsilon)) \} \quad (A4)$$

$$\varphi(t) = \frac{1}{j} \ln e^{j(\phi_+ - \omega_+ t)} + \frac{1}{j} \ln [1 + j\varepsilon \sin(\Delta\omega t - \Delta\phi) + o(\varepsilon)] \quad (A5)$$

$$\varphi(t) = \phi_+ - \omega_+ t + \varepsilon \sin(\Delta\omega t - \Delta\phi) + o(\varepsilon) \quad (A6)$$

A.2. Approximated amplitude-phase modulation

In the considered approximation of $\varepsilon \ll 1$, Fourier series expansion of the amplitude modulation term Eq. (16) can be obtained as its Taylor series expansion:

$$AM_m(t) = 1 + \varepsilon(2m+1) \cos(\Delta\omega t - \Delta\phi) + o(\varepsilon) \quad (A7)$$

Further, to obtain the Fourier series expansion of the frequency modulation term Eq. (17), the following identity is used:

$$\exp(jm \sin x) = \sum_{k=-\infty}^{\infty} J_k(m) \exp(jkx) \quad (\text{A8})$$

where $J_k(m)$ are Bessel functions of the first kind. To meet this identity, the frequency modulation Eq. (17) can be rewritten as follows using equivalent transformations:

$$PM_m(t) = \text{Re}\{\exp[j(2m+1)(\phi_+ - \omega_+ t + \varepsilon \sin(\Delta\omega t - \Delta\phi))]\} \quad (\text{A9})$$

$$PM_m(t) = \text{Re}\{\exp[j(2m+1)(\phi_+ - \omega_+ t)] \exp[j\varepsilon(2m+1) \times \sin(\Delta\omega t - \Delta\phi)]\} \quad (\text{A10})$$

where $\text{Re}(z)$ is the real part of a complex-valued function $z(t)$. Now, application of Eq. (A8)–(A10) followed by equivalent transformations yields the frequency modulation term expanded into Fourier series:

$$PM_m(t) = \text{Re} \left\{ \exp[j(2m+1)(\phi_+ - \omega_+ t)] \sum_{k=-\infty}^{\infty} J_k[\varepsilon(2m+1)] \times \exp[jk(\Delta\omega t - \Delta\phi)] \right\} \quad (\text{A11})$$

$$PM_m(t) = \text{Re} \left\{ \sum_{k=-\infty}^{\infty} J_k[\varepsilon(2m+1)] \times \exp[j(2m+1)(\phi_+ - \omega_+ t) + jk(\Delta\omega t - \Delta\phi)] \right\} \quad (\text{A12})$$

$$PM_m(t) = \sum_{k=-\infty}^{\infty} J_k[\varepsilon(2m+1)] \cos[(2m+1)(\phi_+ - \omega_+ t) + k(\Delta\omega t - \Delta\phi)] \quad (\text{A13})$$

Finally, the obtained Eq. (A13) along with the following asymptotic form and the property of $J_k(m)$:

$$J_k[\varepsilon(2m+1)] \approx \left[\frac{\varepsilon(2m+1)}{2} \right]^k \frac{1}{\Gamma(k+1)}, \quad k \geq 0, \quad 0 < \varepsilon(2m+1) \ll \sqrt{k+1} \quad (\text{A14})$$

$$J_{-k}[\varepsilon(2m+1)] = (-1)^{-k} J_k[\varepsilon(2m+1)] \quad (\text{A15})$$

where $\Gamma(k)$ is the gamma function, provide the phase modulation term in the considered approximation:

$$PM_m(t) = \cos[(2m+1)(\phi_+ - \omega_+ t)] + \frac{\varepsilon(2m+1)}{2} \times \sum_{k=\pm 1} \text{sgn}(k) \cos[(2m+1)(\phi_+ - \omega_+ t) + k(\Delta\omega t - \Delta\phi)] \quad (\text{A16})$$

where $\text{sgn}(k)$ is the signum function.

References

[1] A.G. Marshall, C.L. Hendrickson, High-resolution mass spectrometers, in: *Annual Review of Analytical Chemistry*, 2008, pp. 579–599.
 [2] M. Scigelova, M. Hornshaw, A. Giannakopoulos, A. Makarov, Fourier transform mass spectrometry, *Molecular and Cellular Proteomics* 10 (2011).

[3] F. Xian, C.L. Hendrickson, A.G. Marshall, High resolution mass spectrometry, *Analytical Chemistry* 84 (2012) 708–719.
 [4] B. Bogdanov, R.D. Smith, Proteomics by FTICR mass spectrometry: top down and bottom up, *Mass Spectrometry Reviews* 24 (2005) 168–200.
 [5] Y.O. Tsybin, L. Fornelli, A.N. Kozhinov, A. Vorobyev, S.M. Miladinovic, High-resolution and tandem mass spectrometry – the indispensable tools of the XXI century, *CHIMIA* 65 (2011) 641–645.
 [6] K. Hakansson, H.J. Cooper, R.R. Hudgins, C.L. Nilsson, High resolution tandem mass spectrometry for structural biochemistry, *Current Organic Chemistry* 7 (2003) 1503–1525.
 [7] A.G. Marshall, C.L. Hendrickson, G.S. Jackson, Fourier transform ion cyclotron resonance mass spectrometry: a primer, *Mass Spectrometry Reviews* 17 (1998) 1–35.
 [8] L.S. Brown, G. Gabrielse, Geonium theory – physics of a single electron or ion in a Penning trap, *Reviews of Modern Physics* 58 (1986) 233–311.
 [9] M. Kretzschmar, Single-particle motion in a Penning trap – description in the classical canonical formalism, *Physica Scripta* 46 (1992) 544–554.
 [10] M. Wang, A.G. Marshall, Laboratory-frame and rotating-frame ion trajectories in ion-cyclotron resonance mass spectrometry, *International Journal of Mass Spectrometry* 100 (1990) 323–346.
 [11] M.B. Comisarow, Theory of Fourier-transform ion-cyclotron resonance mass spectroscopy. 2. Signal modeling for ion-cyclotron resonance, *Journal of Chemical Physics* 69 (1978) 4097–4104.
 [12] M. Allemann, H.P. Kellerhals, K.P. Wanczek, Sidebands in the ICR spectrum and their application for exact mass determination, *Chemical Physics Letters* 84 (1981) 547–551.
 [13] E.N. Nikolae, M.V. Gorshkov, Dynamics of ion motion in an elongated cylindrical cell of an ICR spectrometer and the shape of the signal registered, *International Journal of Mass Spectrometry* 64 (1985) 115–125.
 [14] P.B. Grosshans, P.J. Shields, A.G. Marshall, Comprehensive theory of the Fourier transform ion-cyclotron resonance signal for all ion trap geometries, *Journal of Chemical Physics* 94 (1991) 5341–5352.
 [15] C.L. Hendrickson, S.C. Beu, G.T. Blakney, A.G. Marshall, SIMION modeling of ion image charge detection in Fourier transform ion cyclotron resonance mass spectrometry, *International Journal of Mass Spectrometry* 283 (2009) 100–104.
 [16] G. Gabrielse, Why is sideband mass spectrometry possible with ions in a Penning trap? *Physical Review Letters* 102 (2009).
 [17] G. Gabrielse, The true cyclotron frequency for particles and ions in a Penning trap, *International Journal of Mass Spectrometry* 279 (2009) 107–112.
 [18] R.D. Chen, S.H. Guan, A.G. Marshall, Generation and detection of coherent magnetron motion in Fourier-transform ion-cyclotron resonance mass spectrometry, *Journal of Chemical Physics* 100 (1994) 2258–2266.
 [19] P.B. Grosshans, A.G. Marshall, Can Fourier transform mass spectral resolution be improved by detection at harmonic multiples of the fundamental ion-cyclotron orbital frequency, *International Journal of Mass Spectrometry and Ion Processes* 107 (1991) 49–81.
 [20] C. Masselon, A.V. Tolmachev, G.A. Anderson, R. Harkewicz, R.D. Smith, Mass measurement errors caused by local frequency perturbations in FTICR mass spectrometry, *Journal of the American Society for Mass Spectrometry* 13 (2002) 99–106.
 [21] A. Vorobyev, M.V. Gorshkov, Y.O. Tsybin, Towards data acquisition throughput increase in Fourier transform mass spectrometry of proteins using double frequency measurements, *International Journal of Mass Spectrometry* 306 (2011) 227–231.
 [22] H. Ben Hamidane, A. Vorobyev, Y.O. Tsybin, Repeatability and reproducibility of product ion abundances in electron capture dissociation mass spectrometry of peptides, *European Journal of Mass Spectrometry* 17 (2011) 321–331.
 [23] G.T. Blakney, C.L. Hendrickson, A.G. Marshall, Predator data station: a fast data acquisition system for advanced FT-ICR MS experiments, *International Journal of Mass Spectrometry* 306 (2011) 246–252.
 [24] A.N. Kozhinov, S.M. Miladinovic, Y.O. Tsybin, Proceeding of the 59th ASMS Conference on Mass Spectrometry and Allied Topics, Denver, CO, June 5–9, 2011.
 [25] D.L. Rempel, E.B. Ledford, S.K. Huang, M.L. Gross, Parametric mode-operation of a hyperbolic Penning trap for Fourier-transform mass spectrometry, *Analytical Chemistry* 59 (1987) 2527–2532.
 [26] L. Schweikhard, M. Lindinger, H.J. Kluge, Parametric mode excitation/dipole mode detection Fourier transform ion cyclotron resonance spectrometry, *Review of Scientific Instruments* 61 (1990) 1055–1058.
 [27] C. Amano, Detection theory of ion cyclotron resonance phenomena, *International Journal of Mass Spectrometry and Ion Processes* 35 (1980) 47–57.
 [28] P. Kofel, Dissertation, Externe Ionisierung in der Ionen-Cyclotron-Resonanz-Spektrometrie, Universität Bremen, Bremen, Germany, 1987 (in German).
 [29] M. Knobler, K.P. Wanczek, Suppression, amplification and application of the 3rd harmonic of the cyclotron frequency in ion-cyclotron resonance spectrometry, *International Journal of Mass Spectrometry* 125 (1993) 127–134.
 [30] R.C. Dunbar, J.H. Chen, J.D. Hays, Magnetron motion of ions in the cubical ICR cell, *International Journal of Mass Spectrometry* 57 (1984) 39–56.
 [31] A.N. Kozhinov, Y.O. Tsybin, Filter diagonalization method-based mass spectrometry for molecular and macromolecular structure analysis, *Analytical Chemistry* 84 (2012) 2850–2856.
 [32] S.M. Miladinović, A.N. Kozhinov, M.V. Gorshkov, Y.O. Tsybin, On the utility of isotopic fine structure mass spectrometry in protein identification, *Analytical Chemistry* 84 (2012) 4042–4051.

A.2 Paper II

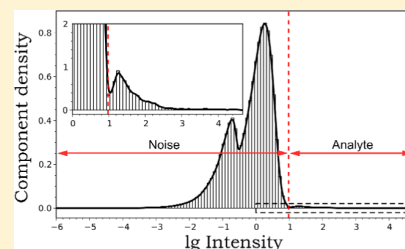
Distinguishing Analyte from Noise Components in Mass Spectra of Complex Samples: Where to Cut the Noise?

Konstantin O. Zhurov, Anton N. Kozhinov, Luca Fornelli, and Yury O. Tsybin*

Biomolecular Mass Spectrometry Laboratory, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

Supporting Information

ABSTRACT: Fourier transform mass spectrometry (FTMS) enables comprehensive analysis of complex molecular mixtures. Given the broad intensity ranges of components in the mass spectra, it is imperative to accurately determine a noise threshold level above which peak assignments will be made. Conventionally, to find the threshold level, the “ N sigma” approach or an equivalent rule is used. However, the “ N sigma” approach cannot be applied to mass spectra stored with partially removed noise (reduced-profile mode). It is also not directly applicable to mass spectra acquired in the absorption mode with removed negative spectral amplitudes. Moreover, N value selection is normally made based on a rule of thumb, meaning that the calculated threshold level may be biased. Here, we present a noise thresholding method which addresses these limitations for analysis of mass spectra of complex molecular mixtures. The introduced data-dependent thresholding method involves analysis of the distribution of logarithmic intensity of all peaks, including noise and analyte, for a given mass spectrum. Selected method applications include FTMS analysis of crude oil fractions as well as tandem MS analysis of intact proteins.



Recent advances in Fourier transform mass spectrometry (FTMS) led to improved analysis of highly complex samples consisting of large numbers of chemical species present in broad ranges of concentrations.^{1–5} Spectral dynamic ranges routinely achieved under reasonable experimental conditions usually span about 3–4 orders of magnitude for the recent generations of high-resolution mass spectrometers.^{5–8} However, current complex mixtures of interest, e.g., petroleomic, proteomic, or metabolomic samples, may have component concentration ranges spanning more than 4–5 orders of magnitude.^{9–22} Therefore, to increase sample component coverage and, ultimately, compound identification, it is necessary to extend the spectral dynamic range typically attained in mass spectra.²³ To that end, numerous associated developments directed at signal processing and hardware optimization were made, including baseline smoothing, differential ion excitation radii, external ion accumulation, etc.^{6,16,24–29} However, for any given experimental setup, it is necessary to know how to distinguish analyte from noise components in a measured mass spectrum. As peak intensity decreases, there comes a point where a significant number of peaks originates from background noise. Hence, it becomes increasingly difficult to differentiate between noise and analyte peaks, particularly if a given peak cannot be related via known mass differences to proximate peaks of higher abundance.³⁰

Identification of chemical compounds using their mass spectrum may be generally formalized as probabilistic determination of their elemental compositions. That is, the expected output for analysis of a mass spectrum is n pairs (elemental composition (c), probability (p)), where the latter measures uncertainty for the reported assignment of the former due to limitations in experiments and data analysis. In situations

when a compound is represented in a mass spectrum by a number of peaks, rather than by a single peak, extra complications arise because a relation between an elemental composition and a set of isotopologues appearing in a mass spectrum should additionally be taken into account. However, it is crucial to realize that for a given type of experiments there exists a certain level of probability, p^* , below which identifications are not considered sufficiently reliable, and hence those identifications, being of no analytical value, are not required. Therefore, for the identification problem, the full set of its solutions $\langle c, p \rangle_i, i = 0, \dots, n^{-1}$ can be *a priori* filtered via applying the condition $p \geq p^*$: only the reduced set of solutions $\langle c, p \rangle_j, j = 0, \dots, k \leq n^{-1}$, which contains the analytically relevant data, is of interest. This way, the large identification problem, which may also be complicated wrt the multiple isotopologues issue, can be easily decomposed into three subsequent separate tasks—thresholding, mass deconvolution, and elemental composition assignment. Importantly, while neglecting the solutions of no analytical value, this decomposition is an equivalent transformation, meaning that it preserves the solutions of the identification problem: the decomposition neither reduces the number of analytically relevant solutions (“false negatives” scenario) nor introduces spurious solutions (“false positives” scenario).

Along with preserving the analytically relevant solutions, the decomposition feature allows for feasible implementation of each of the three steps separately, including a thresholding

Received: October 10, 2013

Accepted: February 28, 2014

Published: February 28, 2014

method. Furthermore, this approach is favorable as it takes advantage of already existing deconvolution and peak assignment methods designed for denoised mass spectra, e.g., Xtract (Thermo Scientific, Bremen, Germany),³¹ SNAP (Bruker Daltonics, Bremen, Germany),³² or MS-Deconv (University of San Diego, San Diego, CA, U.S.A.)³³ for proteomic mass spectra and PetroOrg (Florida State University, Tallahassee, FL, U.S.A.)³⁴ or Composer (Sierra Analytics, Modesto, CA, U.S.A.)⁵ for petroleomic mass spectra. Finally, it is highly desirable to allow for efficient storage of mass spectra as computer files. Particularly, it is relevant in high-throughput research areas such as imaging MS,³⁵ petroleomics,³⁶ and MS-based proteomics,³⁷ where large amounts of data are generated so that keeping file size below some reasonable value is crucial. As a solution to these problems, all peaks in a given mass spectrum residing below the noise level may be removed, resulting in a reduced-profile mass spectrum.

On the other hand, a direct approach to the initial problem of identification with finding a full set of solutions including excessive amounts of those without analytical relevance represents a substantial challenge due to processing of very large amounts of data and potentially dealing with ill-conditioned mathematics, meaning that such an approach may be technically unfeasible for a given experiment. That is also why in most of experimental practices the probabilities of identification are not thoroughly derived for each reported compound, but only estimated for the total set of data. Since the detailed information on probabilities is not available, it is also imperative that any remaining peaks post noise thresholding should belong to the analyte signals, and not noise, in order to make high-confidence assignments and avoid false positives, i.e., peak assignment without noise thresholding is discouraged.

Here, we present a thresholding method that aims to determine an optimal noise threshold level while avoiding the potential limitations of the current strategies. Applications including MS analysis of crude oil fractions as well as tandem MS analysis of intact proteins are considered. Additionally, we analyze the noise distribution theoretically in order to evaluate the obtained results in terms of the conventional “*N* sigma” approach, *vide infra*.

METHODS

Sample Preparation. African and South American crude oil fractions were obtained from IPFEN (Lyon, France) via collaboration. Fuel oil no. 2 and horse myoglobin were obtained from Sigma-Aldrich (Buchs, Switzerland). Humira IgG1 was obtained from Abbott Laboratories (North Chicago, IL, U.S.A.). LC-MS grade dichloromethane, water, and acetonitrile were obtained from Fluka (Buchs, Switzerland). Formic acid was obtained from Merck (Zug, Switzerland). The oil fractions were dissolved in dichloromethane, followed by 1:4 dilution in acetonitrile containing 1% (v/v) of formic acid. Myoglobin and Humira IgG1 were dissolved in 49.9:50:0.1 (v/v) mixture of water, acetonitrile, and formic acid to a final concentration of 5–10 μM .

Mass Spectrometry. MS analysis of petroleomic samples was performed using (i) a hybrid linear ion trap FT-ICR MS (LTQ FT Ultra, Thermo Scientific, Bremen, Germany) equipped with a 10 T superconducting magnet³⁸ and (ii) a hybrid dual linear ion trap Orbitrap FTMS (Orbitrap Elite FTMS, Thermo Scientific) equipped with a high-field compact orbitrap mass analyzer.² Crude oil samples were ionized using a

standard microelectrospray ion source (HESI ion source, Thermo Scientific) and a robotic chip-based nano-ESI ion source (TriVersa Nanomate, Advion Biosciences, Ithaca, NY, U.S.A.). Instrumental parameters and operation were controlled by standard data acquisition systems and software (Thermo Scientific). Total charge target value (AGC value) was set between 10^5 and 5×10^5 ; for both ion sources a 1.6 kV potential was applied to the emitter; flow rates were $\sim 5 \mu\text{L}/\text{min}$ and $\sim 200 \text{ nL}/\text{min}$, respectively. Orbitrap FTMS mass spectra were recorded at a resolution of 480 000 at 400 m/z with the eFT option on; 400–500 individual scans were summed. Magnitude-mode FT-ICR MS mass spectrum was recorded at a resolution of 750 000 at 400 m/z ; 670 individual scans were summed.

Top-down MS was performed on the Orbitrap Elite FTMS following standard procedure.³⁹ Electron-transfer dissociation (ETD) was used for precursor ion fragmentation, with fluoranthene radical anions as electron donors. AGC value was set to 10^6 for MS/MS mode; fluoranthene AGC was set between 10^6 and 2×10^6 . Ion–ion interaction time was 3–5 ms for 23+ charge state precursor of myoglobin and 20 ms for a set of charge states around 50+ of Humira IgG1. The MS/MS spectra were recorded at a resolution of 120 000 at 400 m/z with the eFT option on, each scan being composed of 10 microscans; 20 and 130 scans were averaged for myoglobin and Humira IgG, respectively.

Data Analysis. All mass spectra were stored either in full- or reduced-profile modes. Automatic removal of a significant part of noise peaks took place for the mass spectra acquired in reduced-profile mode (Orbitrap FTMS and FT-ICR MS). Automatic removal of noise peaks with negative amplitudes was observed for the absorption-mode-type (eFT) mass spectra acquired in full-profile mode (Orbitrap FTMS).

For further analysis, binary files with acquired mass spectra were exported into ASCII files. These files were processed using the data analysis framework pyFTMS,⁴⁰ which was developed in house and written in Python programming language. Specifically, for each mass spectrum the quadratic interpolation of local maxima was performed, an appropriate noise threshold was determined using the method described in the current article, and the noise was cut using the determined threshold values. Analyses of the mass spectra were performed using standard procedures, *viz.*, identification or characterization of the analytes under study.^{5,39}

The reported results of the thresholding method were obtained for its computationally efficient implementation, i.e., analysis of bar-plot (peak-picked) mass spectra. Note that, although the comparison of these results to the “*N* sigma” approach includes a theoretical analysis of noise distributions performed in terms of raw spectral amplitudes of the noise (see the Appendix for details), a separate analysis, in which the thresholding method was applied to raw (peak shape) mass spectra, did not show a noticeable bias relative to the presented results.

Throughout this paper, we use the term “intensity” to denote both analyte and/or noise components in a mass spectrum, being equally applicable to both peak-shape (raw data) or bar-plot (peak-picked) mass spectra. Contrastingly, the term “abundance” is used in reference to analyte components only.

RESULTS AND DISCUSSION

Limitations of Common Noise Thresholding Methods. The “*N* sigma” rule, generally employed in FTMS-based

petroleomics, represents the conventional way to cut the noise. To illustrate the “ N sigma” approach, let us consider a transient signal, i.e., the raw data acquired in FTMS, which inevitably contains noise. The major noise component is assumed to be thermal noise from the ion detection electronic circuit and may be modeled as the band-limited Gaussian noise, Figure 1.

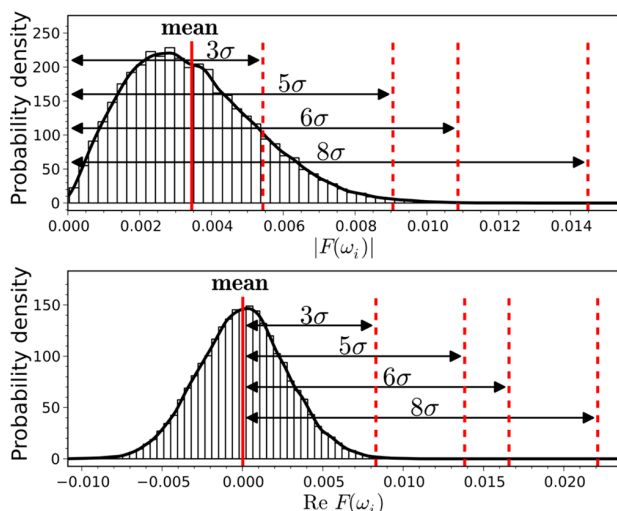


Figure 1. Numerically simulated distribution of spectral amplitudes for noise-only Fourier spectra represented in magnitude (top) and absorption (bottom) modes. Commonly encountered threshold levels expressed in the number of standard deviations, σ , are shown.

Therefore, the noise in the mass spectrum has a Gaussian distribution if the absorption-mode spectral representation is

used, Figure 1 top, or the Rayleigh distribution (the chi-distribution with two degrees of freedom) in case of the magnitude-mode spectral representation, Figure 1 bottom (see the Appendix for further details). The standard deviation, σ , of the noise distribution characterizes how broad the distribution is. The noise threshold level, below which peaks are not considered for assignment, is then determined as $N\sigma$, where N is a positive number. In practice, the standard deviation may be estimated from the “baseline noise” in some analyte-free m/z (frequency) range of the mass (Fourier) spectrum. Additionally, for the thresholding procedure to be consistent, the choice of N should also depend on whether the magnitude or absorption mode is employed to represent the mass spectrum.

In petroleomics, for instance, the 3σ and 6σ rules are commonly applied. However, use of 5σ and 8σ rules has also been reported.^{41–48} For the traditional bottom-up proteomics,⁴⁹ some of the employed methods also consider the estimation of signal-to-noise ratio (S/N),⁵⁰ for instance by approximating the noise level to the average intensity of the lower half of the peaks present in a mass spectrum,⁵¹ or by applying a dynamic noise level algorithm, which considers all the peaks in a given MS/MS spectrum.⁵² Yet no consensus on the noise threshold value has been reached. For the emerging middle-down and top-down proteomics the scenario is similar: MS/MS spectra are normally deconvoluted using standard parameters of the data analysis software, e.g., Xtract³¹ or MS-Deconv.⁵³ Generally, top-down MS/MS spectra are analyzed using the noise threshold value corresponding to $S/N = 3$.^{54,55}

Given the variation in the proposed values on N , the question is how to choose an appropriate threshold level. Although practically useful, the “rule of thumb” basis for choosing a certain value of N is not always justified. For instance, if the actual noise level is overestimated by the $N\sigma$

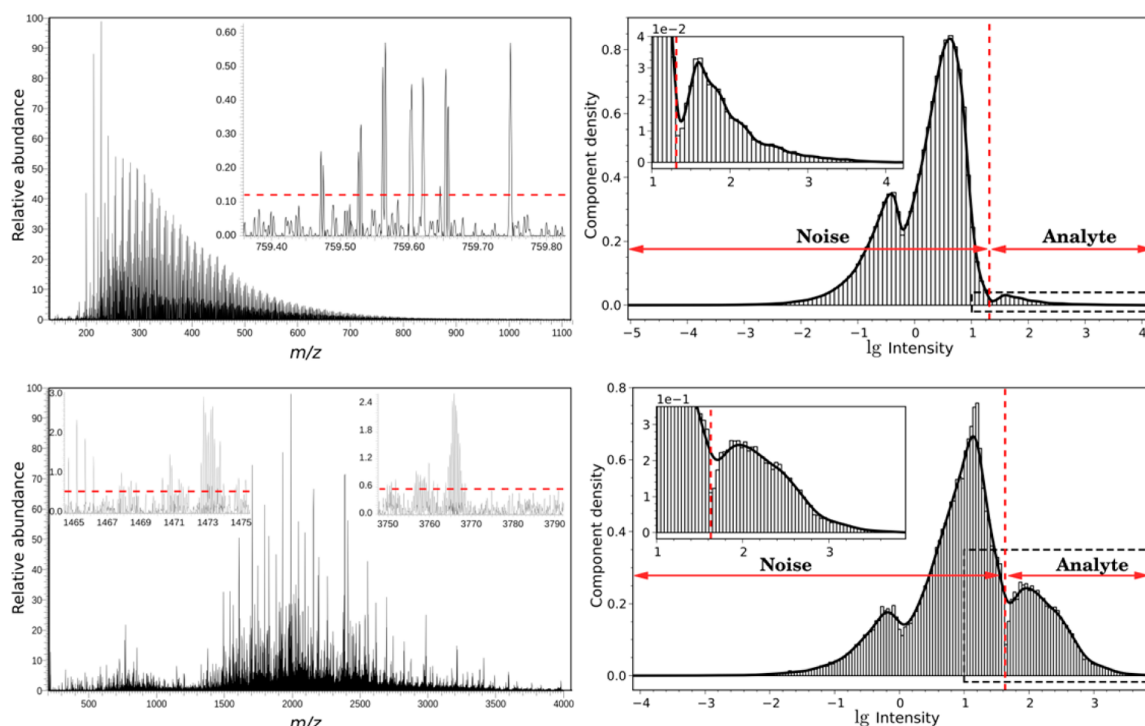


Figure 2. Analysis of full-profile mass spectra of South American crude oil fraction (top) and Humira IgG (bottom). Left panels: absorption-mode-type mass spectra acquired on Orbitrap Elite FTMS. Right panels: distributions of logarithmic intensity. Red lines indicate the noise threshold values.

Table 1. Summary of Threshold Values Derived Using Various Methods for All Samples Presented Herein^{a,b}

figure	sample	MS	storage mode	spectral representation	threshold values by different methods			
					lg A, absolute units/ A, relative units (current method)	A, relative units (visually)	equivalent N for A = Nσ ("N sigma" approach)	equivalent N for A = Nσ (from max of distribution)
2 top	South American crude oil fraction	Orbitrap FTMS	full profile	absorption mode	1.31/0.123%	0.115%	NDA	5.00
2 bottom	Humira IgG	Orbitrap FTMS	full profile	absorption mode	1.63/0.55%	*	NDA	3.13
3 top	African crude oil fraction	Orbitrap FTMS	reduced profile	absorption mode	2.32/0.012%	NA	NA	NA
3 bottom	myoglobin	Orbitrap FTMS	reduced profile	absorption mode	1.85/0.19%	NA	NA	NA
S2	fuel oil no. 2	Orbitrap FTMS	full profile	absorption mode	0.98/0.02%	0.018%	NDA	5.25
S3	African crude oil fraction	FT-ICR MS [#]	reduced profile	magnitude mode	**	0.8%	NA	5.01

^aThe following notation is used: A = intensity; lg A = logarithmic intensity; NA = not applicable; NDA = not directly applicable; # = magnitude mode FT; * = absence of analyte-free regions in mass spectrum; ** = absence of a local minimum in logarithmic intensity distribution plot. ^bAll Orbitrap Elite FTMS mass spectra were acquired with absorption-mode-type spectral representation (eFT signal processing).

value, then low-abundance analyte peaks will be lost at the thresholding stage of the data analysis and will be discarded prior to assignment of elemental compositions ("false negatives" scenario). Importantly, in practice the value of σ may be underivable if some implicit or explicit automatic thresholding procedure is applied during acquisition or storage of mass spectra. Specifically, if the mass spectrum is stored in the reduced-profile mode so that the noise level is automatically removed following data acquisition, then the "N sigma" approach is not applicable since the standard deviation cannot be obtained from the mass spectrum. This is the typical scenario for proteomics data. Additionally, if the absorption-mode spectral representation is used to obtain a mass spectrum with increased resolution and if negative amplitudes are automatically removed, then the "N sigma" approach is not directly applicable since the negative amplitudes are not available for the calculation of σ . Publications indicate that this should be taken into account when dealing with eFT mass spectra acquired on Orbitrap FTMS or absorption-mode mass spectra acquired on FT-ICR MS.^{25,56,57}

Logarithmic Intensity Distribution Noise Thresholding Method. Given that analytes in complex mixtures may span broad ranges of concentrations (so that even the abundance range of detectable ions is nearly saturated for a given mass spectrometer),^{5,58,59} it is convenient to deal with decimal logarithm of peak intensity.⁵ To determine the noise threshold value, we plot the distribution of logarithmic intensity and select the local minimum between the noise and analyte distributions as the threshold level. This point represents the actual relation of the noise and analyte peak distributions as the false positives versus false negatives trade-off. That is, the optimization is made to minimize the amount of noise peaks in the mass spectrum while maximizing the range of abundances of the remaining analytes. We find this optimization condition to be consistent with the above-named objectives of avoiding a loss of analytically relevant data, rejecting most of the noise peaks, and increasing the abundance range of identified compounds.

Orbitrap FTMS Mass Spectra Stored in Full-Profile Mode: Petroleomics and Top-down Proteomics. Figure 2 shows noise level analysis of Orbitrap FTMS mass spectra stored in full-profile mode for representative petroleomic and

top-down proteomic samples. The top left panel shows a broad-band mass spectrum of a South American crude oil fraction; the bottom left panel shows a top-down MS/MS spectrum of an intact monoclonal antibody, Humira IgG1. The corresponding distributions of logarithmic intensity are shown to the right of their respective mass spectra. In both cases, two contributions to the total distribution may be identified. In the lower intensity range, there is a distribution generated by the noise. In the higher intensity range, with significantly smaller number of components, resides a distribution attributable to the analyte peaks.

The logarithmic intensity distribution plots in Figure 2 highlight two significant aspects affecting noise threshold values. First, although the noise and analyte distributions do not have significant overlap for the petroleomics data shown, ca. 12 000 analyte signals are located in the analyte distribution⁵ and an estimated several hundred analyte signals remain hidden in the noise distributions.^{5,9,10} For more complex petroleomic mixtures, the number of analytes can reach 100 000.⁹ Therefore, a significant number of analyte components would intersperse with the noise distribution affecting even more the overall component distribution, resulting in a higher noise threshold value. Second, the noise distributions of these plots exhibit two distinct local maxima. The noise distribution in the higher intensity range is considered to be due to the thermal noise of the detection circuit, whereas the noise distribution in the lower intensity range is presumably due to the quantization effects behind the digital representation of the experimental data. Note that, although the mass spectra have been acquired in full-profile mode and with absorption-mode-type spectral representation, they do not have negative amplitudes. Therefore, if the removal of negative amplitudes is not taken into account, the conventional "N sigma" method would not provide the actual standard deviation, Figure 1 bottom panel, when applied straightforwardly. Additionally, in cases where the number of analytes is non-negligible in comparison to the number of noise components, or where the thermal noise values are comparable with the quantization noise values, caution must be taken when applying any calculations based on deriving the standard deviation for a single noise distribution, as per the standard practice.

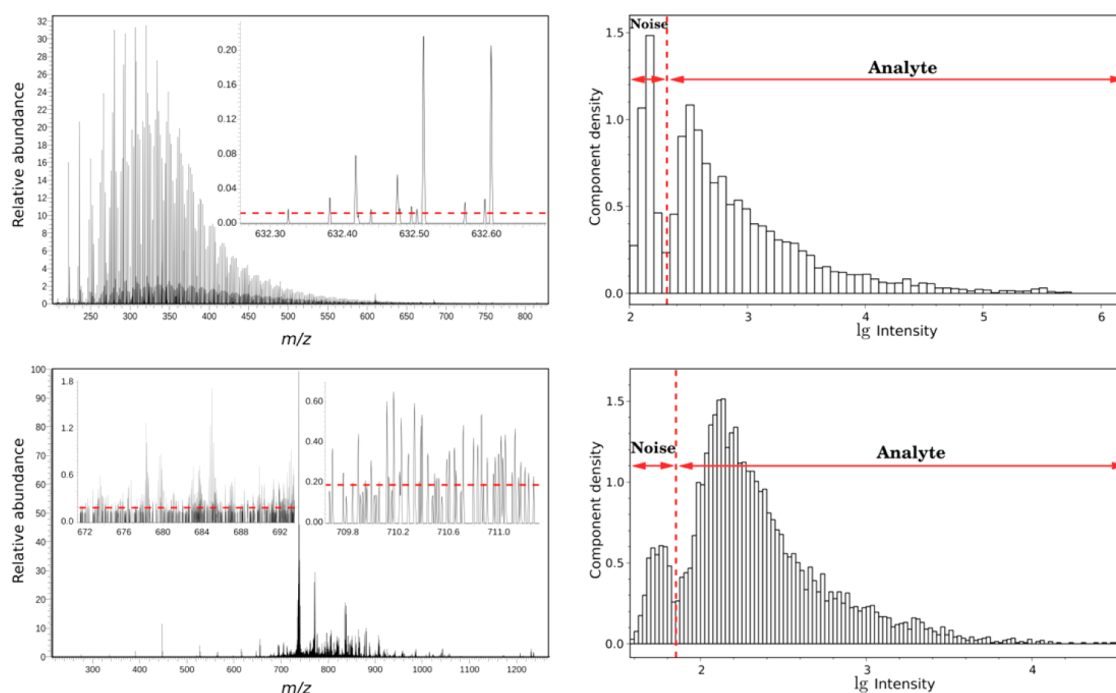


Figure 3. Analysis of reduced-profile mass spectra of African crude oil fraction (top) and myoglobin (bottom). Left panels: absorption-mode-type mass spectra acquired on Orbitrap Elite FTMS. Right panels: distributions of logarithmic intensity. Red lines indicate the noise threshold values.

Hence, we propose to threshold the mass spectral peaks at the local minima between the noise and analyte peak distributions. This way, the dynamic range of analytes is optimized with respect to the noise versus analytes trade-off. Although a number of noise peaks remain, this should not pose a problem in peak assignment in either petroleomic or proteomic samples. Indeed, the highly ordered nature of petroleomic samples ensures that a single peak may be chemically related to a series of other peaks following certain abundance distributions.^{30,60} In case of top-down analysis of large biomolecules, e.g., proteins, vast majority of fragment ions have a number of isotopologues of varying abundance. As such, fragment ion assignment is based on several peaks with specific relative abundance and spacing between them. On the other hand, since the number of noise peaks is significantly higher than the number of analyte peaks, the likelihood of relating several noise peaks via standard mass differences significantly increases when most peaks in a given intensity range originate from noise. Moreover, as stated previously, at the lower end of the intensity scale, relative abundances of isotopic distributions of analytes display larger deviations, with analyte peak shape, abundance, and position affected to a greater extent when noise peak contributions are in comparable intensity range.⁶¹ As such, although a certain percentage of remaining noise peaks is tolerable, thresholding a mass spectrum at too low a value will introduce a non-negligible number of false positives.

In both cases, the logarithmic intensity distribution reveals a minimum between the noise and analyte distributions, Figure 2, right panels. To test the obtained noise threshold value, one may consider a part of the mass spectrum devoid of analyte signals and locate the upper edge of the noise visually—the wider the m/z range considered, the better the estimate. Although, as can be the case in MS/MS mass spectra of top-down analysis of proteins, such regions may be lacking. In cases where such estimation is possible, the noise thresholding values

from visual estimation are in good agreement with local minima intensity values obtained from logarithmic intensity distribution plots, Table 1.

Orbitrap and ICR FTMS Mass Spectra Stored in Reduced-Profile Mode. Figure 3 shows analyses of noise levels of reduced-profile mode mass spectra of petroleomic and top-down proteomic samples. The former is based on a broadband mass spectrum of an African crude oil fraction and the latter on a MS/MS spectrum of a protein, myoglobin. In both cases, the majority of the noise distribution has been cut off during automatic conversion into reduced-profile mass spectra for file storage, Figure 3 left panels. One thus cannot establish a reasonable intensity cutoff value using conventional methods since the standard deviation of the noise intensities cannot be found as the major part of the noise is not available. Likewise, a simple visual approach is not applicable since the noise baseline has been significantly cut off in the analyte-free regions, Figure S1 (Supporting Information). Nevertheless, there are noise peaks remaining in the mass spectrum and the noise and analytes distributions do overlap, Figure 3 right panels. Therefore, a logarithmic intensity distribution plot may still provide the threshold value in the cases presented above, as a distinct local minimum between the noise and analyte distributions is present.

Absence of Minima between Noise and Analyte Distributions. If doubts remain to the degree with which the noise distribution falls off with increasing signal intensity, one solution is to acquire an experimental fingerprint of the noise distribution. If these mass spectra are given with the same vertical scale (that depends on the voltage scales of the corresponding transient signals, any normalizations made to the mass spectra wrt the number of injected ions, etc.) one may expect that the two mass spectra have the same noise distributions. Otherwise, as the noise peak location is irrelevant, the overall shape of the distribution is still the same, provided

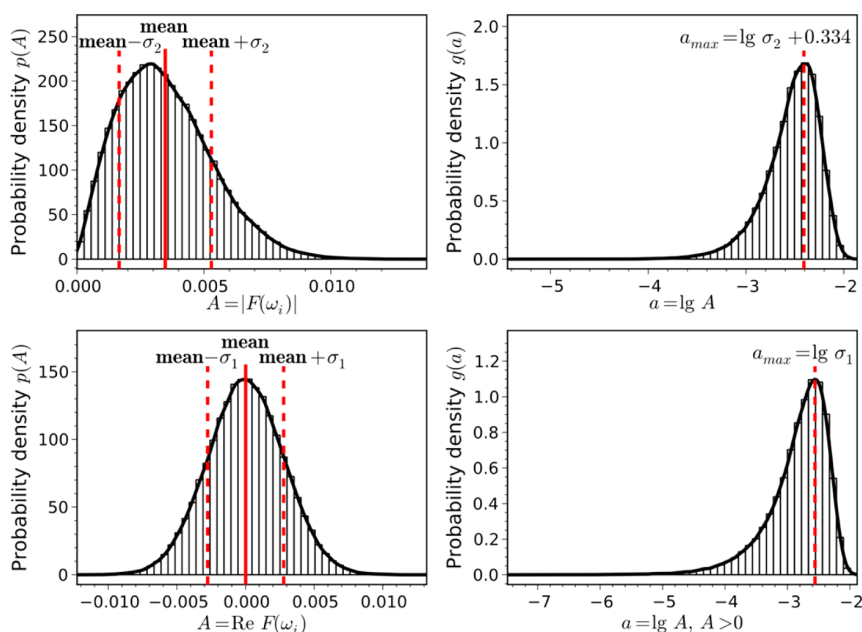


Figure 4. Relation between the distribution of spectral amplitudes and the distribution of logarithmic spectral amplitudes for noise-only Fourier spectra represented in magnitude (top) and absorption (bottom) modes. Left panels: distributions of spectral amplitudes. Right panels: distributions of logarithmic spectral amplitudes; the local maxima provide the standard deviation of interest, σ ; see the Appendix for details.

that the spectral noise is primarily due to thermal noise. Thus, comparing the noise distribution fingerprint with the logarithmic intensity distribution plot of the mass spectrum of interest, one may deduce the threshold level.

To test whether the noise distribution is reproducible, a different petroleomic sample, Figure S2 (Supporting Information), was analyzed on Orbitrap FTMS 4 months after the mass spectrum depicted in Figure 2 top left panel was acquired: the noise distributions are similar in both cases. Now, in the case where the maximum of the analyte distribution lies within the noise distribution resulting in no local minima, Figure S3 (Supporting Information), comparisons can be made to appropriate reference noise distributions, *vide supra*. The difference between the reference noise distribution and the component distribution of the experimental mass spectrum in question may then help to determine the degree of contributions from the analyte and noise distributions, though the analyte versus noise parity is less pronounced in this case.

Comparison of the Noise Thresholding Methods.

Apart from providing a convenient scale for distinguishing analyte and noise distributions for a given mass spectrum, the logarithmic intensity distribution plot has the additional benefit that the local maximum point of the noise distribution provides the standard deviation, σ , Figure 4 (see the Appendix for derivation details). Particularly, the two types of noise distributions appearing in Figures 2 and 3, as well as Supporting Information Figures S2 and S3, now may be assigned to two separate noise distributions with different standard deviations.

Importantly, the logarithmic intensity distribution method provides the standard deviations also in the case when the mass spectrum is acquired in the absorption mode with negative amplitudes cut out, *i.e.*, when the “ N sigma” approach is not directly applicable. For that, we recalculate the obtained threshold values into corresponding values of N via the relation between the standard deviation and the noise distribution maximum, depicted in Figure 4 and detailed in the Appendix.

In doing so, the left-side noise distribution can be neglected since its standard deviation is much smaller than that of the right-side one.

Table 1 summarizes the threshold values obtained in this work and evaluates the thresholding method from the point of view of the “ N sigma” approach and the visual approach. First, for cases where a visual noise baseline could be established, the thresholding values, given as percentages from the base peak, are close to those obtained visually. However, for Humira IgG, the analyte peak density was such that an analyte-free m/z region could not be reliably found, and hence, no approximated thresholding value could be established. Likewise, the visual approach could not be applied to reduced profile mass spectra, as the noise baseline is effectively cut off, Figure S1. Yet the logarithmic intensity distribution method may be used to obtain a thresholding value in all these cases.

Second, the “ N sigma” approach cannot be applied to reduced-profile mass spectra at all, due to partial removal of the noise distribution, and it also cannot be directly applied to absorption-mode-type mass spectra if the noise distribution is altered by removal of peaks with negative amplitudes. As such, the “ N sigma” approach cannot be directly applied to any of the reduced-profile mass spectra considered herein. Nevertheless, the maximum of the logarithmic intensity distribution readily provides the standard deviation in question, Figure 4.

The equivalent “ N sigma” values obtained from the logarithmic intensity distribution plot for the three full-profile and one reduced-profile mass spectra range between 3.13σ and 5.25σ —values close to the ones typically employed using standard noise thresholding procedures. Importantly, for the cases when “ N sigma” can be directly applied, using a single N value for all data sets may not produce optimum results, as interspersions between noise and analyte distributions will result in varying degree to which the noise distribution and its associated parameters deviate from an “ideal” noise distribution. Hence, applying the logarithmic intensity distribution method

in lieu of the “*N* sigma” approach will produce a more optimized thresholding value, by either including more analyte signals or by reducing the number of noise peaks present, when the “*N* sigma” approach overestimates or underestimates, respectively.

Finally, although the optimization condition of the thresholding method is balanced with respect to the noise versus analytes distributions, a more stringent optimization condition may be imposed when required. For instance, the optimization may be alternatively made for the mass accuracy of the remaining peaks, discarding a certain percentage of low-abundance analytes. Indeed, the thresholding procedure governs the appearance of low-intensity peaks in the resultant mass spectra, and the maximum random error of *m/z* values decreases as the peak intensity increases. Then, the total errors of the *m/z* values for low-intensity peaks are mainly defined by random errors, provided that the systematic errors are reduced below the random errors’ level via an appropriate mass calibration.⁶² Therefore, as the threshold level increases, the total errors of the *m/z* values for low-intensity peaks decrease.

CONCLUSIONS

We outline a data-dependent noise thresholding method for improved MS-based analysis of complex mixtures. The method is best applicable for cases where the number of analyte signals in a mass spectrum is non-negligible with respect to the noise distribution. The method is validated on FTMS mass spectra of proteomic and petroleomic samples, showing robustness for optimization of noise thresholding values irrespective of the instrument-specific noise distribution. The results demonstrate that applying a single “*N* sigma” value to all data sets, as is the current convention, does not always result in consistent noise thresholding. Finally, while other commonly employed thresholding methods based on estimating S/N values cannot establish a noise threshold level for a reduced-profile mass spectrum, the presented method is shown to be applicable to reduced-profile mass spectra of complex samples. This alleviates the inherent disadvantage of using efficient storage of mass spectra via prestorage noise removal.

APPENDIX

Here, we establish a relationship between the standard deviation of the distribution of spectral amplitudes and the local maximum of the distribution of logarithmic spectral amplitudes for a transient signal containing random noise with a Gaussian distribution. The problem is solved separately for the absorption-mode and magnitude-mode FT spectral representations. For the sake of brevity, apodization and zero-fillings of the transient signal are not considered and the derivations are given in terms of continuous-variable functions.

First, let us find the distributions of the noise spectral amplitudes. Let the experimental transient signal be represented as

$$f(t) = s(t) + n(t) \quad (\text{A1})$$

where *s*(*t*) is the signal induced by the ions moving in the mass analyzer and *n*(*t*) is a single realization of the noise. From here on, we may ignore the signal, viz., it is assumed that *s*(*t*) ≡ 0. Due to the linearity property of the Fourier transform, this assumption does not change the final result in the case of absorption-mode spectral representation; for the magnitude

mode the result will be an approximation with reasonable accuracy. The Fourier transform of *f*(*t*) can be then given as

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t)e^{-j\omega t} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} n(t)e^{-j\omega t} dt \quad (\text{A2})$$

We introduce the random variables *A* = Re*F*(ω) and *A* = |*F*(ω)|, Figure 1, for the cases of the absorption-mode and magnitude-mode spectral representations, respectively. Let *p*(*A*) denote the probability density function for the corresponding variable *A*. In the case of absorption mode, *p*(*A*) is the Gaussian distribution with zero mean value and some standard deviation σ_1 :

$$p(A) = k_1 \exp\left(-\frac{A^2}{2\sigma_1^2}\right) \quad (\text{A3})$$

In turn, the variable *A* = |*F*(ω)| is the square root of the sum of squares of two variables having the same distribution, eq A3:

$$A = |F(\omega)| = \sqrt{\text{Re}^2 F(\omega) + \text{Im}^2 F(\omega)} \quad (\text{A4})$$

Therefore, for magnitude mode, the function *p*(*A*) is the χ^2 -distribution with two degrees of freedom:

$$p(A) = k_2 \left(\frac{A}{\sigma_1}\right) \exp\left(-\frac{A^2}{2\sigma_1^2}\right) \quad (\text{A5})$$

with the following standard deviation expressed in terms of the standard deviation of the distribution eq A3:

$$\sigma_2 = \sigma_1 \sqrt{2 - 2\Gamma^2(3/2)} \cong 0.655\sigma_1 \quad (\text{A6})$$

where $\Gamma(x)$ is the gamma function.

Now, consider the following change of variables:

$$a = \lg A, \quad A > 0 \quad (\text{A7})$$

$$a = \lg A \quad (\text{A8})$$

for absorption and magnitude modes, respectively. Let us find the corresponding probability density functions *g*(*a*) and their local maxima. From the change of variables theorem⁶³ and eqs A3, A5, and A6 it follows that both functions *g*(*a*) can be written as

$$g(a) = p(A) \frac{dA}{da} = p(A) \frac{A}{\lg e} \quad (\text{A9})$$

where *p*(*A*) is given by eq A3 or eq A5, depending on the spectral representation. The following condition defines the points *a*_{max} of local maxima of the functions *g*(*a*):

$$\left. \frac{dg}{da} \right|_{a^*=a_{\max}} = \left. \frac{dg}{d \lg A} \right|_{\lg A^*=a_{\max}} = 0 \quad (\text{A10})$$

which can be reduced via equivalent transformations:

$$\begin{aligned} \left. \frac{dg}{d \lg A} \right|_{\lg A^*=a_{\max}} &= \left. \frac{1}{\lg e} \left(\frac{d(pA)}{dA} \frac{dA}{d \lg A} \right) \right|_{\lg A^*=a_{\max}} \\ &= \left. \frac{1}{\lg^2 e} \left[A \frac{dp}{dA} + p \right] \right|_{\lg A^*=a_{\max}} \end{aligned} \quad (\text{A11})$$

to the following form:

$$A^* \frac{dp(A^*)}{dA} + p(A^*) = 0 \quad (\text{A12})$$

Using eqs A3 and A5, we obtain the derivative dp/dA for the absorption and magnitude modes, respectively:

$$\frac{dp_1}{dA} = -\frac{A}{\sigma^2} k_1 \exp\left(-\frac{A^2}{2\sigma^2}\right) = -\frac{A}{\sigma^2} p_1 \quad (\text{A13})$$

$$\frac{dp_2}{dA} = -\left(\frac{A}{\sigma^2} + \frac{1}{A}\right) p_2 \quad (\text{A14})$$

From the change of variables eqs A7 and A8, and obtained derivatives eqs A13 and A14, we find the following solutions for eq A12 in case of absorption and magnitude modes, respectively:

$$\sigma_1 = A^* = 10^{a_{\max}} \quad (\text{A15})$$

$$\sigma_1 = \frac{1}{\sqrt{2}} A^* = \frac{1}{\sqrt{2}} 10^{a_{\max}} \quad (\text{A16})$$

Finally, rewriting eq A16 in terms of the standard deviation eq A6 of the magnitude-mode noise distribution gives

$$\sigma_2 \cong 0.463 A^* = 0.463 \times 10^{a_{\max}} \quad (\text{A17})$$

Thus, we have found the expressions of interest: eq A15 and eq A17. For convenience, they can be alternatively represented as follows (see Figure 4):

$$a_{\max} = \lg \sigma_1 \quad (\text{A18})$$

$$a_{\max} \cong \lg \sigma_2 + 0.334 \quad (\text{A19})$$

■ ASSOCIATED CONTENT

📄 Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: yury.tsybin@epfl.ch.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was financially supported by the Swiss National Science Foundation (projects 200021-147006 and 128357) and the European Research Council (ERC Starting Grant 280271).

■ REFERENCES

- (1) Schaub, T. M.; Hendrickson, C. L.; Horning, S.; Quinn, J. P.; Senko, M. W.; Marshall, A. G. *Anal. Chem.* **2008**, *80*, 3985–3990.
- (2) Michalski, A.; Damoc, E.; Lange, O.; Denisov, E.; Nolting, D.; Mueller, M.; Viner, R.; Schwartz, J.; Remes, P.; Belford, M.; Dunyach, J.-J.; Cox, J.; Horning, S.; Mann, M.; Makarov, A. *Mol. Cell. Proteomics* **2012**, *11*, 013698/1–013698/11.
- (3) Savory, J. J.; Kaiser, N. K.; McKenna, A. M.; Xian, F.; Blakney, G. T.; Rodgers, R. P.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2011**, *83*, 1732–1736.
- (4) Qi, Y.; Barrow, M. P.; Li, H.; Meier, J. E.; Van Orden, S. L.; Thompson, C. J.; O'Connor, P. B. *Anal. Chem.* **2012**, *84*, 2923–2929.
- (5) Zhurov, K. O.; Kozhinov, A. N.; Tsybin, Y. O. *Energy Fuels* **2013**, *27*, 2974–2983.

- (6) Belov, M. E.; Nikolaev, E. N.; Anderson, G. A.; Auberry, K. J.; Harkewicz, R.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 38–48.
- (7) Marshall, A. G.; Rodgers, R. P. *Acc. Chem. Res.* **2004**, *37*, 53–59.
- (8) Ding, X.; Ghobarah, H.; Zhang, X.; Jaochico, A.; Liu, X.; Deshmukh, G.; Liederer, B. M.; Hop, C. E. C. A.; Dean, B. *Rapid Commun. Mass Spectrom.* **2013**, *27*, 401–408.
- (9) Marshall, A. G.; Chen, T.; Xian, F.; Kaiser, N. K.; Quinn, J. P.; Beu, S. C.; Hendrickson, C. L. Proceedings of the 60th ASMS Conference on Mass Spectrometry and Allied Topics, Vancouver, Canada, May 20–24, 2012.
- (10) Enke, C. G.; Nagels, L. J. *Anal. Chem.* **2011**, *83*, 2539–2546.
- (11) Picotti, P.; Bodenmiller, B.; Mueller, L. N.; Domon, B.; Aebersold, R. *Cell* **2009**, *138*, 795–806.
- (12) Castro-Marciano, F.; Lobodin, V. V.; Rodgers, R. P.; McKenna, A. M.; Marshall, A. G.; Mathews, J. P. *Fuel* **2012**, *95*, 35–49.
- (13) Marshall, A.; Blakney, G.; Beu, S.; Hendrickson, C.; McKenna, A.; Purcell, J.; Rodgers, R.; Xian, F. *Eur. J. Mass Spectrom.* **2010**, *16*, 367–371.
- (14) Nagels, L. J.; Creten, W. L.; Vanpeperstraete, P. M. *Anal. Chem.* **1983**, *55*, 216–220.
- (15) Ricard, E.; Pecheyran, C.; Sanabria Ortega, G.; Prinzhofner, A.; Donard, O. X. *Anal. Bioanal. Chem.* **2011**, *399*, 2153–2165.
- (16) Swearingen, K. E.; Hoopmann, M. R.; Johnson, R. S.; Saleem, R. A.; Aitchison, J. D.; Moritz, R. L. *Mol. Cell. Proteomics* **2012**, *11*, M111.014985.
- (17) Liu, H.; Sadygov, R. G.; Yates, J. R. *Anal. Chem.* **2004**, *76*, 4193–4201.
- (18) Jmeian, Y.; El Rassi, Z. *Electrophoresis* **2009**, *30*, 249–261.
- (19) Corthals, G. L.; Wasinger, V. C.; Hochstrasser, D. F.; Sanchez, J.-C. *Electrophoresis* **2000**, *21*, 1104–1115.
- (20) Eriksson, J.; Fenyö, D. *J. Proteome Res.* **2005**, *4*, 387–393.
- (21) Southam, A. D.; Payne, T. G.; Cooper, H. J.; Arvanitis, T. N.; Viant, M. R. *Anal. Chem.* **2007**, *79*, 4595–4602.
- (22) Becker, S.; Kortz, L.; Helmschrodt, C.; Thiery, J.; Ceglarek, U. *J. Chromatogr., B* **2012**, *883–884*, 68–75.
- (23) Wong, C. L.; Cociorva, D.; Venable, J. D.; Xu, T.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1405–1414.
- (24) Riedo, A.; Bieler, A.; Neuland, M.; Tulej, M.; Wurz, P. *J. Mass Spectrom.* **2013**, *48*, 1–15.
- (25) Xian, F.; Corilo, Y. E.; Hendrickson, C. L.; Marshall, A. G. *Int. J. Mass Spectrom.* **2012**, *325–327*, 67–72.
- (26) Jungmann, J. H.; MacAleese, L.; Visser, J.; Vrakking, M. J. J.; Heeren, R. M. A. *Anal. Chem.* **2011**, *83*, 7888–7894.
- (27) Abplanalp, D.; Wurz, P.; Huber, L.; Leya, I. *Int. J. Mass Spectrom.* **2010**, *294*, 33–39.
- (28) Belov, M. E.; Anderson, G. A.; Angell, N. H.; Shen, Y.; Tolic, N.; Udseth, H. R.; Smith, R. D. *Anal. Chem.* **2001**, *73*, 5052–5060.
- (29) Kaiser, N. K.; McKenna, A. M.; Savory, J. J.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2013**, *85*, 265–272.
- (30) Kilgour, D. P. A.; Mackay, C. L.; Langridge-Smith, P. R. R.; O'Connor, P. B. *Anal. Chem.* **2012**, *84*, 7431–7435.
- (31) Senko, M.; Beu, S.; McLafferty, F. J. *Am. Soc. Mass Spectrom.* **1995**, *6*, 52–56.
- (32) Fornelli, L.; Parra, J.; Hartmer, R.; Stoermer, C.; Lubeck, M.; Tsybin, Y. *Anal. Bioanal. Chem.* **2013**, *405*, 8505–8514.
- (33) Liu, X.; Inbar, Y.; Dorrestein, P. C.; Wynne, C.; Edwards, N.; Souda, P.; Whitelegge, J. P.; Bafna, V.; Pevzner, P. A. *Mol. Cell. Proteomics* **2010**, *9*, 2772–2782.
- (34) Klitzke, C. F.; Corilo, Y. E.; Siek, K.; Binkley, J.; Patrick, J.; Eberlin, M. N. *Energy Fuels* **2012**, *26*, 5787–5794.
- (35) Schramm, T.; Hester, A.; Klinkert, I.; Both, J.-P.; Heeren, R. M. A.; Brunelle, A.; Laprévotte, O.; Desbenoit, N.; Robbe, M.-F.; Stoeckli, M.; Spengler, B.; Römpf, A. *J. Proteomics* **2012**, *75*, 5106–5110.
- (36) McKenna, A. M.; Donald, L. J.; Fitzsimmons, J. E.; Juyal, P.; Spicer, V.; Standing, K. G.; Marshall, A. G.; Rodgers, R. P. *Energy Fuels* **2013**, *27*, 1246–1256.
- (37) Wilhelm, M.; Kirchner, M.; Steen, J. A. J.; Steen, H. *Mol. Cell. Proteomics* **2012**, *11*, O111.011379.

- (38) Ben Hamidane, H.; Vorobyev, A.; Tsybin, Y. O. *Eur. J. Mass Spectrom.* **2011**, *17*, 321–331.
- (39) Fornelli, L.; Damoc, E.; Thomas, P. M.; Kelleher, N. L.; Aizikov, K.; Denisov, E.; Makarov, A.; Tsybin, Y. O. *Mol. Cell. Proteomics* **2012**, *11*, 1758–1767.
- (40) The open-source release is available at <https://bitbucket.org/kozhinov/pyfims>.
- (41) Smith, D. F.; Schaub, T. M.; Rodgers, R. P.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2008**, *80*, 7379–7382.
- (42) Purcell, J. M.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G. *Anal. Chem.* **2006**, *78*, 5906–5912.
- (43) Shi, Q.; Zhao, S.; Xu, Z.; Chung, K. H.; Zhang, Y.; Xu, C. *Energy Fuels* **2010**, *24*, 4005–4011.
- (44) Liao, Y.; Shi, Q.; Hsu, C. S.; Pan, Y.; Zhang, Y. *Org. Geochem.* **2012**, *47*, 51–65.
- (45) Hsu, C.-S.; Hendrickson, C. L.; Rodgers, R. P.; McKenna, A. M.; Marshall, A. G. *J. Mass Spectrom.* **2011**, *46*, 337–343.
- (46) Chiaberge, S.; Fiorani, T.; Savoini, A.; Bionda, A.; Ramello, S.; Pastori, M.; Cesti, P. *Fuel Process. Technol.* **2013**, *106*, 181–185.
- (47) Marshall, A. G.; Hendrickson, C. L. *Annu. Rev. Anal. Chem.* **2008**, *1*, 579–599.
- (48) Podgorski, D. C.; Corilo, Y. E.; Nyadong, L.; Lobodin, V. V.; Robbins, W. K.; McKenna, A. M.; Marshall, A. G.; Rodgers, R. P. *Energy Fuels* **2013**, *27*, 1268–1276.
- (49) McAlister, G. C.; Berggren, W. T.; Griep-Raming, J.; Horning, S.; Makarov, A.; Phanstiel, D.; Stafford, G.; Swaney, D. L.; Syka, J. E. P.; Zabrouskov, V.; Coon, J. J. *J. Proteome Res.* **2008**, *7*, 3127–3136.
- (50) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 320–332.
- (51) Purvine, S.; Kolker, N.; Kolker, E. *OMICS* **2004**, *8*, 255–265.
- (52) Xu, H.; Freitas, M. A. *BMC Bioinf.* **2010**, *11*, 436.
- (53) Wu, C.; Qian, K.; Nefliu, M.; Cooks, R. G. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 261–267.
- (54) Guan, S.; Burlingame, A. L. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 455–459.
- (55) Sun, L.; Knierman, M. D.; Zhu, G.; Dovichi, N. J. *Anal. Chem.* **2013**, *85*, 5989–5995.
- (56) Lange, O. *Methods and Apparatus for Producing a Mass Spectrum*. U.S. Patent 20110240841, 2011.
- (57) Lange, O.; Damoc, E.; Wiegand, A.; Makarov, A. Proceedings of the 59th ASMS Conference on Mass Spectrometry and Allied Topics, Denver, CO, USA, June 5–9, 2011.
- (58) Kekäläinen, T.; Pakarinen, J. M. H.; Wickström, K.; Lobodin, V.; McKenna, A. M.; Jänis, J. *Energy Fuels* **2013**, *27*, 2002–2009.
- (59) Zubarev, R. A. *Proteomics* **2013**, *13*, 723–726.
- (60) Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G.; Qian, K. *Anal. Chem.* **2001**, *73*, 4676–4681.
- (61) Kozhinov, A. N.; Miladinovic, S. M.; Tsybin, Y. O. Proceedings of the 59th ASMS Conference on Mass Spectrometry and Allied Topics, Denver, CO, USA, June 5–9, 2011.
- (62) Kozhinov, A. N.; Zhurov, K. O.; Tsybin, Y. O. *Anal. Chem.* **2013**, *85*, 6437–6445.
- (63) Bendat, J. S.; Piersol, A. G. *Random Data: Analysis and Measurement Procedures*, 4th ed.; Wiley: Hoboken, NJ, 2010; p 640.

Supplementary Information.

Distinguishing analyte from noise components in mass spectra of complex samples: where to cut the noise?

Konstantin O. Zhurov, Anton N. Kozhinov, Luca Fornelli, and Yury O. Tsybin

Figure S1. Expanded view of a reduced-profile mass spectrum of an African crude oil fraction acquired on Orbitrap Elite FTMS with absorption mode-type FT signal processing. Note the absence of noise peaks in the analyte-free regions, which usually may be used to establish a visual noise baseline estimate.

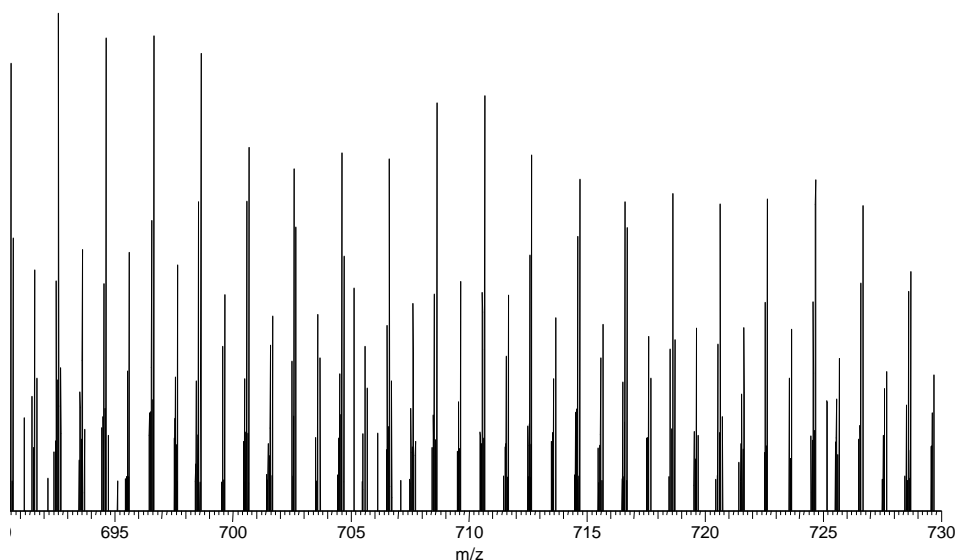


Figure S2. Analysis of a full-profile mass spectrum of fuel oil No.2. Left: full profile mass spectrum acquired on an Orbitrap Elite FTMS with absorption-mode-type FT signal processing. Right: logarithmic intensity distribution plot. Red line indicates the noise threshold value.

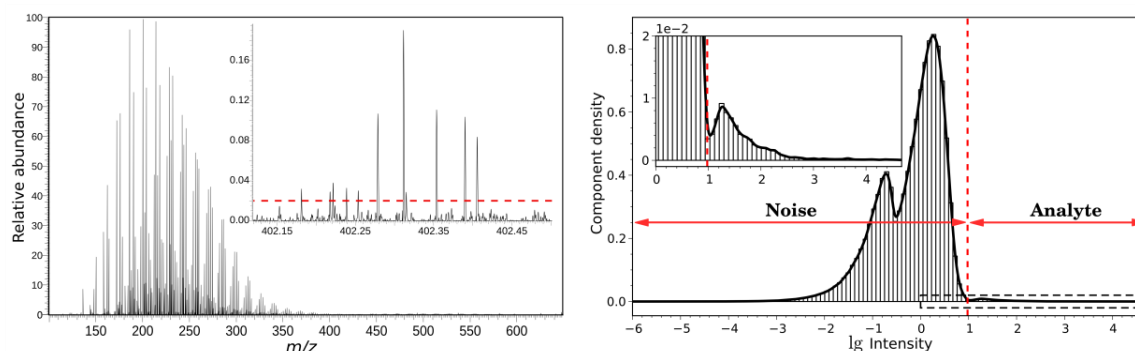
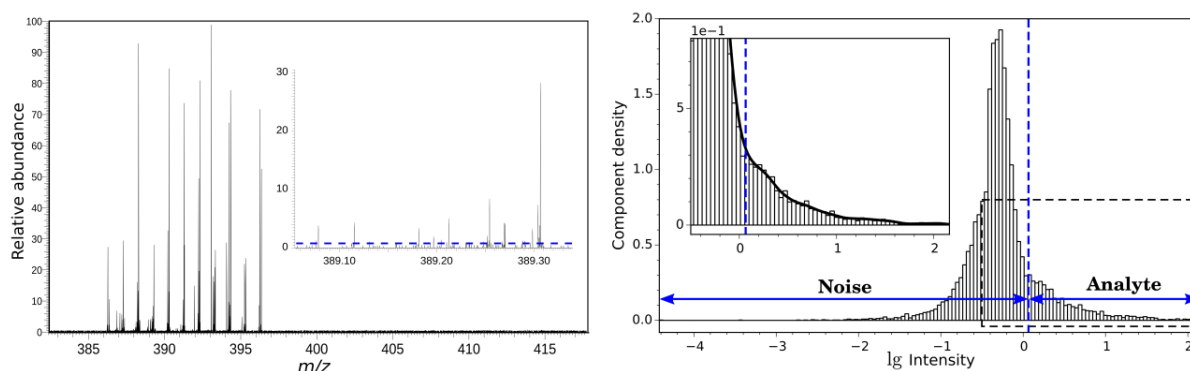


Figure S3. Analysis of a reduced-profile mass spectrum of African crude oil fraction. Left: reduced-profile mass spectrum acquired on a 10 T FT-ICR MS with magnitude-mode FT. Right: logarithmic intensity distribution plot. Blue line indicates the noise threshold value estimated visually. Presence of a maximum in the noise distribution allows the derivation of a noise thresholding value for “*N* sigma” method, Table 1.



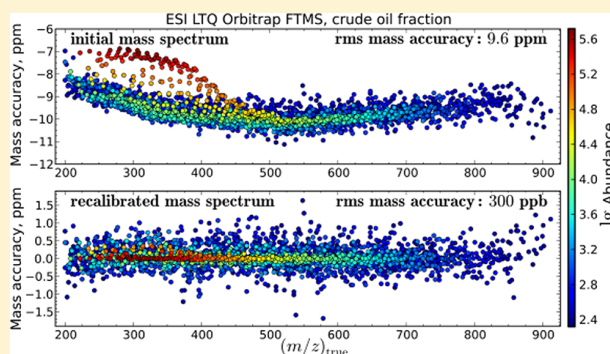
A.3 Paper III

Iterative Method for Mass Spectra Recalibration via Empirical Estimation of the Mass Calibration Function for Fourier Transform Mass Spectrometry-Based Petroleomics

Anton N. Kozhinov, Konstantin O. Zhurov, and Yury O. Tsybin*

Biomolecular Mass Spectrometry Laboratory, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

ABSTRACT: We describe a mass spectra recalibration method, which enables analysis of petroleum samples with Orbitrap FTMS. In this method, the mass calibration function is estimated on the basis of mass-to-charge ratios and abundances of internal calibrants without a need for theoretical description of residual mass errors. Importantly, to maximize the estimation accuracy of the mass calibration function, an iterative approach is implemented to obtain sufficiently high number of internal calibrants covering the entire ranges of mass-to-charge ratios and abundances of interest. For petroleomic samples, the method routinely provides root-mean-square (RMS) mass accuracies at sub-ppm level and hence allows for reliable assignment of elemental compositions. Moreover, since the achieved mass accuracies are normally limited only by random errors of low-abundance analytes, the method maximizes the range of abundances of assignable species for a given signal-to-noise ratio of experimental data. Additionally, despite being initially developed for Orbitrap FTMS, the method is likewise applicable for ion cyclotron resonance FTMS.



High-resolution mass analysis of complex petroleomic samples is mainly performed using high-field Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS).^{1,2} Additionally, high-field Orbitrap FTMS has recently been shown to be suitable for analysis of petroleomic samples of relatively high complexity, e.g., up to and including maltene fractions.³ However, in FTMS-based petroleomics, even if the analytes of interest are well-resolved in a mass spectrum, it does not necessarily mean that the mass measurement accuracy achieved is sufficient for correct identification of their elemental compositions. Specifically, the mass accuracy at the level of hundreds of ppb is the current requirement for high-resolution mass analysis in petroleomics.^{4,5} Nevertheless, such mass accuracy may still be achieved if systematic errors in mass measurements are reduced significantly below the random errors level, provided that the latter is within the required mass accuracy level.

To identify the source of errors, recall that the FTMS mass measurement principle may be generally written via the mapping F from the mass-to-charge ratios m/z and other relevant physical quantities p_i , $1 \leq i \leq m$ (such as ion abundances, ion trajectories, etc.), to the corresponding ion motion frequencies f for a given mass analyzer:

$$f = F(m/z, p_1, \dots, p_m) \quad (1)$$

Then, to obtain the mass-to-charge ratios, eq 1 may be solved for m/z :

$$m/z = M_0(f, p_1, \dots, p_m) \quad (2)$$

where M_0 is the mass calibration function. Below are the theoretical forms of the mass calibration function for the model case of a single ion moving in: (i) a uniform magnetic field and an electric field with quadratic potential of a Penning trap,^{6,7} eq 3; (ii) an electric field with quadro-logarithmic potential of an Orbitrap,⁸ eq 3':

$$M_{\text{theor}}(f) = \frac{C_1}{f} + \frac{C_2}{f^2} \quad (3)$$

$$M_{\text{theor}}(f) = \frac{C_2}{f^2} \quad (3')$$

where C_1 and C_2 are constants defined by magnetic and electric fields, respectively, for a given mass analyzer. However, due to various unaccounted effects, the fields may be perturbed such that the theoretical mass calibration functions, eqs 3 and 3', do not provide sufficient mass accuracy even for routine applications. Specifically, one of the most significant perturbations of the electric field results from ion-ion interactions. For a first-order approximation, these were taken into account initially in FT-ICR MS and later in Orbitrap FTMS by introducing the total charge of ions in the mass analyzer, Q , into the electric-field parameter C_2 : $C_2 = C_2(Q)$.^{9,10}

Received: April 3, 2013

Accepted: June 3, 2013

Published: June 3, 2013

It should be noted, however, that conventional approaches for mass calibration in FTMS introduce additional features. Specifically, in FT-ICR MS, the magnetic-field constant C_1 is implied to be loose, contrary to eq 3, and hence should be considered as: $C_1 = C_1(Q)$. Additionally, in Orbitrap FTMS, an extra term with a total charge-dependent parameter $C_3 = C_3(Q)$ is added to eq 3'.¹¹ Thus, the mass calibration functions conventionally utilized in FT-ICR MS and Orbitrap FTMS, respectively, may be written via the following, or similar, semi-theoretical forms:

$$M_{S/\text{theor}}(f) = \frac{C_1(Q)}{f} + \frac{C_2(Q)}{f^2} \quad (4)$$

$$M_{S/\text{theor}}(f) = \frac{C_2(Q)}{f^2} + \frac{C_3(Q)}{f^3} \quad (4')$$

We note that, in the case of recent non-LTQ-based Orbitrap FTMS platforms, e.g., Exactive series Orbitrap FTMS, a spline-based calibration is implemented, meaning that the parameters depend also on mass-to-charge ratios.

The default way to obtain an estimate of the mass calibration function M_0 is to evaluate the semi-theoretical mass calibration functions, eqs 4 and 4', on the basis of calibrants analyzed in separate measurements ("external" calibration). In this case, the systematic errors in the observed mass-to-charge ratios, $(m/z)_{\text{obs}}$, are governed by the accuracy of the obtained estimate $M_{\text{def}}(f)$ relative to the actual mass calibration function M_0 , eq 2:

$$(m/z)_{\text{obs}} = M_{\text{def}}(f) \quad (5)$$

With implementation of the automatic gain control (AGC) system to control the ion population in the mass analyzer,^{12,13} this calibration approach, eq 5, allows one to achieve mass accuracies on the order of 1 ppm in routine applications.^{14,15} Nevertheless, such performance is insufficient for analysis of complex petroleum samples.

Thus, to reduce the systematic errors in mass measurements, an estimate of the mass calibration function with improved accuracy is required. For that, the semi-theoretical mass calibration functions, eqs 4 and 4', may be evaluated on the basis of calibrants present within a given mass spectrum ("internal" calibration). However, to obtain more accurate estimates of the mass calibration function, further theoretical developments for the residual effects are required. For instance, among alternative mass calibration functions for FT-ICR MS,¹⁶ the calibration functions incorporating ion abundances aim to provide a second-order correction for ion-ion interactions.^{17–20} These studies resulted in a number of calibration approaches implemented and evaluated for FT-ICR MS-based applications.^{21–24} In Orbitrap FTMS, importance of the residual effects has also been realized;²⁵ however, further studies are required.

Alternatively, unlike the mass calibration functions with theoretical bases, purely empirical estimation of the mass calibration function may also be performed, e.g. in regression-type approaches,^{26–28} as well as methods where the mass calibration function is not obtained explicitly,^{29–32} including the lock-mass approach.^{33–36} Finally, once a sufficiently accurate estimate $M_{\text{est}}(f, p_1, \dots, p_m)$ of the mass calibration function is available, the observed mass-to-charge ratios $(m/z)_{\text{obs}}$ can be recalculated to obtain mass-to-charge ratios of improved accuracy $(m/z)_{\text{corr}}$ by means of solving eq 5 for the ion frequency: $f = M_{\text{def}}^{-1}((m/z)_{\text{obs}})$, where M_{def}^{-1} denotes the inverse function of M_{def} :

$$(m/z)_{\text{corr}} = M_{\text{est}}(f, p_1, \dots, p_m) \quad (6)$$

In the context of mass analysis of petroleomic samples, recalibration is performed on the basis of internal calibrants obtained from up to several prominent homologous series in the mass spectrum.^{37,38} Particularly, in FT-ICR MS-based petroleomics, several recalibration methods are available: (i) simple recalibration, such as the lock-mass approach, which implies empirical estimation of the mass calibration function M_0 , (ii) conventional recalibration, wherein the mass calibration function M_0 is evaluated in the semi-theoretical form, eq 4,³⁷ and (iii) advanced recalibration using the "walking" calibration equation,⁵ where the mass calibration function, eq 2, is evaluated in the form of a piecewise-defined function that matches a semi-theoretical form incorporating a term with ion abundance¹⁹ for each m/z segment. However, for analysis of complex petroleomic samples by Orbitrap FTMS, high performance recalibration methods that routinely deliver suitable mass accuracies have not been published to date.

Thus, in the current work, we present a recalibration method developed for Orbitrap FTMS-based petroleomics. Since a comprehensive theory for residual errors of the default mass calibration, eq 5, is not available for Orbitrap FTMS, the method implements empirical estimation of the mass calibration function M_0 . Importantly, to maximize the accuracy of an estimate of the mass calibration function, an iterative approach with a number of assignment-recalibration steps is implemented to obtain a sufficiently high number of internal calibrants covering the entire ranges of mass-to-charge ratios and abundances of interest. Finally, the described recalibration method is applicable for FT-ICR MS-based petroleomics since the method only takes advantage of the smooth property of the mapping F , eq 1, without the need for theoretical assumptions about the form of this mapping.

EXPERIMENTAL SECTION

Sample Description and Preparation. Of the several petroleomic samples that have been analyzed using the recalibration method described here,³ the resin fraction from a Nigerian crude oil was chosen for the purpose of the current article. The crude oil fraction was obtained via collaboration from Institut Français du Pétrole Energies Nouvelles (IFPEN, Lyon, France). LC-MS grade dichloromethane and acetonitrile were obtained from Fluka (Buchs, Switzerland). Formic acid was obtained from Merck (Zug, Switzerland). The crude oil fraction was dissolved in dichloromethane followed by 4-fold dilution in acetonitrile and addition of 1% (v/v) formic acid.

Instrumentation. MS experiments were performed on a dual linear ion trap Orbitrap Fourier transform mass spectrometer (Orbitrap Elite FTMS, Thermo Scientific, Bremen, Germany) with a high-field compact Orbitrap mass analyzer and the eFT signal processing algorithm. Ions were generated by a nano-electrospray ionization (nESI) ion source (TriVersa Nanomate, Advion Biosciences, Ithaca, NY, USA) operated in chip-based infusion mode, with the ion source's emitter voltage of +1.5 kV and at a flow rate of ca. 200 nL/min. Standard instrument control and data acquisition system was utilized (Thermo Scientific). Temperature of the heated metal capillary was kept at 200 °C. The MS data was acquired in broadband mode, $100 < m/z < 2000$. The automatic gain control (AGC) function was activated with a target value of 10^6 charges to be accumulated in the linear ion trap before the ions were transmitted to the C-trap and then injected into the Orbitrap. The mass spectra were acquired with resolving power of ca. 480 000 at 400 m/z , with the eFT option enabled, which corresponds to a transient signal's acquisition

time of 1.536 s. For a mass spectrum obtained in each MS scan, the frequency-to- m/z conversion was automatically processed using the default mass calibration function, eq 5, obtained via external mass calibration, which was performed following standard procedures for instrument tuning and calibration. To yield the final mass spectrum, 400 single MS scans were acquired and averaged.

Preprocessing of Mass Spectra. From the acquired raw files, the mass spectra were exported into text-format files available for further processing. The PyFTMS library, which was developed in house and written in Python and C/C++ programming languages, was used to: (i) locate peak apexes via parabolic interpolation of local maxima, (ii) cut the noise level at a preselected appropriate threshold value,³ and (iii) recalibrate the mass spectra using the method presented in the current article.

Assignment of Elemental Compositions. For the initial mass spectrum, which was obtained using the default mass calibration, eq 5, two prominent homologous series spanning the whole m/z range of interest and belonging to the nitrogen heteroatom class were assigned manually. In the manual assignment, ChemCalc web-resource (Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland) was used to generate elemental compositions for a given monoisotopic mass and mass tolerance window.^{3,39} Additionally, to validate the assignments, the crude oil fraction was spiked with calibrants (caffeine and Ultramark 1621) in separate experiments.

In all consequent analyses, mass spectral assignments were performed using Composer (Sierra Analytics, Modesto, CA, USA) in “hydrocarbon rules” and “compute all compositions” modes as detailed elsewhere,³ with manual verification of selected parts of the mass spectrum. Kendrick plots as well as spectral slices plots were used to control the identification consistency.^{40,41}

The recalibration was performed in three iterations. After the final iteration, a full-scale assignment of elemental compositions for the obtained mass spectrum was performed to characterize the heteroatom classes composition of the sample under consideration.

RESULTS AND DISCUSSION

Mass Calibration. When the default mass calibration, eq 5, is utilized, its residual error leads to systematic errors in observed mass-to-charge ratios, $(m/z)_{\text{obs}}$. We introduce the mass error function ε , which is the error of the default mass calibration, M_{def} relative to the actual mass calibration function, M_0 , eq 2:

$$\varepsilon(f, p_1, \dots, p_m) = [M_{\text{def}}(f) - M_0(f, p_1, \dots, p_m)] / M_0(f, p_1, \dots, p_m) \quad (7)$$

At this point, we limit the scope of analysis and consider the contribution of all other parameters in eq 1 to be negligible with respect to that of abundance A of a given analyte with frequency f . Therefore, the mass error function, eq 7, is a function of ion frequencies and abundances only: $\varepsilon = \varepsilon(f, A)$. Additionally, from eqs 1 and 5, it follows that $f = F(m/z, A)$ and $f = M_{\text{def}}^{-1}((m/z)_{\text{obs}})$. Therefore, when required in further analysis, the mass error function may be alternatively expressed in terms of either true or observed mass-to-charge ratios: $\varepsilon = \varepsilon(m/z, A)$ and $\varepsilon = \varepsilon((m/z)_{\text{obs}}, A)$, respectively.

Note that, within the estimation of the mass error function, an estimate M_{est} of the mass calibration function M_0 is automatically

provided, as follows from eq 7. Thus, once the mass error function is determined, the recalibration procedure, eq 6, may be applied to the mass spectrum in order to obtain mass-to-charge ratios of improved accuracy:

$$(m/z)_{\text{corr}} = (m/z)_{\text{obs}} / [\varepsilon((m/z)_{\text{obs}}, A) + 1] \quad (8)$$

Internal Calibrants. Here, we consider only monoisotopic species as internal calibrants in a mass spectrum since m/z values and abundances measured for non-monoisotopic species may be additionally shifted due to fine structure interference. From eq 7, it follows that, for estimation of the mass error function, a set of $n + 1$ experimental points $\{(m/z)_j, A_j, \varepsilon_j\}$, $j = 0, \dots, n$, can be provided by $n + 1$ internal calibrants:

$$\varepsilon_j = \varepsilon((m/z)_j, A_j) = [(m/z)_{\text{obs},j} - (m/z)_j] / (m/z)_j \quad (9)$$

The accuracy of such an estimate depends on the number of calibrants, $n + 1$, but is ultimately limited by their random errors. Therefore, we aim to maximize the accuracy by increasing the number of calibrants in eq 9 until a certain value, when the accuracy reaches its limit due to insufficient signal-to-noise ratio (SNR) of the remaining candidates for calibrants.

We note that, if the number of employed calibrants is insufficient to maximize the accuracy of M_{est} , then the potential of experimental data is not entirely released. Therefore, to estimate the mass calibration function with maximized accuracy, we implement an iterative approach of assignment followed by recalibration. In each iterative step, (i) assignment of elemental compositions is performed anew within the limits of available mass accuracy for the mass spectrum obtained in a previous iteration (in the case of the first iteration, the assignment is performed for the initial mass spectrum), (ii) then, an estimate of the mass calibration function is obtained via estimation of the mass error function using all of the assigned monoisotopic peaks as internal calibrants, (iii) and finally, the initial mass spectrum is recalibrated using the obtained estimate of the mass calibration function. In doing so, in each iterative step, the mass calibration function's estimate is obtained with increased accuracy, which in turn provides an increased number of calibrants, as peak assignment is limited by mass accuracy constraints on each iteration. Finally, the accuracy of the estimate reaches its limit since random errors become prevailing in the mass errors of the remaining monoisotopic peaks. Thus, after the final iteration, the accuracy of the estimate of the mass calibration function is maximized. The recalibrated mass spectrum is then available for a full-scale MS analysis consisting of assignment of elemental compositions.

Binomial Average. Since the mapping F is smooth, eq 1, the mass error function ε is a smooth function. Taking this into account, we estimate the function ε on the basis of available points ε_j , eq 9. First, we consider the effect of all parameters p_i in eq 1 to be negligible. Thus, suppose for an unknown smooth mass error function $\varepsilon(m/z)$, a set of $n + 1$ points $\{(m/z)_j, \varepsilon_j^{(\text{exp})}\}$, $\varepsilon_j^{(\text{exp})} = \varepsilon((m/z)_j) + \Delta_j$, $j = 0, \dots, n$, is given, where Δ_j represents the random errors in measurements of the sample points $\varepsilon_j = \varepsilon((m/z)_j)$ of the mass error function; it is then required to find an estimate $\varepsilon_{\text{est}}(m/z)$ of $\varepsilon(m/z)$ such that the error of the estimate is minimized:

$$|\varepsilon_{\text{est}}(m/z) - \varepsilon(m/z)| \rightarrow 0 \quad (10)$$

However, Δ_j is not constant and may vary with the calibrants. Thus, the solution for eq 10 is complicated by an unknown variation of Δ_j . An approximate solution for this problem may be obtained using various approaches including averages,

polynomials, splines, etc. However, the downside of the approximation is that the condition in eq 10 might be highly compromised such that the obtained estimate $\varepsilon_{\text{est}}(m/z)$ is not of acceptable accuracy. Therefore, we aim to find an approximate solution for eq 10 of sufficient accuracy.

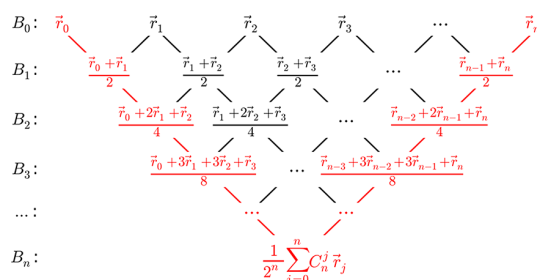
Perhaps the simplest approximate solution for eq 10 is the central moving average (CMA) of order k , where $k + 1$ is the number of data points to be averaged in a moving window. For lower k , the CMA is prone to overfitting, i.e., a situation when the noise Δ_j is fitted instead of the underlying relationship. Contrastingly, for higher k , the CMA is not able to adequately represent the approximated relationship due to being non-local, meaning that for any given point the CMA strongly depends on distant points as well as the point of interest (data not shown). Following this consideration, one may show that the solution for eq 10 can be written as a weighted average, with appropriately selected weights w_j as a function of random errors Δ_j or, equivalently, the m/z values:

$$\varepsilon(m/z) = \sum_{j=0}^n w_j(m/z) \varepsilon_j^{\text{(exp)}} \quad (11)$$

Qualitatively, eq 11 shows that the function of interest ε is determined by means of passing the set of its samples $\varepsilon_j^{\text{(exp)}}$ through a noise filter suppressing the random errors Δ_j .

Next, we approximate eq 11 as illustrated in Scheme 1. Let vectors \vec{r}_j denote the points $\{(m/z)_j, \varepsilon_j^{\text{(exp)}}\}$, $j = 0, \dots, n$. To each

Scheme 1. Binomial Average for Estimation of the Mass Error Function^a



^aThe upper row shows all $n + 1$ data points. Each B_k row, where $0 \leq k \leq n$, represents the binomial central moving average of order k , i.e., the weighted central moving average of order k with the weights being the normalized binomial coefficients. The binomial average function applied in this work is shown in red and constructed as the binomial central moving average of variable order k such that for each approximated point the maximum available k is employed.

pair \vec{r}_l and \vec{r}_{l+1} , $l = 0, \dots, n - 1$, we assign the average $\vec{r}_l^{(1)} = (\vec{r}_l + \vec{r}_{l+1})/2$; further, to each pair $\vec{r}_l^{(1)}$ and $\vec{r}_{l+1}^{(1)}$, $l = 0, \dots, n - 2$, we assign the average $\vec{r}_l^{(2)} = (\vec{r}_l^{(1)} + \vec{r}_{l+1}^{(1)})/2$; at a step k , to each pair $\vec{r}_l^{(k-1)}$ and $\vec{r}_{l+1}^{(k-1)}$, $l = 0, \dots, n - k$, we assign the average $\vec{r}_l^{(k)} = (\vec{r}_l^{(k-1)} + \vec{r}_{l+1}^{(k-1)})/2$ and so forth. At the final step n , to the pair $\vec{r}_l^{(n-1)}$ and $\vec{r}_{l+1}^{(n-1)}$, $l = 0, \dots, n - 1$, we assign the average $\vec{r}_l^{(n)} = (\vec{r}_l^{(n-1)} + \vec{r}_{l+1}^{(n-1)})/2$. Constructed this way, the function $B_k: \{\vec{r}_0, \vec{r}_1, \dots, \vec{r}_n\} \rightarrow \{\vec{r}_0^{(k)}, \vec{r}_1^{(k)}, \dots, \vec{r}_{n-k}^{(k)}\}$, $0 \leq k \leq n$, is the binomial central moving average, i.e., the weighted central moving average with the weights being the normalized binomial coefficients:

$$\vec{r}_l^{(k)} = \frac{1}{2^k} \sum_{j=0}^k C_k^j \vec{r}_{l+j}, \quad 0 \leq l \leq n - k \quad (12)$$

where C_k^j is a binomial coefficient indexed by k and j , $0 \leq j \leq k$.

Finally, we implement the binomial central moving average of variable order k such that for each $(m/z)_j$, the maximum available value of k is employed, as shown in red in Scheme 1. This way, for each $(m/z)_j$, the estimation is performed using the maximized number of points \vec{r}_j . The values of ε corresponding to intermediate points m/z are obtained via linear approximation. Henceforth, the obtained function is referred to as the binomial average, for the sake of brevity. We find the weights in the form of binomial coefficients with maximized order to be a suitable approximation of eq 11 for the purpose of this work since the binomial average is locally flexible enough to adequately represent the approximated mass error function without overfitting taking place.

Piecewise Approximation and Assignment-Recalibration Iterations. Recall that we parametrize the mass error function ε such that only the mass-to-charge ratio m/z and abundance A of a given analyte, eq 1, are considered: $\varepsilon = \varepsilon(m/z, A)$. Since it is common for chemical species from the same homologous series of petroleomic samples to show gradual dependence of their abundances with variation of their m/z value, a convenient piecewise approximation of the function ε is available.⁵ Thus, we approximate the function $\varepsilon(m/z, A)$ with a piecewise function $\varepsilon_j(m/z, A)$, $j = 0, \dots, n$, defined throughout n intervals, and break the estimation problem into two independent estimation problems for the two parameters, m/z and A , as follows:

$$\varepsilon_j(m/z, A) = \varepsilon(m/z) + \varepsilon_j(A) \quad (13)$$

where $\varepsilon(m/z)$ and $\varepsilon_j(A)$ represent the m/z parametrization and abundance parametrization, respectively, with the latter being piecewise-defined for each j^{th} m/z segment. Specifically, for the m/z parametrization, we consider calibrants differing by one CH_2 unit, $\sim 14 m/z$, from a prominent homologous series; for the abundance parametrization, calibrants of different double bond equivalent (DBE) types and up to several heteroatom classes are selected in each $\pm 7 m/z$ window relative to the calibrants employed for the m/z parametrization. The function $\varepsilon(m/z)$ is estimated using the binomial average as described above; the function $\varepsilon_j(A)$ is estimated analogously for each segment j .

The first assignment-recalibration iteration is performed as follows. Only m/z parametrization is considered since the high number of calibrants needed for the abundance parametrization cannot be efficiently obtained using manual assignment only. Specifically, Figure 1 shows the first assignment step, the aim of which is to obtain calibrants for the following m/z -based recalibration such that they are of sufficiently high abundance and span as broad m/z range as possible. The total number of 48 calibrants from two homologous series, $[\text{C}_{15+n}\text{H}_{14+2n}\text{N}_1]^+$ and $[\text{C}_{19+n}\text{H}_{24+2n}\text{N}_1]^+$, was assigned manually and shown in Figure 1, top panel. The calibrants span the m/z range of $208.1 < m/z < 854.8$.

To validate the assignments, we introduce the concept of spectral slices, where for a given $(m/z)_{\text{true}}$ value and a half-width Δ of m/z window of interest mass errors of candidate calibrants with respect to $(m/z)_{\text{true}}$ are plotted for a slice of the mass spectrum for $[(m/z)_{\text{true}} - \Delta \leq m/z \leq (m/z)_{\text{true}} + \Delta]$. The located calibrants demonstrate gradual dependence of their mass errors on the m/z values and abundances, Figure 1 bottom panel. Thus, the data is in agreement with the smoothness property of the mass error function and hence supports the identification consistency. Note that picking analytes on the basis of the lowest mass errors would lead to mis-identification as may be seen for the two m/z values below 300 m/z . Therefore, the spectral slices plot aids in correct assignment of analyte. 129

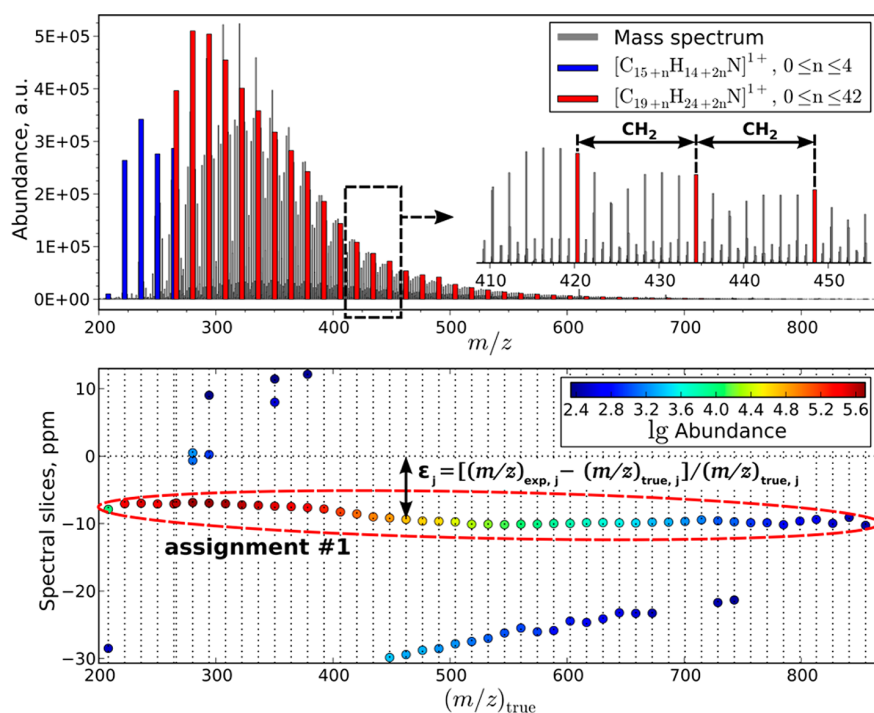


Figure 1. Calibrants assigned in the first assignment-recalibration iteration. The calibrants are shown in the mass spectrum (top panel) and the spectral slices plot (bottom panel). Two prominent homologous series were identified manually. There are 48 calibrants of 2 DBE types belonging to the nitrogen heteroatom class. The spectral slices are plotted for windows of 30 ppm. Vertical dotted lines indicate the mass-to-charge ratios for the analytes of interest. Outlier dots indicate neighboring peaks in spectral slices.

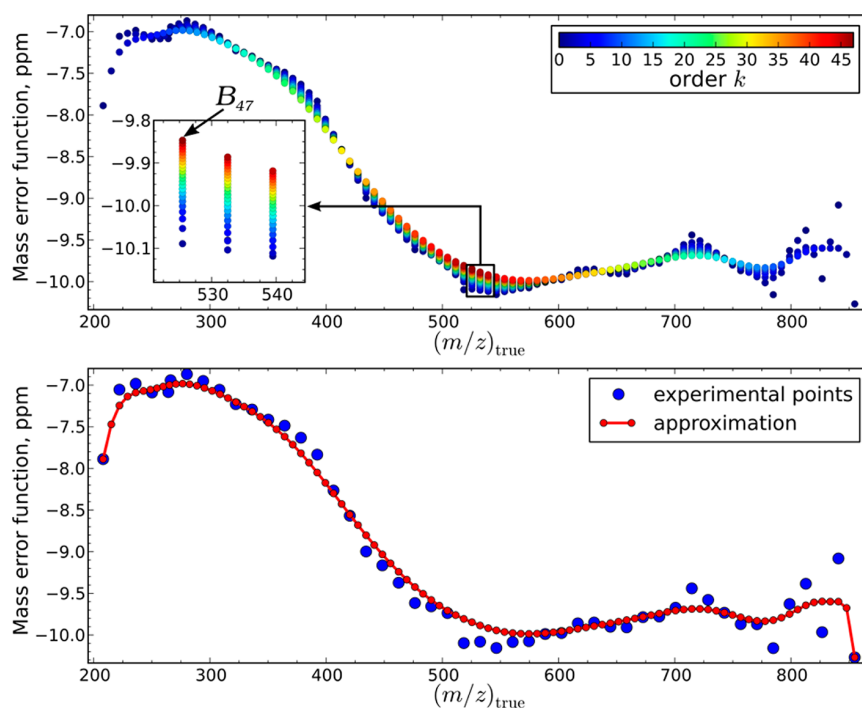


Figure 2. Estimation of the mass error function in the first assignment-recalibration iteration. The top panel demonstrates a set of binomial central moving averages of order k , $0 \leq k \leq 47$, obtained on the basis of the calibrants available. The inset shows the behavior of the binomial average B_k as k increases. The bottom panel shows the binomial average constructed from the set of the binomial central moving averages.

The assignment is followed by the estimation of the mass error function using the binomial average, Figure 2. Specifically, Figure 3, top panel, shows a set of binomial central moving

averages of order k , $0 \leq k \leq 47$, whereas the final binomial average is shown in Figure 2, bottom panel. When the initial mass spectrum is recalibrated using the estimated mass calibration

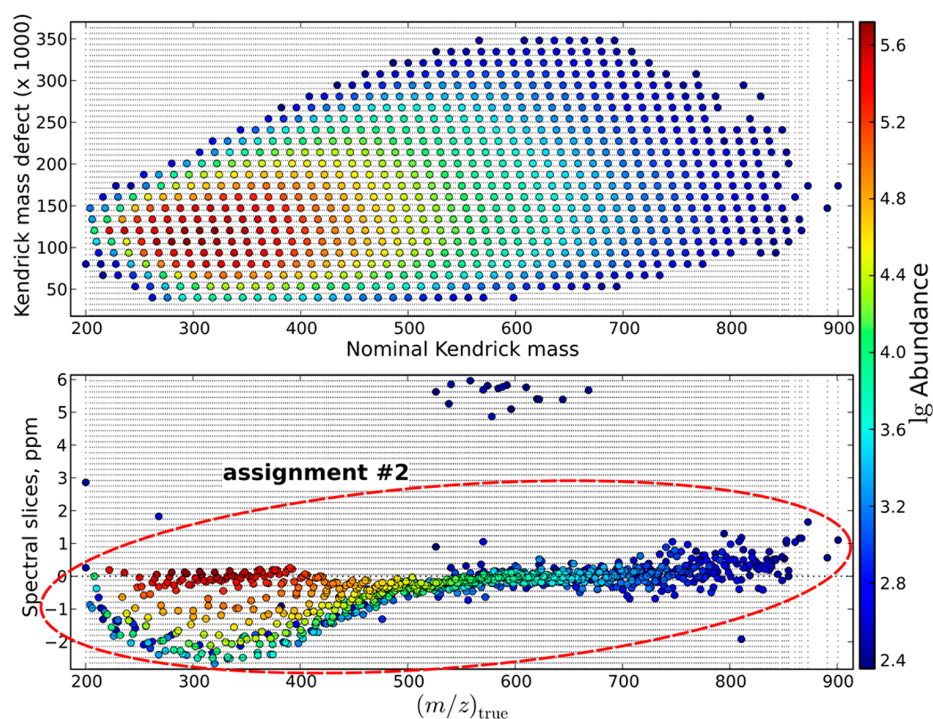


Figure 3. Calibrants assigned in the second assignment-recalibration iteration. The calibrants are shown in the Kendrick plot (top panel) and the spectral slices plot (bottom panel). There are 843 calibrants of 24 DBE types belonging to the nitrogen heteroatom class. The spectral slices are plotted for windows of 6 ppm. Vertical dotted lines indicate the mass-to-charge ratios for the analytes of interest.

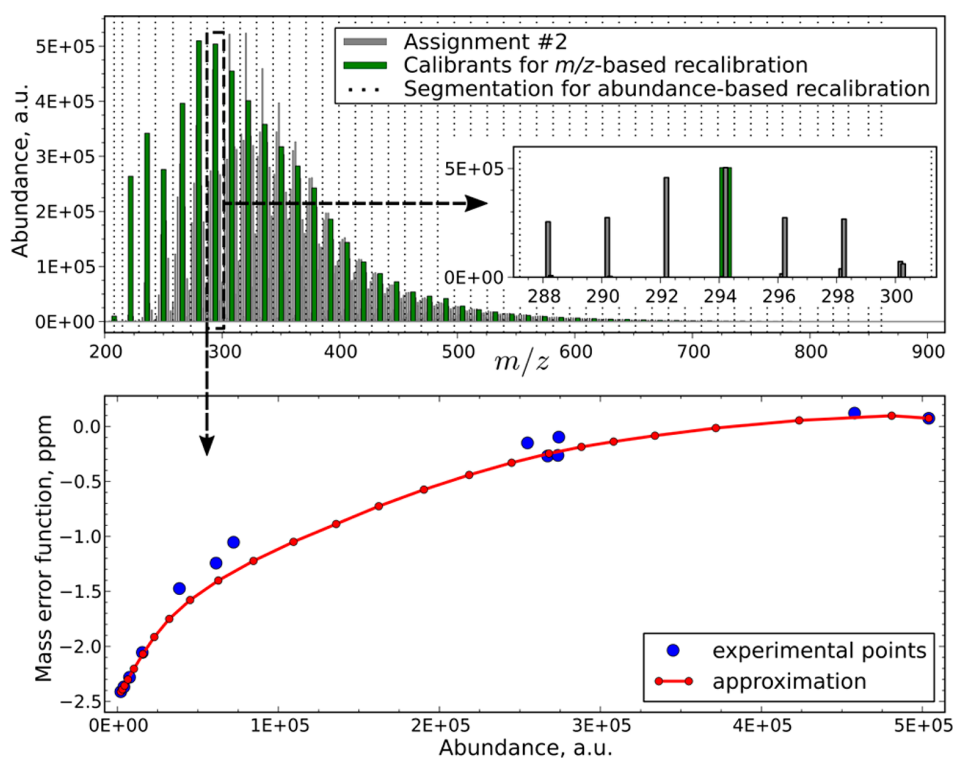


Figure 4. Estimation of the mass error function in the second assignment-recalibration iteration. The top panel demonstrates the test of locally most abundant calibrants for the m/z -based recalibration and shows the segmentation for the piecewise estimation of the mass error function. Dashed lines indicate the segments for estimation of $\epsilon_j(A)$. The bottom panel exemplifies the estimation of the mass error function for the m/z window of $287.2 < m/z < 301.2$.

function, the obtained mass spectrum is available for further analysis.

To further improve the accuracy of the estimate of the mass calibration function, the next assignment-recalibration iteration

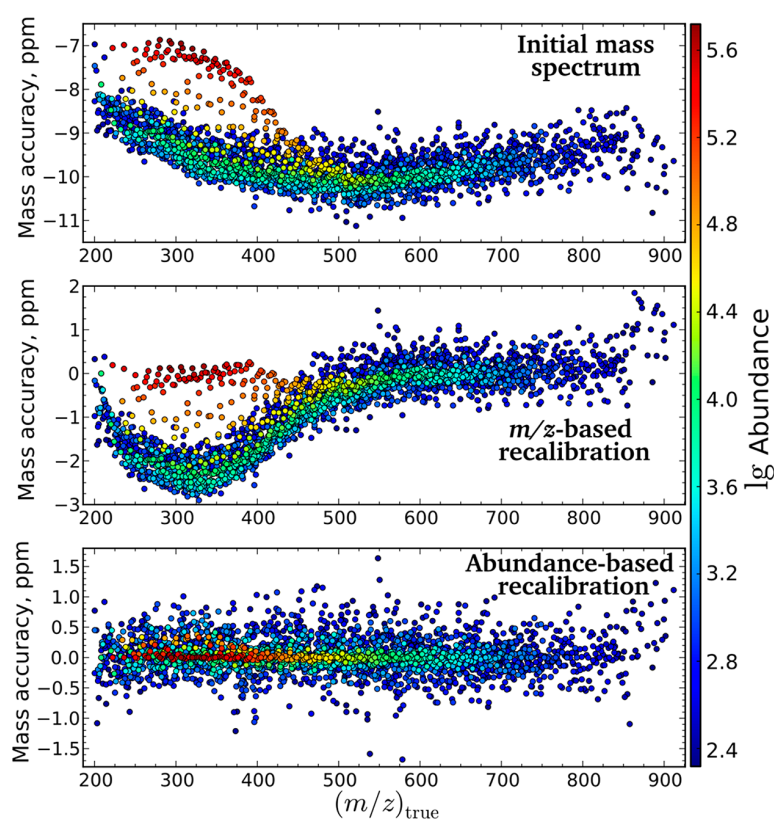


Figure 5. Validation of the final estimate of the mass calibration function. Mass accuracy vs m/z and abundance plots are shown for: the initial mass spectrum (top panel), the mass spectrum obtained after an intermediate m/z -based recalibration (middle panel), and the final mass spectrum which was obtained after a subsequent abundance-based recalibration (bottom panel). The monoisotopic peaks of all elemental compositions assigned in this work were tracked throughout the three recalibration stages.

is performed. In this iteration, a sufficient number of calibrants is available for estimation of the mass error function in terms of both m/z values and abundances. Specifically, Figure 3 shows the assignment performed for the mass spectrum obtained in the previous iteration. The total number of 843 calibrants of 24 DBE types, $4 \leq \text{DBE} \leq 27$ (values given for neutral species), from the N heteroatom class was assigned and is shown in a Kendrick plot in Figure 3, top panel. Gradual variation of the abundances in the Kendrick plot supports the identification consistency. The calibrants span the m/z range of $200.1 < m/z < 900.8$ and abundance range of $230 < A < 5.2 \times 10^5$. The spectral slices plot for the found calibrants, Figure 3, bottom panel, indicates the smoothness property of the mass error function with respect to m/z and abundance values. Note that a dependence of the mass errors on abundances is also demonstrated.

On the basis of the calibrants obtained, the mass error function is estimated anew, Figure 4. Here, both m/z and abundance parametrization are implemented. The m/z -based parametrization is largely similar to that of the first iteration with the following difference: although the calibrants employed in the first iteration were of high abundance, their abundances might not be the local maxima. Therefore, since a high number of calibrants is now available, the calibrants to be chosen for the m/z -based recalibration here are selected on the basis of being locally most abundant peaks in order to ensure that the m/z -based calibration is not occasionally performed on the basis of low abundant monoisotopic peaks if a high-abundance peak is available instead. For that, we scan the m/z and abundances of the calibrants obtained in the manual assignment of the first iteration against

those of the calibrants obtained in the current iteration. If a tested calibrant from the former set is a local abundance maximum in the latter set, such calibrant is selected as shown in Figure 4, top panel. Otherwise, a local maximum around the tested calibrant is determined in the latter set, and the found calibrant is selected. Additionally, Figure 4 top panel, shows m/z windows of abundance parametrization for piecewise approximation of the mass error function. An example of the abundance-based estimation of the mass error function for the m/z window of $287.2 < m/z < 301.2$ is illustrated in Figure 4, bottom panel. Additionally, this example demonstrates a limitation of the binomial average: since the curve has no inflection points, the binomial average shows reduced performance around the center region.

The initial mass spectrum is now recalibrated using the estimated mass calibration function, and the obtained mass spectrum is available for the next iterations. The consequent assignment-recalibration iterations are repeated precisely as in the second iteration, Figures 3 and 4, until there is no notable dependence of the mass error function on the m/z values and abundances; i.e., the mass error spread in spectral slices, Figure 3 bottom, is largely due to random errors. In this particular analysis, three assignment-recalibration iterations were performed.

Recalibration Performance. With the final estimate of the mass calibration function obtained, the recalibrated mass spectrum is available for assignment of elemental compositions. As a result, 2804 elemental compositions were identified,³ the assigned analytes belong to 10 heteroatom classes (N, NO, N₂, NO₂, OS, NS, N₂O, O, O₂, HC). Their monoisotopic peaks

span a range of ion abundances of 2520:1 and the m/z range of $200.1 < m/z < 911.8$. For reference, we note that the total number of peaks within this m/z range in the mass spectrum is 5424.

Importantly, once the elemental compositions are assigned, we are able to track the location of each individual peak back to all stages and iterations of the recalibration. Thus, to assess the method performance, we trace how the mass accuracy for all identified monoisotopic peaks is changed as a function of true m/z and experimental ion abundances at three key stages of the recalibration: the initial mass spectrum (with default mass calibration); the mass spectrum after m/z -based recalibration; and the mass spectrum after abundance-based recalibration, as illustrated in Figure 5.

Figure 5, top panel, shows mass accuracies for the initial mass spectrum. In this particular case, there was a significant, 7–11 ppm, mass error. Also, a noticeable difference in mass accuracy spread can be observed for peaks of various abundances. For the peaks of intermediate abundance, approximately $3.2 < \lg A < 4.4$, the local spread in mass accuracy is less than 1 ppm. For the peaks of low abundance, with $\lg A < 3.2$, the local spread of mass accuracies is higher, 1–2 ppm, and presumably is due to the increasing significance of random errors for peaks with low signal-to-noise ratios. Finally, mass accuracies of peaks of high abundance, with $\lg A > 4.4$, deviate from the general trend and show dependence on the abundances, which likely results from the increased effect of ion–ion interactions for high-abundance ions.

Next, the result of the intermediate m/z -based recalibration is shown in Figure 5, middle panel. Overall, the mass accuracy is improved, as expected. However, since this intermediate step is based on high-abundance analyte peaks only, the abundance-dependent spread of ~ 3 ppm in mass accuracy remains. Finally, the abundance-based recalibration corrects the abundance-dependent spread in mass accuracies, Figure 5, bottom panel.

Thus, the recalibration shows efficient elimination of both m/z - and abundance-dependent systematic shifts in mass accuracies for the tested set of all monoisotopic peaks assigned, Figure 5. Importantly, the data indicates that the achieved mass accuracy is limited mainly by random errors in m/z measurements for low-abundance peaks with ca. $\lg A < 3.2$. Albeit, a small number of abundant peaks demonstrated worse mass accuracy values compared to that of other high-abundance peaks with similar m/z , Figure 5 bottom panel. This indicates that the accuracy of the obtained mass calibration function in those local regions of abundances is limited by the approximation function itself, as discussed above. Nevertheless, we consider that acceptable since said mass accuracy is still better than the random errors level, which is the actual limiting factor in the data analysis.

Mass Accuracy Distributions. Figure 6 compares the mass accuracy distributions corresponding to the three recalibration stages outlined above. Figure 6, top panel, shows the mass accuracy distribution for the initial mass spectrum, with the root-mean-square (RMS) mass accuracy of 9.63 ppm, the mean of -9.61 ppm, and the standard deviation of 0.625 ppm. Notably, the distribution has a small sidelobe, which is due to high-abundance monoisotopic peaks, as follows from Figure 5, top panel. The intermediate m/z -based recalibration improved the RMS mass accuracy to 1.22 ppm, Figure 6, middle panel. However, an additional maximum in the distribution appeared, which is due to the abundance correction not being taken into account, resulting in low-abundance peaks having significant

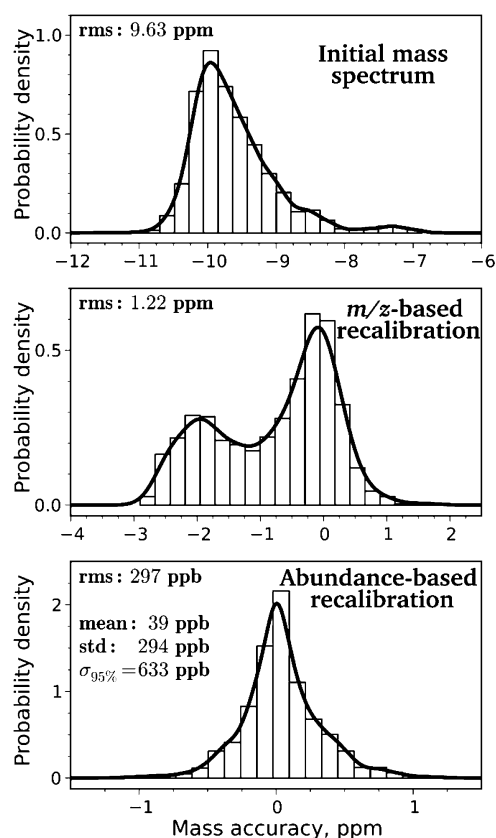


Figure 6. Mass accuracy distributions throughout the final recalibration. The distributions are shown for: the initial mass spectrum (top panel), the mass spectrum obtained after an intermediate m/z -based recalibration stage (middle panel), and the final mass spectrum which was obtained after a subsequent abundance-based recalibration (bottom panel). The probability density functions were calculated using kernel density estimation (KDE) with Gaussian kernels. The histograms were calculated with 20, 20, and 28 bins, respectively, and are shown for reference.

deviations in mass-to-charge ratios in regions of the mass spectrum where m/z -only calibration was performed on high-abundance peaks, mainly in the 200–500 m/z range, as follows from Figure 5, middle. Finally, the abundance-based recalibration provided the mass accuracy distribution, Figure 6, bottom panel, with the RMS mass accuracy of 0.297 ppm, the mean of 0.039 ppm, the standard deviation of 0.294 ppm, and with 95% of the monoisotopic peaks being within the ± 0.633 ppm range away from the mean.

CONCLUSIONS

The described recalibration method routinely provides RMS mass accuracies at the level of hundreds ppb, typically ~ 300 ppb, for analysis of petroleum-type samples with a high-field Orbitrap FTMS. Hence, the method allows for adequate assignment of chemical species from petroleomic samples of relatively high complexity, e.g., at least up to and including maltene fractions.³ Importantly, the method reduces the mass measurement errors to the level of random errors and hence maximizes the range of species' abundances over which reliable assignments can be made. The achieved mass accuracies are limited mainly by random errors level, which is imposed partially by the limited

scope of the parametrization employed in this work and largely by the SNR of experimental MS data.

AUTHOR INFORMATION

Corresponding Author

*Phone: +41 21 693 97 51. Fax: +41 21 693 98 95. E-mail: yury.tsybin@epfl.ch.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Luc Patiny of EPFL for the ChemCalc web-resource, Jérémie Ponthus of IFPEN for the petroleum samples provided, and David Stranz of Sierra Analytics for technical support. This work was financially supported by the Swiss National Science Foundation (projects 200021-125147 and 200021-147006), the Joint Russia–Switzerland Research Program (grant 128357), and the European Research Council (ERC Starting grant 280271).

REFERENCES

- (1) Rodgers, R. P.; McKenna, A. M. *Anal. Chem.* **2011**, *83*, 4665–4687.
- (2) Hsu, C. S.; Hendrickson, C. L.; Rodgers, R. P.; McKenna, A. M.; Marshall, A. G. *J. Mass Spectrom.* **2011**, *46*, 337–343.
- (3) Zhurov, K. O.; Kozhinov, A. N.; Tsybin, Y. O. *Energy Fuels* **2013**, DOI: 10.1021/ef400203g.
- (4) Kim, S.; Rodgers, R. P.; Marshall, A. G. *Int. J. Mass Spectrom.* **2006**, *251*, 260–265.
- (5) Savory, J. J.; Kaiser, N. K.; McKenna, A. M.; Xian, F.; Blakney, G. T.; Rodgers, R. P.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2011**, *83*, 1732–1736.
- (6) Brown, L. S.; Gabrielse, G. *Rev. Mod. Phys.* **1986**, *58*, 233–311.
- (7) Shi, S. D. H.; Drader, J. J.; Freitas, M. A.; Hendrickson, C. L.; Marshall, A. G. *Int. J. Mass Spectrom.* **2000**, *195*, 591–598.
- (8) Makarov, A. *Anal. Chem.* **2000**, *72*, 1156–1162.
- (9) Francl, T. J.; Sherman, M. G.; Hunter, R. L.; Locke, M. J.; Bowers, W. D.; McIver, R. T. *Int. J. Mass Spectrom. Ion Processes* **1983**, *54*, 189–199.
- (10) Ledford, E. B.; Rempel, D. L.; Gross, M. L. *Anal. Chem.* **1984**, *56*, 2744–2748.
- (11) Gorshkov, M. V.; Good, D. M.; Lyutvinskiy, Y.; Yang, H. Q.; Zubarev, R. A. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1846–1851.
- (12) Belov, M. E.; Zhang, R.; Strittmatter, E. F.; Prior, D. C.; Tang, K.; Smith, R. D. *Anal. Chem.* **2003**, *75*, 4195–4205.
- (13) Syka, J. E. P.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; Ueberheide, B.; Garcia, B.; Busby, S.; Muratore, T.; Shabanowitz, J.; Hunt, D. F. *J. Proteome Res.* **2004**, *3*, 621–626.
- (14) Makarov, A.; Denisov, E.; Kholomeev, A.; Baischun, W.; Lange, O.; Strupat, K.; Horning, S. *Anal. Chem.* **2006**, *78*, 2113–2120.
- (15) Makarov, A.; Denisov, E.; Lange, O.; Horning, S. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 977–982.
- (16) Zhang, L. K.; Rempel, D.; Pramanik, B. N.; Gross, M. L. *Mass Spectrom. Rev.* **2005**, *24*, 286–309.
- (17) Easterling, M. L.; Mize, T. H.; Amster, I. J. *Anal. Chem.* **1999**, *71*, 624–632.
- (18) Burton, R. D.; Matuszak, K. P.; Watson, C. H.; Eyler, J. R. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1291–1297.
- (19) Masselon, C.; Tolmachev, A. V.; Anderson, G. A.; Harkewicz, R.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 99–106.
- (20) Taylor, P. K.; Amster, I. J. *Int. J. Mass Spectrom.* **2003**, *222*, 351–361.
- (21) Muddiman, D. C.; Oberg, A. L. *Anal. Chem.* **2005**, *77*, 2406–2414.
- (22) Wong, R. L.; Amster, I. J. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1681–1691.
- (23) Wong, R. L.; Amster, I. J. *Int. J. Mass Spectrom.* **2007**, *265*, 99–105.
- (24) Smith, D. F.; Kharchenko, A.; Konijnenburg, M.; Klinkert, I.; Pasa-Tolic, L.; Heeren, R. M. A. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1865–1872.
- (25) Kharchenko, A.; Vladimirov, G.; Heeren, R. M. A.; Nikolaev, E. N. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 977–987.
- (26) Williams, D. K.; Muddiman, D. C. *Anal. Chem.* **2007**, *79*, 5058–5063.
- (27) Becker, C. H.; Kumar, P.; Jones, T.; Lin, H. *Anal. Chem.* **2007**, *79*, 1702–1707.
- (28) Petyuk, V. A.; Jaitly, N.; Moore, R. J.; Ding, J.; Metz, T. O.; Tang, K.; Monroe, M. E.; Tolmachev, A. V.; Adkins, J. N.; Belov, M. E.; Dabney, A. R.; Qian, W. J.; Camp, D. G.; Smith, R. D. *Anal. Chem.* **2008**, *80*, 693–706.
- (29) O'Connor, P. B.; Costello, C. E. *Anal. Chem.* **2000**, *72*, 5881–5885.
- (30) Bruce, J. E.; Anderson, G. A.; Brands, M. D.; Pasa-Tolic, L.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 416–421.
- (31) Wu, S.; Kaiser, N. K.; Meng, D.; Anderson, G. A.; Zhang, K.; Bruce, J. E. *J. Proteome Res.* **2005**, *4*, 1434–1441.
- (32) Egertson, J.; Eng, J.; Bereman, M.; Hsieh, E.; Merrihew, G.; MacCoss, M. J. *Am. Soc. Mass Spectrom.* **2012**, *23*, 2075–2082.
- (33) Olsen, J. V.; de Godoy, L. M. F.; Li, G. Q.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. *Mol. Cell. Proteomics* **2005**, *4*, 2010–2021.
- (34) Wenger, C. D.; McAlister, G. C.; Xia, Q. W.; Coon, J. J. *Mol. Cell. Proteomics* **2010**, *9*, 754–763.
- (35) Cox, J.; Michalski, A.; Mann, M. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1373–1380.
- (36) Zhang, Y.; Wen, Z. H.; Washburn, M. P.; Florens, L. *Anal. Chem.* **2011**, *83*, 9344–9351.
- (37) Hughey, C. A.; Rodgers, R. P.; Marshall, A. G. *Anal. Chem.* **2002**, *74*, 4145–4149.
- (38) Klitzke, C. F.; Corilo, Y. E.; Siek, K.; Binkley, J.; Patrick, J.; Eberlin, M. N. *Energy Fuels* **2012**, *26*, 5787–5794.
- (39) Patiny, L.; Borel, A. *J. Chem. Inf. Model.* **2013**, *53* (5), 1223–1228.
- (40) Kendrick, E. *Anal. Chem.* **1963**, *35*, 2146–2154.
- (41) Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G.; Qian, K. N. *Anal. Chem.* **2001**, *73*, 4676–4681.

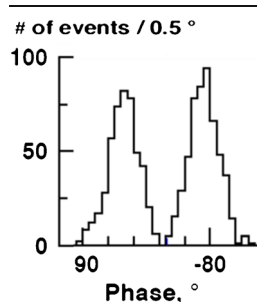
A.4 Paper IV

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

Tagir Aushev,¹ Anton N. Kozhinov,² Yury O. Tsybin²

¹Institute for Theoretical and Experimental Physics, 117218 Moscow, Russia

²Biomolecular 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.

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).

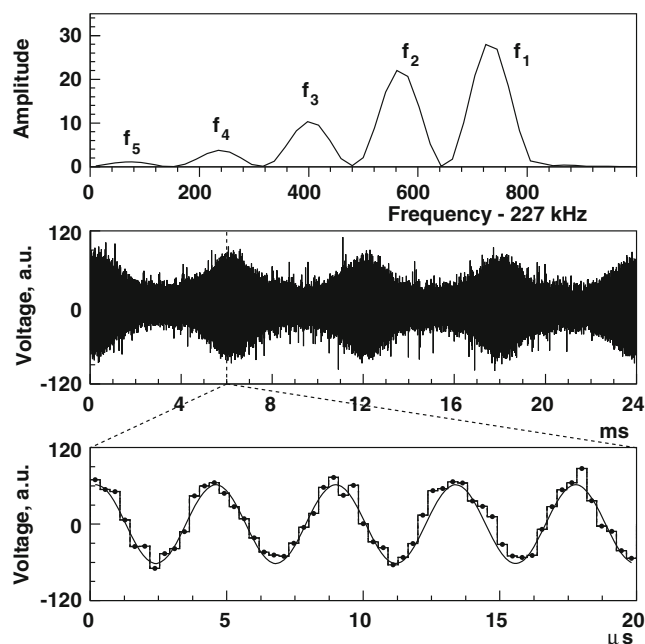


Figure 1. The principle of least-squares fitting (LSF) of transient signals. Top panel shows the magnitude-mode Fourier spectrum of a 24 ms-long experimental transient signal shown in the middle panel. Bottom panel shows an expanded view of the transient with sampled points and a curve corresponding 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 $\Delta_{df} = 1/(2T)$ in the sampled spectrum (in our case $T = 24$ ms and hence $\Delta_{df} \approx 21$ 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

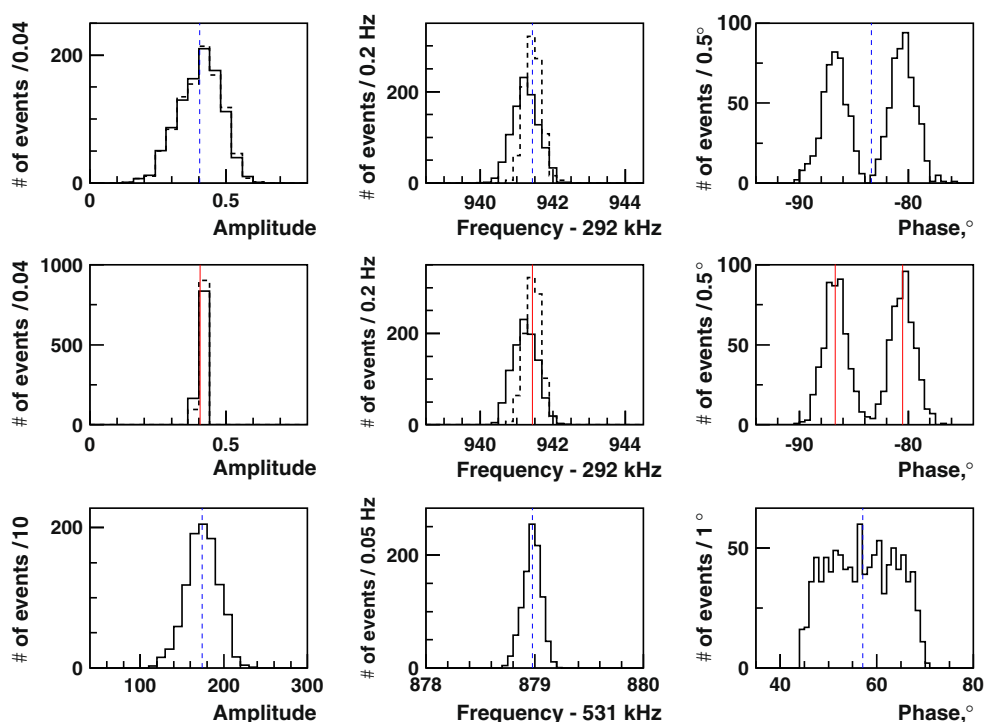


Figure 2. Results of LSF (solid lines) and FT (dashed lines) processing of single-component transients: (top panels) 1000 ICR transients; (middle panels) 1000 simulated transients with frequency $f = 292941.44$ Hz, amplitude $A = 0.405$, standard deviation of noise $\sigma = 0.65$, randomly assigned 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 $\Delta_{\text{wnd}}f / \Delta df = 4 \cdot 2 = 8$ 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 $f_0 \leq f_n \leq f_0 + \Delta df$, 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

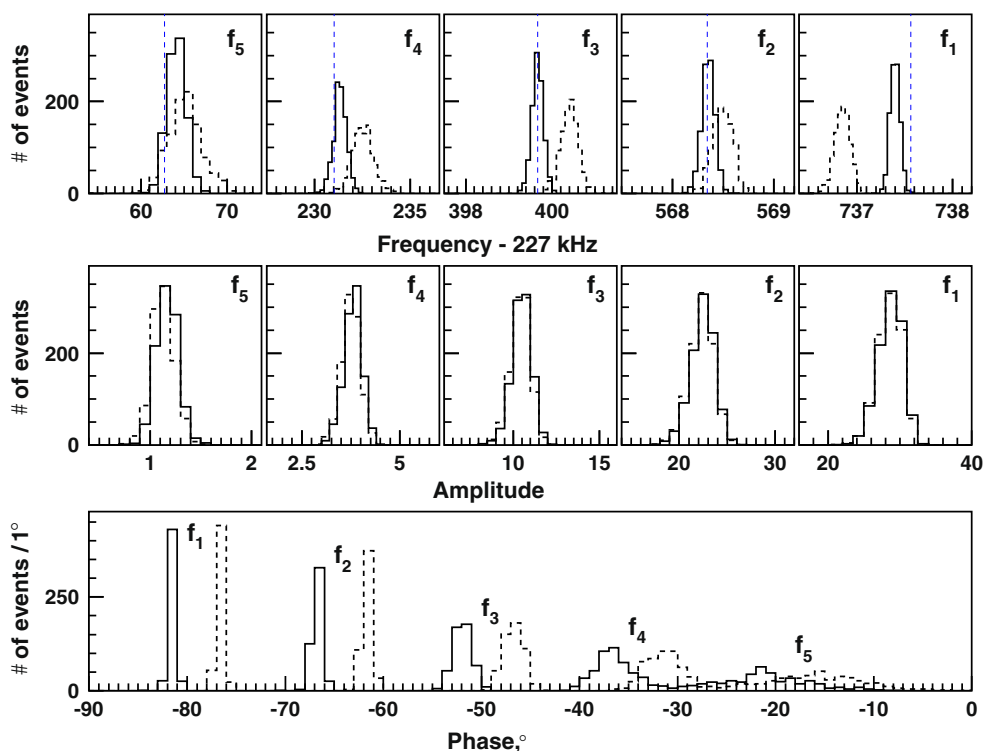


Figure 3. Results of LSF and FT processing of ICR multiple-component transients. Top and middle panels show frequency and amplitude distributions obtained with LSF (solid lines) and FT (dashed lines). Bottom panel shows phase distributions obtained with LSF; solid and dashed lines correspond to the two-mode 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 42. 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 the central electrode potential) may lead to a scan-to-scan variation of

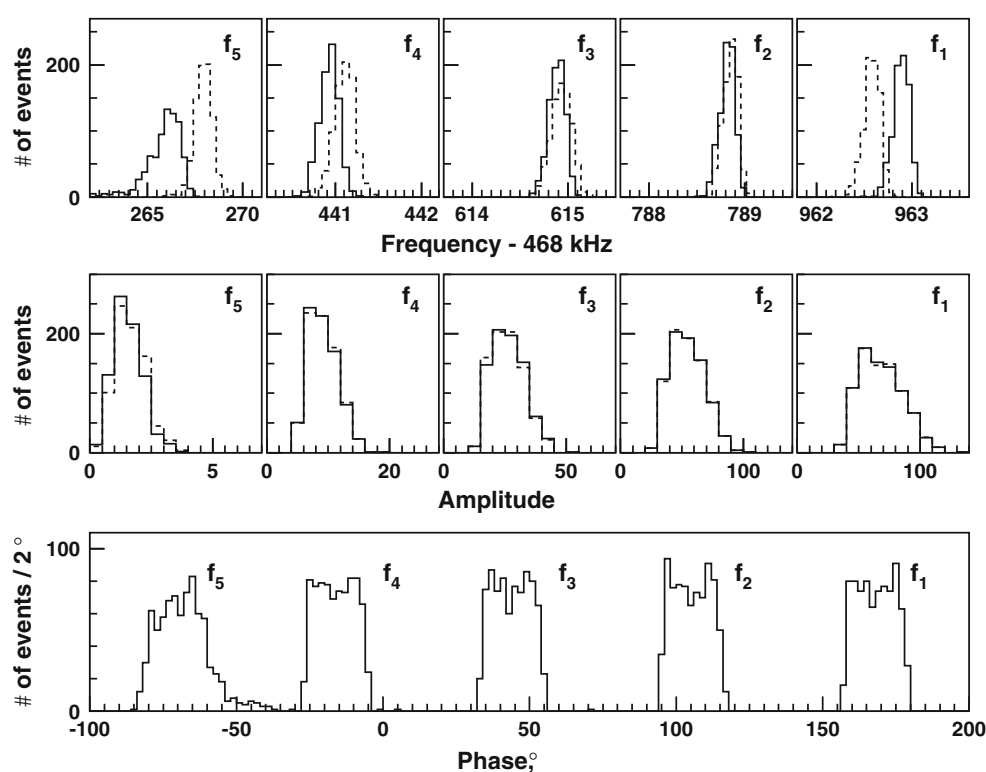


Figure 4. Frequency, amplitude, and phase distributions obtained with LSF (solid lines) and FT (dashed lines) processing of multiple-component Orbitrap transients. The set of 1000 transients with five components of an isotopic distribution of a doubly charged peptide Substance P was obtained experimentally 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 $B \rightarrow Ds1(2536)^+(D)\overline{\nu}^*(\nu)$. *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)

Supporting Information

Least-squares fitting of FTMS transients

Tagir Aushev, Anton N. Kozhinov, and Yury O. Tsybin

Figure S1. Least-squares fitting (LSF) of an experimental transient signal acquired for isolated monoisotopic ions of singly charged peptide MRFA (524 m/z) analyzed on the 10 T FT-ICR MS. Top panel shows a 24 ms long single-scan transient signal. 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.

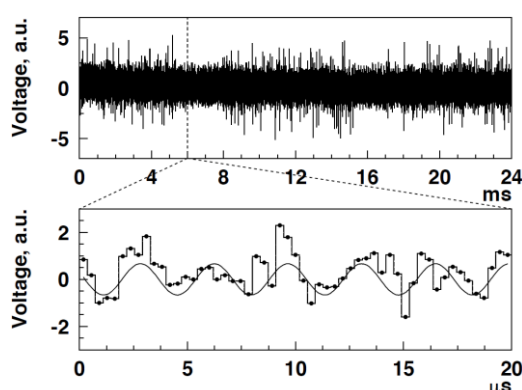


Figure S2. Correlation between the measurement error, Δf , of FT signal processing and the true frequency, f_n , in numerical simulations of frequency measurements performed for modeled sinusoidal signals. Conventional FT signal processing workflow was used, including Hann apodization window and single zero-filing. Out of $N = 10$ points on the plot, the n^{th} point, $n = 0, \dots, N-1$, shows the analysis of the transient signal with frequency $f_n = f_0 + n \Delta f / (N-1)$, where $\Delta f = 1 / (2T)$ is the frequency spacing in the discrete Fourier spectra of these signals. The following parameters corresponding to LSF analysis of the ICR transient signals, Figure 2 and Table 1, were used: length $T = 24$ ms, frequency $f_0 = 292,941.44$ Hz (the leftmost point on the plot), amplitude $A = 0.405$, and phase $\varphi = -86.7^\circ$. Additional results obtained for the other phase $\varphi = -80.5^\circ$ did not differ noticeably compared w.r.t. the effect under consideration (data not shown).

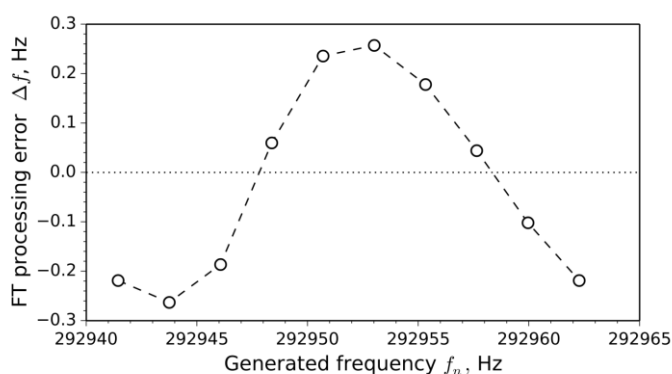


Figure S3. Phase jittering of the monoisotopic ion in the LSF analysis of the set of 1,000 ICR transient signals acquired for the isolated isotopic distribution of a doubly charged peptide substance P. No correlation between the phase and the scan number was found.

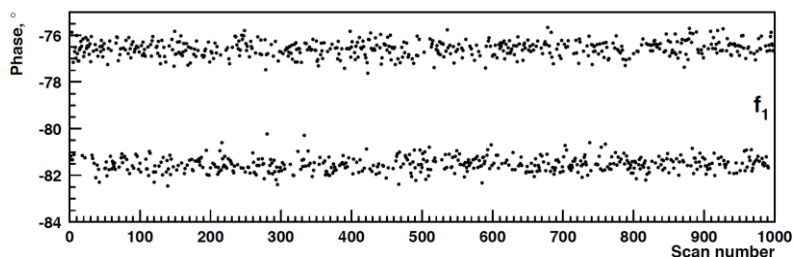


Figure S4. Frequency (top) and amplitude (middle) distributions obtained with LSF (solid lines) and FT (dashed lines) processing of 1,000 modeled transient signals with 5 spectral components corresponding to isotopic distribution of a doubly charged peptide substance P. Phase distributions (bottom) were obtained with LSF. The modeled transient signals were generated using experimental parameters (shown with red lines) derived from the LSF results for the ICR transients, Figure 3. Note, only a single initial phase value was used for the modeling.

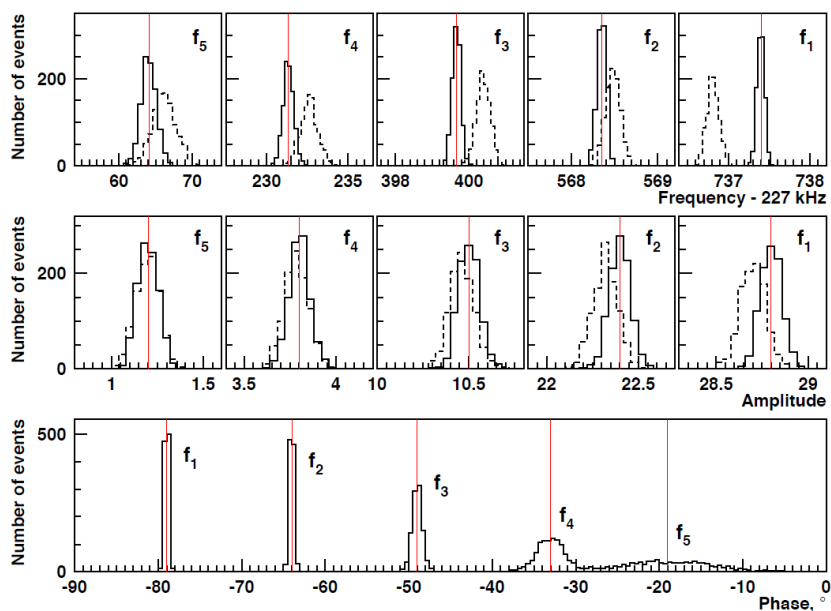


Figure S5. Frequency distributions obtained with LSF processing of 5-component ICR transient signals at different lengths: (top) 48 ms; (middle) 24 ms; and (bottom) 12 ms. The set of 1,000 transients was acquired for the isolated isotopic distribution of a doubly charged peptide substance P analyzed on the 10 T FT-ICR MS.

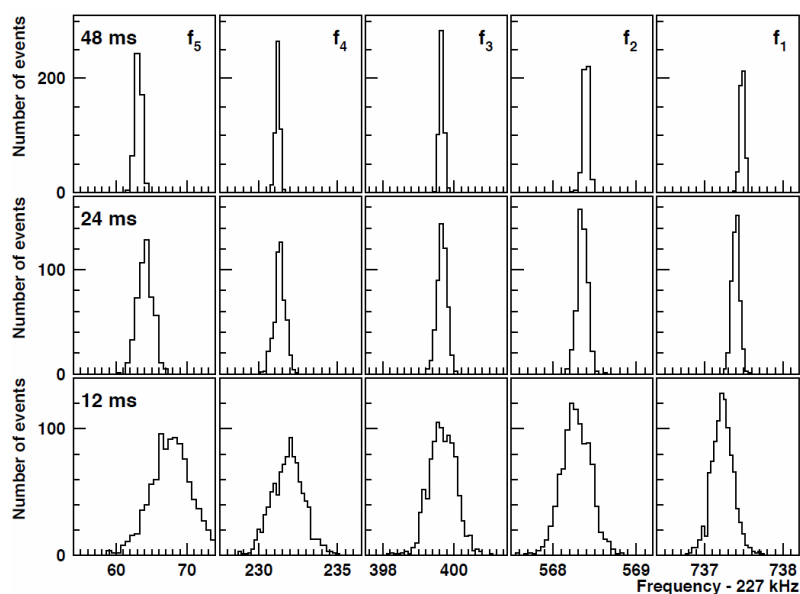


Table S1. Validation of the LSF method for a transient signal whose Fourier spectrum contains unresolved fine-structure clusters of isotopologues. The transient signal, with length $T = 24$ ms and sampling frequency $f_s = 2.7306(6)$ MHz, was generated as a sum of sinusoidal components with zero initial phases and values of amplitudes and frequencies corresponding to the theoretical abundances and m/z of all isotopologues from the 5 clusters of doubly charged peptide substance P. Ion frequencies were calculated for 10 T magnetic field, Figure 1 and Figure 3.

Cluster	Generated signal		Expected values		LSF results		
	Frequency, Hz	Amplitude, a.u.	Frequency, Hz	Amplitude, a.u.	Frequency, Hz	Amplitude, a.u.	Phase, °
0	227737.1654	100.00	227737.1654	100.00	227737.165 ±0.001	99.996 ±0.005	-0.003 ±0.005
1	227568.9374	6.58	227567.9613	77.15	227567.960 ±0.001	77.119 ±0.005	-0.001 ±0.007
	227568.5407	7.90E-01					
	227567.8719	68.14					
	227567.7265	4.95E-01					
	227567.3792	1.15					
2	227400.9578	2.04E-01	227399.2299	36.52	227399.228 ±0.003	36.460 ±0.005	0.002 ±0.014
	227400.6672	4.47					
	227400.5617	5.19E-02					
	227399.8938	4.48					
	227399.7487	3.26E-02					
	227399.4977	5.38E-01					
	227399.4019	7.56E-02					
	227399.3525	3.91E-03					

	227399.2445	2.67					
	227399.0058	9.08E-03					
	227398.8298	22.85					
	227398.6847	3.37E-01					
	227398.5395	1.13E-03					
	227398.3379	7.84E-01					
	227398.1928	5.70E-03					
	227397.8460	6.55E-03					
	227233.2260	3.98E-03					
	227232.9358	2.94E-01					
	227232.8304	1.61E-03					
	227232.1636	1.39E-01					
	227232.0186	1.01E-03					
	227231.8734	3.05					
	227231.7680	3.54E-02					
	227231.7284	2.22E-02					
	227231.6724	2.35E-03					
	227231.6231	2.57E-04					
	227231.5152	1.76E-01					
	227231.3822	5.15E-02					
	227231.2768	5.97E-04					
	227231.1197	2.11E-02					
	227231.1011	1.50					
	227230.9562	2.22E-02					
	227230.8113	7.44E-05					
	227230.7056	1.80E-01					
	227230.6100	5.15E-02					
3	227230.5607	2.66E-03	227230.7855	12.84	227230.790 ±0.008	12.800 ±0.005	0.009 ±0.041
	227230.4650	3.75E-04					
	227230.4528	1.82					
	227230.4157	8.94E-06					
	227230.3079	1.22E-02					
	227230.2144	6.19E-03					
	227230.1188	4.31E-04					
	227230.0695	4.50E-05					
	227230.0387	5.02					
	227229.9617	3.07E-02					
	227229.8938	1.13E-01					
	227229.7489	7.71E-04					
	227229.7233	5.17E-05					
	227229.6039	1.58E-06					
	227229.5476	2.63E-01					
	227229.4026	3.88E-03					
	227229.2577	1.30E-05					
	227229.0564	4.46E-03					
	227228.9115	3.24E-05					
	227228.5652	2.46E-05					
4	227065.7414	5.45E-05	227062.5716	3.60	227062.59 ±0.03	3.584 ±0.005	0.05 ±0.15
150	227065.4517	9.14E-03					
	227065.3464	3.14E-05					
	227064.6805	2.71E-03					
	227064.5882	1.05E-02					

227064.5358	1.97E-05				
227064.3908	2.00E-01				
227064.2856	1.10E-03				
227064.2461	1.46E-03				
227064.1901	4.58E-05				
227064.1409	7.98E-06				
227064.0332	5.46E-03				
227063.9003	3.38E-03				
227063.7951	1.85E-05				
227063.7434	1.20E-01				
227063.6382	1.39E-03				
227063.6197	4.67E-02				
227063.4750	6.89E-04				
227063.3302	2.31E-06				
227063.3300	1.02				
227063.2248	1.19E-02				
227063.1852	1.51E-02				
227063.1292	1.60E-03				
227063.0800	1.75E-04				
227063.0405	5.06E-05				
227062.9845	1.16E-05				
227062.9723	1.20E-01				
227062.9353	5.88E-07				
227062.8395	3.51E-02				
227062.8276	8.03E-04				
227062.7343	4.07E-04				
227062.6948	2.55E-04				
227062.6388	1.34E-05				
227062.5896	2.96E-06				
227062.5774	1.44E-02				
227062.5589	3.30E-01				
227062.4819	2.02E-03				
227062.4327	9.64E-05				
227062.4141	7.44E-03				
227062.3491	2.93E-04				
227062.3250	3.29E-02				
227062.2694	5.07E-05				
227062.2438	3.40E-06				
227062.1639	3.97E-02				
227062.1247	1.04E-07				
227062.0869	2.43E-04				
227062.0684	1.73E-02				
227062.0192	8.93E-04				
227061.9237	2.55E-04				
227061.9115	6.10E-01				
227061.8745	6.09E-06				
227061.7790	8.56E-07				
227061.7668	8.32E-03				
227061.7297	1.25E-08				
227061.6735	2.07E-03				
227061.6221	2.56E-05				
227061.5780	2.93E-04				

227061.5287	3.06E-05					
227061.4980	8.15E-01					
227061.4332	2.13E-06					
227061.4211	2.09E-02					
227061.3840	1.03E-07					
227061.3533	2.49E-02					
227061.2763	1.40E-04					
227061.2086	2.59E-04					
227061.1830	3.52E-05					
227061.0875	1.62E-06					
227061.0639	1.08E-06					
227061.0383	2.56E-07					
227061.0076	5.78E-02					
227060.9306	1.75E-04					
227060.9191	1.51E-09					
227060.8629	1.30E-03					
227060.7181	8.87E-06					
227060.6926	1.94E-07					
227060.5734	1.82E-08					
227060.5171	1.50E-03					
227060.3724	2.21E-05					
227060.2277	7.41E-08					
227060.0267	1.68E-05					
227059.8820	1.22E-07					
227059.5363	6.86E-08					

Table S2. Mean values and standard deviations of frequencies, amplitudes, and initial phases obtained with LSF and FT methods for the modeled transients with 5 spectral components corresponding to isotopic distribution of a doubly charged peptide substance P. The modeled transient signals were generated using experimental parameters derived from the ICR transients. The leftmost column lists the values used for transient signal modeling. Note, only a single initial phase value was used for the modeling. FT bias values are given relative to the FT standard deviations. N/A indicates that ion phases were calculated using LSF only.

Parameters	True value	LSF mean	LSF std	FT mean/bias	FT std
Amplitude 1	28.8	28.80	0.06	28.70/-1.4	0.07
Frequency 1, Hz	227737.4	227737.40	0.05	227736.81/-8.4	0.07
Phase 1, °	-79	-79.0	0.2	N/A	N/A
Amplitude 2	22.4	22.40	0.05	22.32/-1.1	0.07
Frequency 2, Hz	227568.35	227568.35	0.05	227568.48/1.4	0.09
Phase 2, °	-64	-64.0	0.3	N/A	N/A
Amplitude 3	10.5	10.50	0.05	10.46/-0.7	0.06
Frequency 3, Hz	227399.67	227399.66	0.11	227400.39/3.8	0.19
Phase 3, °	-49	-49.0	0.6	N/A	N/A
Amplitude 4	3.8	3.80	0.05	3.78/-3.3	0.06
Frequency 4, Hz	227231.32	227231.34	0.32	227232.63/2.4	0.54
Phase 4, °	-33	-33.1	1.5	N/A	N/A
Amplitude 5	1.2	1.20	0.05	1.20/0.0	0.06
Frequency 5, Hz	227064.1	227064.1	1.0	227066.2/1.3	1.6
Phase 5, °	-19	-19	5	N/A	N/A

Table S3. Mean values and standard deviations of frequencies, amplitudes, and initial phases obtained with LSF and FT methods for the modeled transients with 5 spectral components corresponding to isotopic distribution of a doubly charged peptide substance P. The modeled transient signals were generated using experimental parameters derived from the Orbitrap transients. The leftmost column lists the values used for transient signal modeling. FT bias values are given relative to the FT standard deviations. N/A indicates that ion phases were calculated using LSF only.

Parameters	True value	LSF mean	LSF std	FT mean/bias	FT std
Amplitude 1	68	68.00	0.02	67.93/-2.9	0.02
Frequency 1, Hz	468962.89	962.890	0.006	468962.68/-21.4	0.01
Phase 1, °	168	168.00	0.03	N/A	N/A
Amplitude 2	54	54.00	0.02	53.992/-0.3	0.024
Frequency 2, Hz	468788.82	788.820	0.009	468788.97/11.7	0.01
Phase 2, °	105	105.00	0.05	N/A	N/A
Amplitude 3	26	26.00	0.02	25.899/-4.1	0.025
Frequency 3, Hz	468614.89	614.891	0.016	468614.84/-1.4	0.03
Phase 3, °	44	44.00	0.08	N/A	N/A
Amplitude 4	9.1	9.1	0.02	9.12/0.8	0.02
Frequency 4, Hz	468440.92	440.92	0.05	468440.58/-4.4	0.08
Phase 4, °	-16	-16.00	0.25	N/A	N/A
Amplitude 5	1.6	1.6	0.02	1.60/0.2	0.02
Frequency 5, Hz	468265.97	265.97	0.27	468265.95/-0.05	0.43
Phase 5, °	-67	-67.0	1.4	N/A	N/A

A.5 Paper V

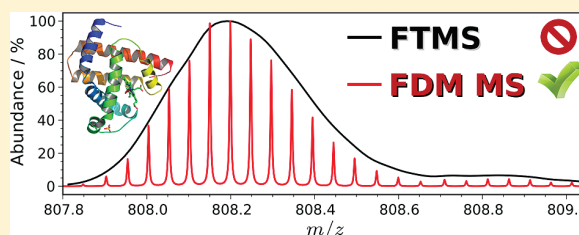
Filter Diagonalization Method-Based Mass Spectrometry for Molecular and Macromolecular Structure Analysis

Anton N. Kozhinov and Yury O. Tsybin*

Biomolecular Mass Spectrometry Laboratory, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

Supporting Information

ABSTRACT: Molecular and macromolecular structure analysis by high resolution and accurate mass spectrometry (MS) is indispensable for a number of fundamental and applied research areas, including health and energy domains. Comprehensive structure analysis of molecules and macromolecules present in the extremely complex samples and performed under time-constrained experimental conditions demands a substantial increase in the acquisition speed of high resolution MS data. We demonstrate here that signal processing based on the filter diagonalization method (FDM) provides the required resolution for shorter experimental transient signals in ion cyclotron resonance (ICR) MS compared to the Fourier transform (FT) processing. We thus present the development of a FDM-based MS (FDM MS) and demonstrate its implementation in ICR MS. The considered FDM MS applications are in bottom-up and top-down proteomics, metabolomics, and petroleomics.



High performance mass spectrometry (MS) provides structural information of organic, inorganic, and biological molecules by accurately measuring the masses of the ions of interest and their fragments.¹ Mass spectrometers with signal processing based on Fourier transform (FT) deliver superior analytical performance among the palette of MS instruments and are employed for the comprehensive structural analysis of molecules and macromolecules, both isolated and contained in the extremely complex mixtures.² However, to provide high resolution and high mass accuracy mass spectra, FT signal processing employed in the modern MS platforms requires long data acquisition time, typically on the order of hundreds of milliseconds. Thus, the advancement of the FTMS-based application areas including bottom-up,³ middle-down, and top-down proteomics,⁴ as well as metabolomics,⁵ petroleomics,⁶ and imaging,⁷ is severely hindered since these applications not only require high resolution but also set the very tight time constraints for the MS experiment. To overcome these limitations and advance molecular structure analysis, the high resolution MS data must be acquired faster.

The two main FT-based mass spectrometers today, ion cyclotron resonance (ICR) and Orbitrap, measure the masses by employing the relation between m/z ratios and abundances of ions with the frequencies and amplitudes of components of the time-domain transient signal.^{8–11} In general, the frequencies and amplitudes of the transient signal components can be obtained by either spectral or parameter estimation. The FT signal processing is arguably the most robust solution, albeit, being a spectral estimator, it has several drawbacks, including the resolution limitation. Super-resolution (SR) signal processing methods have been developed to overcome the FT limitations.^{12–14} The SR methods are either non-FT spectral estimators, e.g., the maximum entropy method and regularized

resolvent transform, or parameter estimators, e.g., linear prediction, the Prony method, and the filter diagonalization method (FDM).^{12–15} The FDM is a SR parameter estimator recently developed to be devoid of numerical limitations which other SR methods normally suffer from.¹⁵ Historically, the FDM was first introduced for quantum mechanics and further developed for nuclear magnetic resonance (NMR) spectroscopy.^{15–22} Currently, along with the standard FDM,^{18–20} there are several FDM-related methods, e.g., decimated signal diagonalization (DSD) and regularized resolvent transform (RRT).^{23,24} DSD-related methods were examined in FT-ICR MS earlier;^{25,26} however, they have an inferior performance compared to the FDM, primarily due to the fundamental limitations of the DSD-related methods. Recently, Aizikov and O'Connor applied the FDM to trace frequency modulation effects in experimental transient signals in FT-ICR MS and demonstrated super-resolution mass spectra obtained by FDM for simulated transients.^{27,28} Here, we introduce FDM-based mass spectrometry (FDM MS) to accelerate high resolution FT-ICR MS data acquisition and demonstrate its performance on the experimental data obtained for common classes of molecules and macromolecules.

EXPERIMENTAL SECTION

Sample Preparation. Cisplatin, substance P, and equine myoglobin were obtained from Sigma-Aldrich (Buchs, Switzerland). LC-MS grade water and acetonitrile were obtained from Fluka (Buchs, Switzerland). Formic acid was obtained from

Received: December 20, 2011

Accepted: February 27, 2012

Published: February 27, 2012

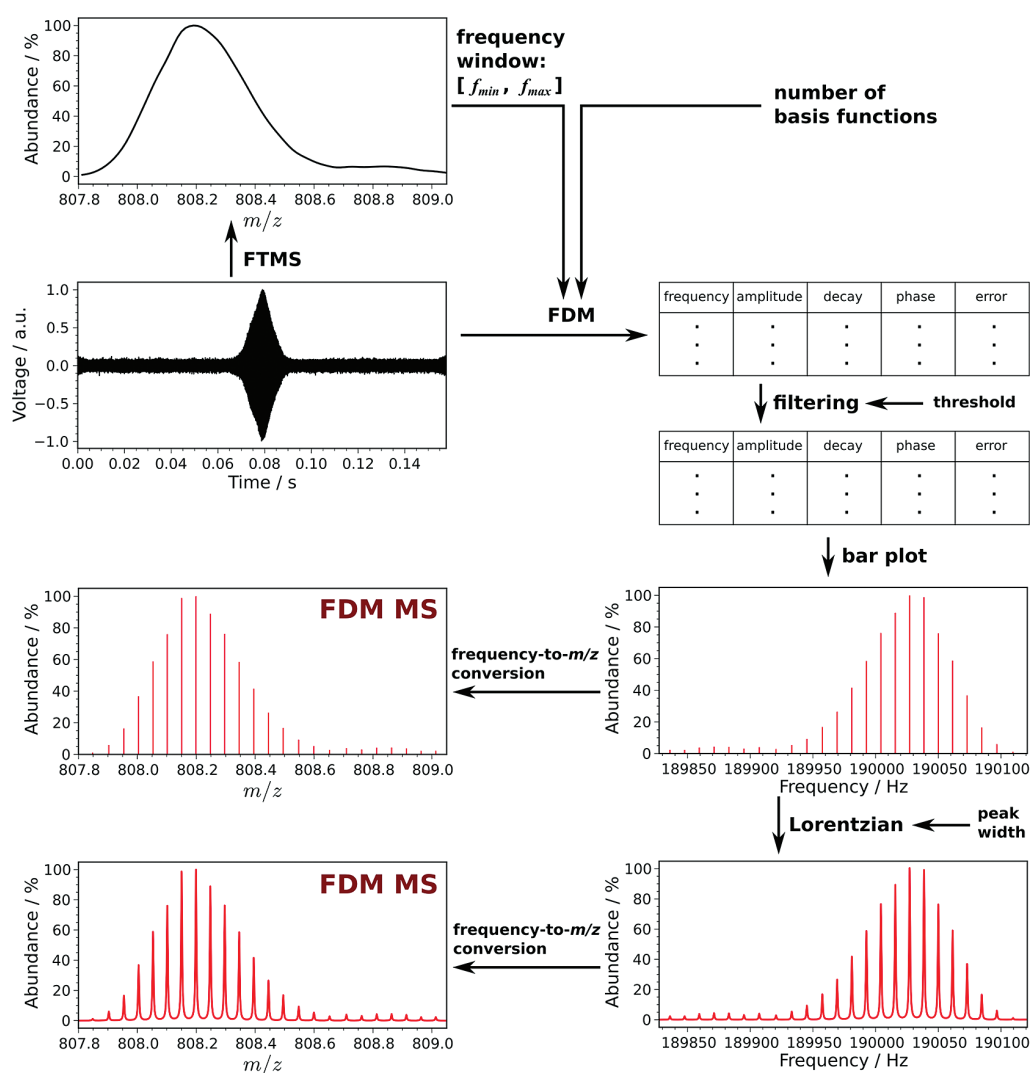


Figure 1. The main steps in the suggested here FDM-based mass spectrometry (FDM MS) methodology. The FDM processing is the standard FDM algorithm described elsewhere.^{15,18–20} Other steps are realized in the custom-written PyFTMS package.

Merck (Zug, Switzerland). Solution of hydrated cisplatin was prepared in water. Solutions of substance P and myoglobin 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 10 T electrospray ionization linear ion trap Fourier transform ion cyclotron resonance (ESI LTQ FT-ICR) mass spectrometer (Thermo Scientific, Bremen, Germany). The instrumental parameters and operation were controlled by standard data acquisition software. Acquisition of transient signals was performed in MIDAS data format. Transient signal of a European crude oil was acquired on a custom 9.4 T FT-ICR MS (NHMFL, FL, USA).⁶

Data Analysis. Data analysis was performed using the in-house developed Python-based software. The calculations were carried out on a standard desktop computer with a quad-core processor. The FTMS was performed following the standard algorithm.²⁹ First, the transient signal is submitted to the procedures of apodization with Hann's window and single zero-filling. Second, Fourier transformation is employed to convert the transient signal into the frequency domain spectrum with

conventional magnitude-mode spectral representation. Finally, calibration procedure based on the relation between frequency and m/z values is performed to provide the mass spectrum.³⁰ The FDM MS was performed as described in the main text. To find the monoisotopic mass of substance P from the experimental data, a standard deconvolution procedure was performed on the basis of the m/z of the monoisotopic peak and m/z spacing between the peaks in the mass spectra. To find the experimental monoisotopic mass of myoglobin, the isotopic ion distributions in the mass spectra were deconvoluted using the THRASH algorithm.³¹

RESULTS AND DISCUSSION

Methodology of Filter Diagonalization Method-Based Mass Spectrometry. The FDM-based MS methodology as employed here is described in Figure 1. Briefly, the harmonic inversion problems are solved for the selected frequency (mass-to-charge) window and chosen number of the basis functions following the standard FDM algorithm.¹⁵ The frequency window selection can be performed without (data-independent approach) or with (data-dependent approach) use of the

preliminary FTMS step performed to find the location of the frequency regions with peaks of interest. In the latter case, since FT processing does not necessarily resolve the peaks of interest, FTMS is used only to reveal the corresponding frequency regions for further analysis. The transient signal is subjected to the FDM calculations for the selected frequency window. The FDM calculations are carried out with a user-defined number of basis functions. The number of basis functions is a constraint of the implemented numerical algorithm of FDM. For the appropriate results, the number of basis functions should exceed the actual number of ion peaks to be resolved in a given m/z (frequency) window, whereas the upper limit of the number of basis functions is restricted by the available computational resources. The number of basis functions can be selected on the basis of the particular sample being analyzed, MS experimental conditions, or stability of the calculations performed for a set of different numbers of basis functions. In the calculations performed here, the number of basis functions was between 15 and 240 per a frequency window, vide infra. In addition, the control calculations for the sets of different numbers of basis functions were carried out to verify the calculations stability and to confirm the obtained results. The FDM processing provides the solutions as FDM table that contains information on frequency, amplitude, phase, decay, and roughly estimated complex frequency error as a figure of merit for each solution. In the following step, the possible spurious solutions (false positives) are filtered out on the basis of the threshold values specified for user-defined acceptable peak parameters. If the applied threshold is not sufficiently strong, the spurious peaks can remain in the spectra and, vice versa, if the threshold is too strong, the real ion peaks can be lost. Therefore, method optimization should be performed for the particular experimental conditions and type of samples employed. Further, on the basis of the frequencies and amplitudes from the FDM table, the bar plot in the frequency scale is created and then reconstructed into the spectrum with Lorentzian-mode spectral representation with user-defined peak width. Finally, the bar plot and the spectrum in the frequency scale are subjected to the standard frequency-to- m/z conversion.³⁰ Note, compared to the typical procedure in the FDM NMR,¹⁵ here, the spectral representation with a given peak width is performed for visual convenience only. The m/z ratios and abundances of interest are parameters to be found from the transient signal. Therefore, the reported m/z ratios and abundances are taken directly from the final bar plot in the m/z scale as the results of the considered parameter estimation problem. The methodology described in Figure 1 can be similarly applied to provide analysis of the broadband mass spectra, which would be considered as a number of adjacent windows.

Application of Filter Diagonalization Method-Based Mass Spectrometry to Isolated Compounds. The performance comparison between the FTMS and the FDM MS as a function of data acquisition time is shown in Figure 2. The presented examples correspond to the MS application areas of metabolomics (small molecule analysis),⁵ bottom-up proteomics (peptide analysis),³ and top-down proteomics (protein analysis).⁴ Figure S1, Supporting Information, provides the theoretical isotopic distributions of the selected compounds. The corresponding ICR transient signals for all compounds analyzed are shown in Figure S2, Supporting Information. The FTMS and FDM MS accuracies of the monoisotopic masses and m/z ratios, obtained with an external

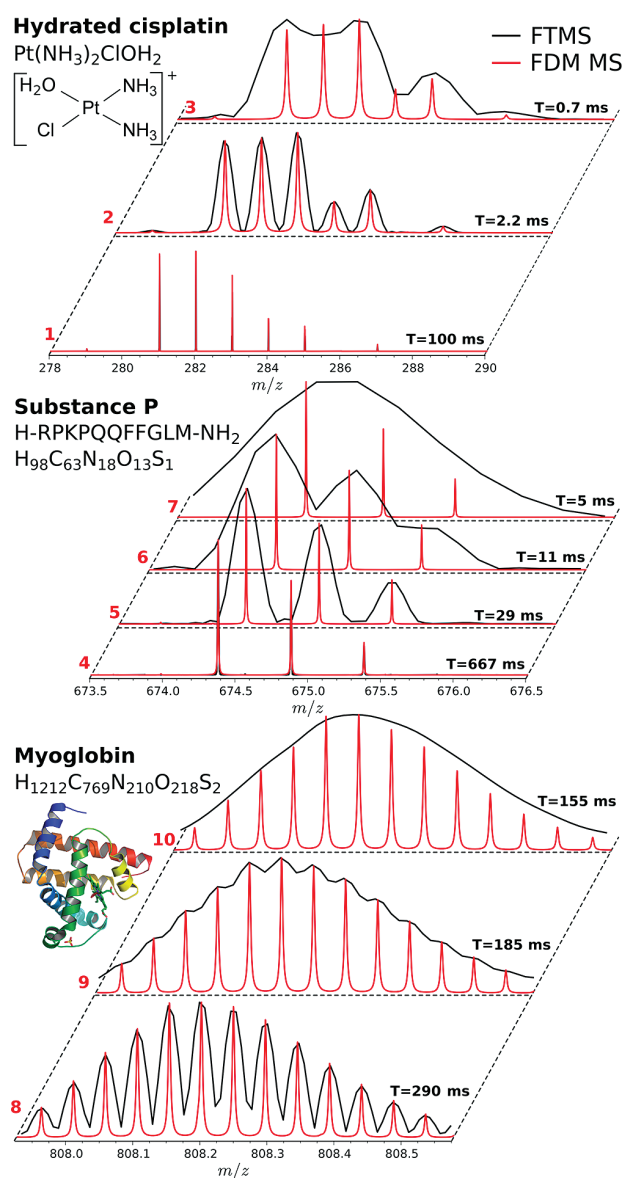


Figure 2. Filter diagonalization method (FDM) MS versus magnitude mode Fourier transform (FT)-based ICR MS for molecular structural analysis as a function of data acquisition time (transient length, T). (Top) analysis of a singly charged Pt-based compound; (Middle) analysis of an 11 amino acid long peptide, substance P, doubly charged; (Bottom) analysis of a 16.9 kDa protein, myoglobin, carrying 21 charges. The particular advantage of FDM MS application here is the potential ability to resolve the isotopic distributions of molecules and macromolecules in the experiments with (i) time constraints, e.g., in the online liquid chromatography-MS and (ii) finite lifetimes of transient signals, e.g., in a top-down proteomics.

calibration, are given in Tables S1–S3 (Supporting Information). In the structural analysis of a singly charged hydrated derivative of a typical metal-based anticancer therapeutic compound, cisplatin,³² FDM MS resolves the isotopic distribution even at an acquisition time of 0.7 ms, Figure 2 top. Here, the ion abundances and peak widths on the panel 1 ($T = 100$ ms) differ from those on the panels 2 ($T = 2.2$ ms) and 3 ($T = 0.7$ ms) since the isotopic fine structures are resolved only for $T = 100$ ms case (see Figure S3, Supporting

Information, for more details). For a given acquisition time, the FDM MS-provided m/z accuracies are normally superior to those obtained by FTMS for the same acquisition times, especially for the short ones.

In the structural analysis of a doubly protonated 11 amino acid long peptide substance P (H-RPKPQQFFGLM-NH₂), all FDM MS mass spectra shown, even the one for the 5 ms-long transient signal, reveal the isotopic ion distribution, in contrast to FTMS, Figure 2 middle. For all FDM MS mass spectra, the determination of the monoisotopic mass resulted in mass accuracies better than 1.3 ppm (external calibration).

Finally, all FDM MS mass spectra of a protein shown in Figure 2 bottom, even the one for a transient signal containing only one isotopic beat (Figure S2, Supporting Information), clearly reveal the isotopic distribution of myoglobin charged with 21 protons. In contrast, FTMS requires at least 4 isotopic beats of the transient signal to yield baseline resolution in the magnitude mode (Figure S4, Supporting Information). The deconvolution of the FDM MS mass spectra of myoglobin provided the correct protein charge state of 21 and the monoisotopic masses with the accuracies of 1.6 ppm (external calibration), whereas the FT-based mass spectrum from 3 or more isotopic beats transient was required. Comparable values of mass accuracies in both methods indicate that the mass calibration error prevailed over the errors of the signal processing methods and the deconvolution procedure. Thus, for all cases considered in Figure 2, compared to FTMS, FDM MS presents superior performance along with acceptable mass accuracy.

Application of Filter Diagonalization Method-Based Mass Spectrometry to a Complex Mixture of Compounds. The ultimate test for the analytical performance of a mass spectrometer and the associated method of data treatment involves the structural analysis of extremely complex mixtures of molecules as contained, for example, in crude oil.⁶ Figure 3 shows a comparative application of FTMS (top) and FDM MS

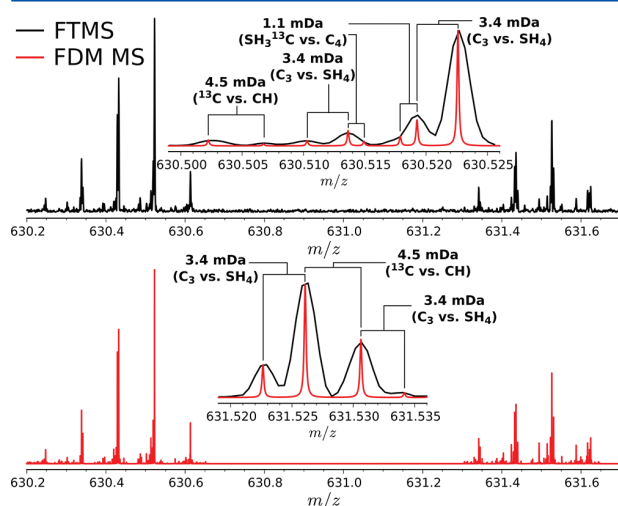


Figure 3. Filter diagonalization method (FDM) MS versus magnitude mode Fourier transform (FT) ICR MS for structural analysis of very complex mixtures of organic molecules, such as a European crude oil.⁶ Length of the transient signal employed was 2.86 s. The particular features of FDM MS here are the potential abilities to reveal low abundance ions and resolve closely spaced ion peaks that could not be separated by conventional FT-ICR MS at the same acquisition times.

(bottom) to a crude European oil sample. Importantly, FDM MS confirms the molecular ion peaks suggested by FTMS, including those with low abundances, Figure 3 insets. In addition, molecular ion peaks differing by only 1.1 mDa are resolved in the FDM MS mass spectrum. Thus, FDM MS confirms its applicability even in the case of high spectral complexity of analyzed transient signals.

Analytical Characteristics of Filter Diagonalization Method-Based Mass Spectrometry. Figure 4 shows the

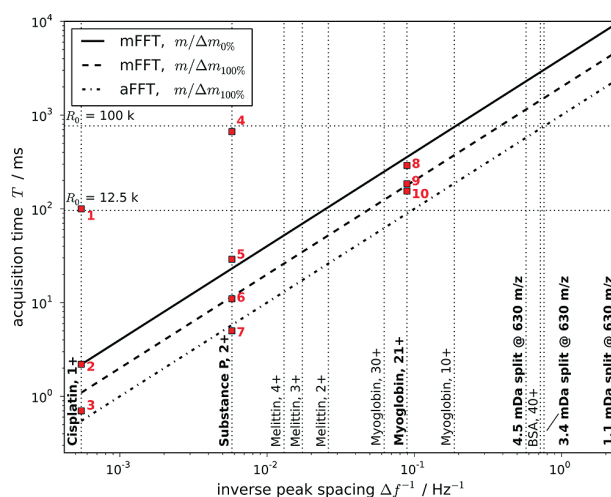


Figure 4. Comparison of filter diagonalization method (FDM) and fast Fourier transform (FFT)-based signal processing in ICR MS. The linear dependences for magnitude (mFFT, baseline resolution $m/\Delta m_{0\%}$ and crossover regime $m/\Delta m_{100\%}$) and absorption (aFFT, crossover regime $m/\Delta m_{100\%}$) modes were calculated for a 10 T magnetic field (see Figure S6, Supporting Information). The resolution levels $R_0 = 100k$ and $12.5k$ represent acquisition times typically employed in FTMS-based applications. The FDM MS results are given with the panel numbers that correspond to the numbers given to the mass spectra in Figure 2.

dependence of the required acquisition time of the transient signals on the desired resolution for molecules in a wide range of molecular weights and different charge states. The required acquisition time is determined by the frequency spacing between the ion peaks to be resolved (see Figure S5, Supporting Information, for further details). In FTMS, the required acquisition time depends on the smallest frequency spacing between the ion peaks in a given m/z ratio range to be resolved and on the desired degree of resolution (see Figure S6, Supporting Information, for further details). Up to twice shorter acquisition times may be required to yield a comparable resolution in FTMS when the absorption mode of spectral representation is used instead of the typically employed magnitude mode.^{33–36} Furthermore, detection of ICR frequency harmonics or multiples may reduce the required acquisition times in FTMS by several times.³⁷ Importantly, and in contrast to the FTMS, in the FDM MS, the minimum transient duration required to resolve the ion peaks of interest is determined by the average peak spacing in the considered m/z (frequency) narrow window (see Figure S5, Supporting Information, for further details). This time is shorter or even significantly shorter than the one required in FTMS, depending on the particular peak distribution within the considered m/z (frequency) spectrum window.^{17,18,38} The obtained FDM MS

Table 1. Parameters of the Employed FDM MS Methodology: Frequency Windows, Numbers of Basis Functions, and the Corresponding FDM and FDM MS Calculation Times

analyte	length of transient signal, ms	<i>m/z</i> window	FDM parameters		calculation time, s	
			frequency window, kHz	number of basis functions	FDM	FDM MS
cisplatin	0.7	278.0–290.0	529.59–552.45	15	<1 ^a	12
	2.2				<1	15
	100				<1	17
substance P	5	673.0–677.0	226.86–228.21	50	<1	13
	11				<1	15
	29				<1	22
	667				4	54
myoglobin	195	807.8–809.2	189.79–190.12	240	3	38
	185				5	47
	290				9	55
crude oil	2 860	630.20–630.27	228.848–228.873	200	392	920
		630.27–630.37	228.811–228.848	200		
		630.37–630.47	228.775–228.811	200		
		630.47–630.57	228.739–228.775	200		
		630.57–630.67	228.702–228.739	200		
		631.30–631.37	228.449–228.474	200		
		631.37–631.47	228.413–228.449	200		
		631.47–631.57	228.377–228.413	200		
631.57–631.67	228.340–228.377	200				

^aThe accuracy of time measurements was 1 s; calculation times below this value are shown as <1.

results, Figure 3, indicate the limitations in the provided computational resources and quality of the employed transient signals. Improved computational and experimental conditions should translate into higher FDM MS performance, e.g., even shorter transients required to resolve the peaks of interest with the appropriate mass accuracies, as in Figure 4, and better.

One of the anticipated limitations of FDM MS is a substantially higher demand for the computational resources required for FDM signal processing compared to FTMS. Among the non-FT methods, FDM in its original version scales quasi-linearly as $O(N \log N)$ for a signal of N sample points compared to other non-FT methods, which scale as $O(N^3)$ at best. Note, the FFT scales as $O(N \log N)$ compared to the straightforward discrete FT, DFT, which scales as $O(N^2)$. Thus, both FFT and FDM scale quasi-linearly with time-domain signal length. However, depending on the number of basis functions used for a certain window, the scaling prefactor for FDM can be larger than the one for FFT. To evaluate the performance of the FDM signal processing employed in the current work, the number of basis functions was varied for cisplatin between 15 and 40, for substance P between 45 and 70, for myoglobin between 235 and 260, and for crude oil sample between 175 and 300. For each analyte, the obtained mass spectra were consistent with each other providing similar frequency (mass) and amplitude (abundance) accuracies. The frequency windows and numbers of basis functions employed to achieve results in Figures 2 and 3 as well as the corresponding calculation times are summarized in Table 1. The calculation times are shown separately for the FDM calculations as well as for the complete FDM MS procedure, as described in Figure 1. The described above FDM computational speed as a function of a number of data points N refers to the FDM calculation time. The results in Table 1 show that in this work the shortest FDM calculation time was less than 1 s for the panels 1–3 and 5–7, Figure 2 when, respectively, 15 and 50 basis functions were employed, whereas the longest FDM calculation time was 9 s when 240 basis functions were

considered for the panel 8, Figure 2. In addition, when a broadband mass spectrum is to be processed, there is a set of frequency windows to calculate in the FDM-based approach. For example, the total time to calculate the set of 9 windows from Figure 3 was 392 s when 200 basis functions per each window were considered, Table 1. Thus, FDM is slower than FFT. There are several directions to speed up the FDM calculations, including (i) calculation of a certain window using parallel algorithm instead of the original one and (ii) calculation of a set of windows simultaneously using parallel computing. Currently, we are working on the numerically effective version of the algorithm based on both approaches.

Another FDM MS limitation of numerical nature is the reduction of mass and abundance accuracies or resolution performance in case of low signal-to-noise ratio transient signals. This limitation requires not only software but also FTMS hardware improvements to be made. Nevertheless, a recently achieved substantial progress in FTMS instrumentation accompanied by further FDM algorithm characterization and development for weak signals support the implementation of FDM MS.^{39–42}

Other anticipated FDM MS limitations include effects that currently reduce FTMS performance as well and are due to incoherence of ion motion, which develops during signal acquisition: space charge effect in ICR cell, in particular, ion motion perturbation immediately after ion excitation; effect of inhomogeneity or instability of the magnetic field in ICR MS; effect of inharmonicity of the electric field in ICR and Orbitrap MS. Nevertheless, currently, space charge effect is under study and can be controlled;^{28,43} magnetic field drift or significant inhomogeneity are not typical for the state-of-the-art ICR magnets, and finally, novel ICR cells with harmonized electric field enter the routine use.^{44–46} In addition, the effect of signal frequency (and amplitude) modulation should be mentioned.²⁷ The modulation would lead to a set of peaks in the frequency (mass) spectrum compared to a single peak that would be given by a constant frequency. In the current work, such sets of peaks

were the isotopic distributions, Figure 2, and isotopic fine structures, Figure S3 in the Supporting Information. Interestingly, it was FDM that allowed studying the modulation effects in the experimental FT-ICR MS transient signals using frequency tracing calculations.^{27,28}

Finally, it should be noted that, by definition, the term “super-resolution” does not mean “ultrahigh resolution” and should not be considered as such. It is more appropriate to consider super-resolution methods-based MS as a technique to resolve the peaks of interest employing shorter acquisition times than the ones required for the FTMS. The ability of FDM MS to deliver ultrahigh resolution in its standard definition as employed in FT-ICR MS is still to be shown for the real experimental data.

CONCLUSIONS

We introduce FDM-based MS, which has a potential to significantly improve the structure analysis of (macro-) molecules and their complex mixtures by substantially decreasing measurement times in FTMS. Surprisingly, and despite the encouraging preliminary results by Aizikov and O'Connor, up to now, there has been no significant effort dedicated to the implementation of FDM MS for routine mass measurements. Compared to 1D FDM NMR, the benefits of FDM MS application to 1D ICR MS are more pronounced. Despite the described above limitations of the current FDM MS implementation, using readily available powerful workstations combined with recent advances in FTMS hardware supports the implementation of FDM MS for routine molecular structure analysis in the near future. The expected FDM MS-delivered improvements should be both qualitative, e.g., higher throughput, higher dynamic range of sample concentrations, lower sample consumption, and quantitative, e.g., more accurate liquid chromatography (LC) peak representation by MS data, especially important in label-free quantitative proteomics and metabolomics. Complementary molecular structure information, e.g., isotopic fine structures,¹ could be provided by FDM MS on a LC time scale (see Figure S3 in Supporting Information). To maximize these envisioned FDM MS benefits, an improved time management of FTMS experimental sequences is required. Importantly, the close analogy between ICR and Orbitrap FTMS ion detection principles allows one to expect a straightforward application of the FDM MS methodology to the Orbitrap FTMS data. Moreover, specifically, Orbitrap FDM MS-based bottom-up and top-down proteomics are expected to significantly advance molecular and macromolecular structure analysis. Finally, FDM MS may benefit other, currently FT-based and mainly research-grade, techniques, e.g., charge detection time-of-flight MS.⁴⁷

ASSOCIATED CONTENT

Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Address: EPFL ISIC LSMB, BCH 4307, 1015 Lausanne, Switzerland. Phone: +41216939751. E-mail: yury.tsybin@epfl.ch.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Sasa Miladinovic, Unige Laskay, Luca Fornelli, Aleksey Vorobyev, Konstantin Zhurov, Ryan P. Rodgers, Gregory T. Blakney, Alan G. Marshall, Paul Dyson, Oleg Tsybin, and Geoffrey Bodenhausen for the discussions and technical support. The work was supported by the Swiss National Science Foundation (Projects 200021-125147/1 and 128357) and the European Research Council (ERC Starting Grant 280271).

REFERENCES

- (1) Marshall, A. G.; Hendrickson, C. L.; Shi, S. D.-H. *Anal. Chem.* **2002**, *74*, 252A–259A.
- (2) Marshall, A. G.; Hendrickson, C. L. *Annu. Rev. Anal. Chem.* **2008**, *1*, 579–599.
- (3) Cox, J.; Mann, M. *Annu. Rev. Biochem.* **2011**, *80*, 273–299.
- (4) Kelleher, N. L. *Anal. Chem.* **2004**, *76*, 196a–203a.
- (5) Lei, Z. T.; Huhman, D. V.; Sumner, L. W. *J. Biol. Chem.* **2011**, *286*, 25435–25442.
- (6) Marshall, A. G.; Rodgers, R. P. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18090–18095.
- (7) van Hove, E. R. A.; Smith, D. F.; Heeren, R. M. A. *J. Chromatogr., A* **2010**, *1217*, 3946–3954.
- (8) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1–35.
- (9) Amster, I. J. *J. Mass Spectrom.* **1996**, *31*, 1325–1337.
- (10) Makarov, A.; Denisov, E.; Lange, O. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1391–1396.
- (11) Scigelova, M.; Hornshaw, M.; Giannakopoulos, A.; Makarov, A. *Mol. Cell. Proteomics* **2011**, *10*, M111 009431.
- (12) Roy, R.; Sumpter, B. G.; Pfeffer, G. A.; Gray, S. K.; Noid, D. W. *Phys. Rep.* **1991**, *205*, 109–152.
- (13) Marple, S. R., Jr. *Digital Spectral Analysis with Applications*; Prentice-Hall: Englewood Cliffs, N.J., 1987.
- (14) Kay, S. M. *Modern Spectral Estimation: Theory and Application*; Prentice Hall: Englewood Cliffs, N.J., 1988.
- (15) Mandelshtam, V. A. *Prog. Nucl. Magn. Reson. Spectrosc.* **2001**, *38*, 159–196.
- (16) Neuhauser, D. *J. Chem. Phys.* **1990**, *93*, 2611–2616.
- (17) Wall, M. R.; Neuhauser, D. *J. Chem. Phys.* **1995**, *102*, 8011–8022.
- (18) Mandelshtam, V. A.; Taylor, H. S. *J. Chem. Phys.* **1997**, *107*, 6756–6769.
- (19) Mandelshtam, V. A.; Taylor, H. S. *J. Chem. Phys.* **1998**, *109*, 4128–4128.
- (20) Mandelshtam, V. A.; Taylor, H. S. *J. Chem. Phys.* **1997**, *106*, 5085–5090.
- (21) Hu, H. T.; Van, Q. N.; Mandelshtam, V. A.; Shaka, A. J. *J. Magn. Reson.* **1998**, *134*, 76–87.
- (22) Pang, J. W.; Dieckmann, T.; Feigon, J.; Neuhauser, D. *J. Chem. Phys.* **1998**, *108*, 8360–8368.
- (23) Belkic, D.; Dando, P. A.; Taylor, H. S.; Main, J. *Chem. Phys. Lett.* **1999**, *315*, 135–139.
- (24) Chen, J. H.; Shaka, A. J.; Mandelshtam, V. A. *J. Magn. Reson.* **2000**, *147*, 129–137.
- (25) Belkic, D.; Dando, P. A.; Main, J.; Taylor, H. S.; Shin, S. K. *J. Phys. Chem. A* **2000**, *104*, 11677–11684.
- (26) Belkic, D.; Dando, P. A.; Main, J.; Taylor, H. S. *J. Chem. Phys.* **2000**, *113*, 6542–6556.
- (27) Aizikov, K.; O'Connor, P. B. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 836–843.
- (28) Aizikov, K.; Mathur, R.; O'Connor, P. B. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 247–256.
- (29) Marshall, A. G.; Verdun, F. R. *Fourier Transforms in NMR, Optical, and Mass Spectrometry: a User's Handbook*; Elsevier: Amsterdam, 1990.

- (30) Shi, S. D.-H.; Drader, J. J.; Freitas, M. A.; Hendrickson, C. L.; Marshall, A. G. *Int. J. Mass Spectrom.* **2000**, *195*, 591–598.
- (31) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 320–332.
- (32) Hartinger, C. G.; Tsybin, Y. O.; Fuchser, J.; Dyson, P. J. *Inorg. Chem.* **2008**, *47*, 17–19.
- (33) Beu, S. C.; Blakney, G. T.; Quinn, J. P.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2004**, *76*, 5756–5761.
- (34) Xian, F.; Hendrickson, C. L.; Blakney, G. T.; Beu, S. C.; Marshall, A. G. *Anal. Chem.* **2010**, *82*, 8807–8812.
- (35) Qi, Y. L.; Thompson, C. J.; Van Orden, S. L.; O'Connor, P. B. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 138–147.
- (36) Qi, Y.; Barrow, M. P.; Van Orden, S. L.; Thompson, C. J.; Li, H.; Perez-Hurtado, P.; O'Connor, P. B. *Anal. Chem.* **2011**, *83*, 8477–8483.
- (37) Vorobyev, A.; Gorshkov, M. V.; Tsybin, Y. O. *Int. J. Mass Spectrom.* **2011**, *306*, 227–231.
- (38) Barinova, G.; Nyman, G. *Chem. Phys.* **2002**, *281*, 23–31.
- (39) Kaiser, N. K.; Quinn, J. P.; Blakney, G. T.; Hendrickson, C. L.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1343–1351.
- (40) Michalski, A.; Damoc, E.; Lange, O.; Denisov, E.; Nolting, D.; Mueller, M.; Viner, R.; Schwartz, J.; Remes, P.; Belford, M.; Dunyach, J.-J.; Cox, J.; Horning, S.; Mann, M.; Makarov, A. *Mol. Cell. Proteomics* **2011**, DOI: <http://dx.doi.org/10.1074/mcp.O111.013698>.
- (41) Benko, U.; Juricic, D. *Signal Process.* **2008**, *88*, 1733–1746.
- (42) Celik, H.; Shaka, A. J. *J. Magn. Reson.* **2010**, *207*, 17–23.
- (43) Vladimirov, G.; Hendrickson, C. L.; Blakney, G. T.; Marshall, A. G.; Heeren, R. M.; Nikolaev, E. N. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 375–384.
- (44) Nikolaev, E. N.; Boldin, I. A.; Jertz, R.; Baykut, G. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1125–1133.
- (45) Kaiser, N. K.; Savory, J. J.; McKenna, A. M.; Quinn, J. P.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2011**, *83*, 6907–6910.
- (46) Tolmachev, A. V.; Robinson, E. W.; Wu, S.; Smith, R. D.; Pasachonis, L. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1334–1342.
- (47) Doussineau, T.; Bao, C. Y.; Clavier, C.; Dagany, X.; Kerleroux, M.; Antoine, R.; Dugourd, P. *Rev. Sci. Instrum.* **2011**, *82*, 084104.

Supplementary Information

Filter diagonalization method-based mass spectrometry for molecular and macromolecular structure analysis

Anton N. Kozhinov and Yury O. Tsybin

Biomolecular Mass Spectrometry Laboratory, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.

Figure S1. Theoretical isotopic distributions of the analyzed compounds. (Top left) isotopic distribution of hydrated cisplatin at resolution $R = 2.7k$ (for 10 T magnitude mode FTMS equivalent acquisition time $T = 0.01$ s); (Bottom left) isotopic distribution of hydrated cisplatin at $R = 270k$ ($T = 1.0$ s), isotopic fine structures are resolved; (Top right) isotopic distribution of doubly protonated substance P at $R = 22k$ ($T = 0.192$ s); (Bottom right) isotopic distribution of 21 times protonated myoglobin at $R = 36k$ ($T = 0.384$ s). The theoretical isotopic distributions were calculated for the Hann frequency response function.

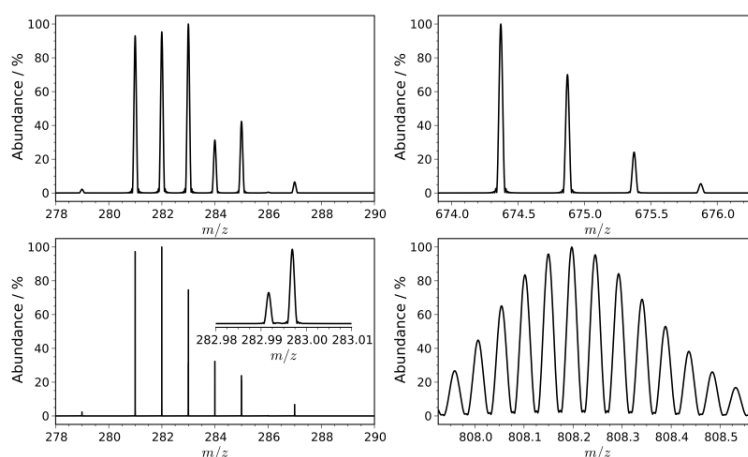


Figure S2. Experimental ICR MS transient time-domain signals of the analyzed compounds for different acquisition times. (Left) transient signals of hydrated cisplatin; (Center) transient signals of doubly protonated substance P; (Right) transient signals of 21 times protonated myoglobin. To reduce the detrimental effect of the transient signal perturbation immediately after ion excitation in the employed ICR cell, an appropriate time delay (10 ms-100 ms) was introduced between ion excitation and detection for all cases represented.

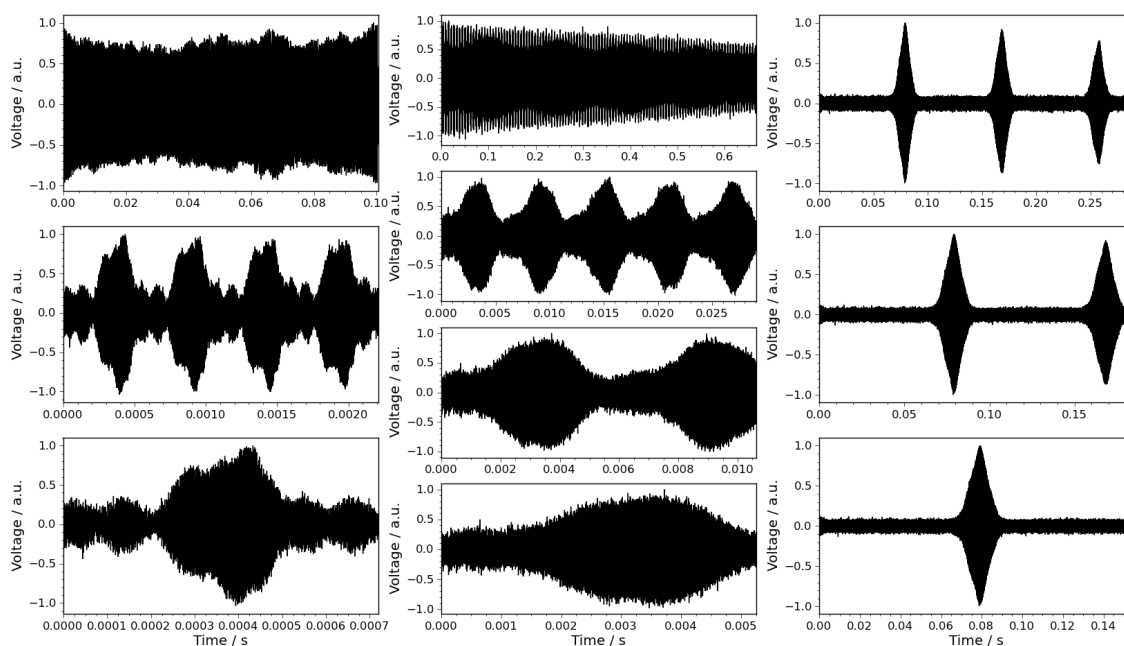


Figure S3. FDM MS versus FTMS in the analysis of hydrated cisplatin. Zoom-in to the isotopic fine structures (doublets) for 100 ms long transient signal. (Left) Isotopic fine structure at nominal m/z 283; (Right) isotopic fine structure at nominal m/z 285. Isotopic fine structure explains both the abundance and peak widths differences between panel 1 and panels 2 and 3 in Figure 2 (main text). According to the FDM MS methodology, the Lorentzian peak shape is applied to each bar from the bar spectrum in the frequency scale. When isotopic fine structure is unresolved (too short transient duration for a given peak density), Lorentzian peak shape is made broad. In case isotopic fine structure is resolved, Lorentzian peak shape is made narrow to visually reveal the peaks. Note, as peak shape is applied specifically for the visual representation of the MS data, a narrower peak shape can be applied to the data in panel 1, Figure 2.

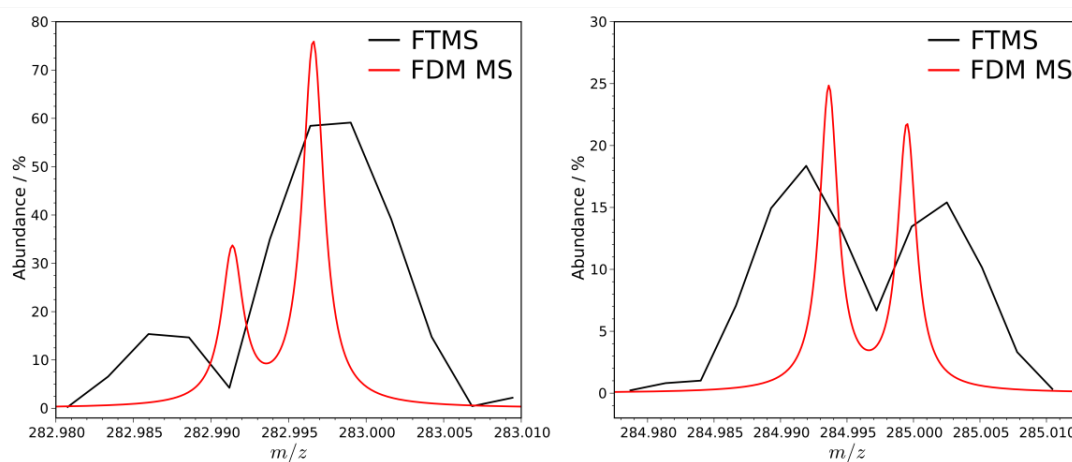


Figure S4. Baseline resolution of the isotopic distribution of 21 times protonated myoglobin in FT ICR-MS. (Top) experimental transient time-domain signal with 4 isotopic beats; (Bottom) baseline resolved mass spectrum of the transient shown in the top panel.

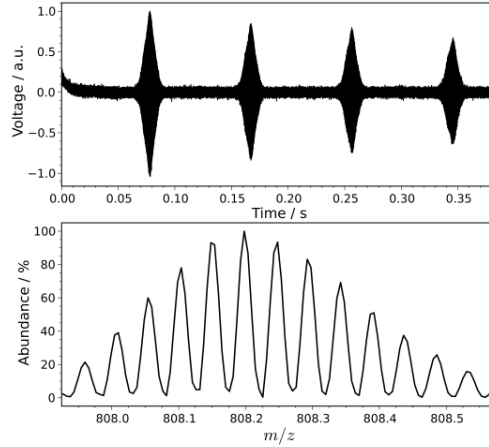
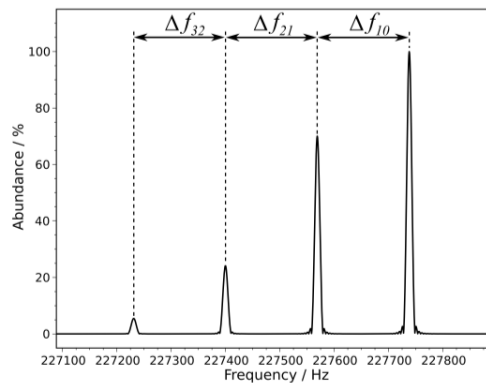


Figure S5. Peak spacing in frequency spectra. The definitions for "peak spacing" and "average peak spacing" are given below.

Peak spacing. An example for isotopic distribution of doubly protonated substance P is shown. According to the standard two parameter calibration equation, $m/z = \frac{A}{f} + \frac{B}{f^2}$, for given ions with m_i/z and m_j/z ratios with close mass-to-charge ratios ($\Delta m_{ij} = |m_i - m_j| \ll m_i, m_j$), the peak spacing Δf_{ij} is related with the corresponding mass difference Δm_{ij} in the following way: $\Delta f_{ij} = A \frac{\Delta m_{ij}/z}{[m_j/z]^2}$.

For isotopic distributions $\Delta m_{ij} \approx 1$ so that for any i^{th} and j^{th} successive ions $\Delta f_{ij} \approx A \frac{z}{m_j^2}$ and

furthermore $\Delta f_{ij} \approx \Delta f_{k,k+1} \approx A \frac{z}{m_k^2}$, where m_k is the mass of any ion from the isotopic distribution, e.g. the monoisotopic one.



Note, (i) higher ion charge state for the same ion mass increases peak frequency spacing and thus decreases the inverse peak spacing, and (ii) higher ion mass for the same ion charge state reduces peak frequency spacing and thus increases the inverse peak spacing, as reflected in Figure 4.

Average peak spacing. For given frequency window ΔF and number of peaks N , the average peak density is $\langle \rho \rangle = \frac{N}{\Delta F}$ so that the average peak spacing is $\langle \Delta f \rangle = \langle \rho \rangle^{-1} = \frac{\Delta F}{N}$.

Figure S6. Definition of resolution in FTMS. (Top) General definition of resolution with the valley at a specified fraction of the peak height for 2 functions, F1 and F2; (Middle) Resolution with the valley at 0% of the peak height, the peaks are baseline resolved; (Bottom) Resolution with the valley at 100% of the peak height, crossover regime (the peaks are not resolved).

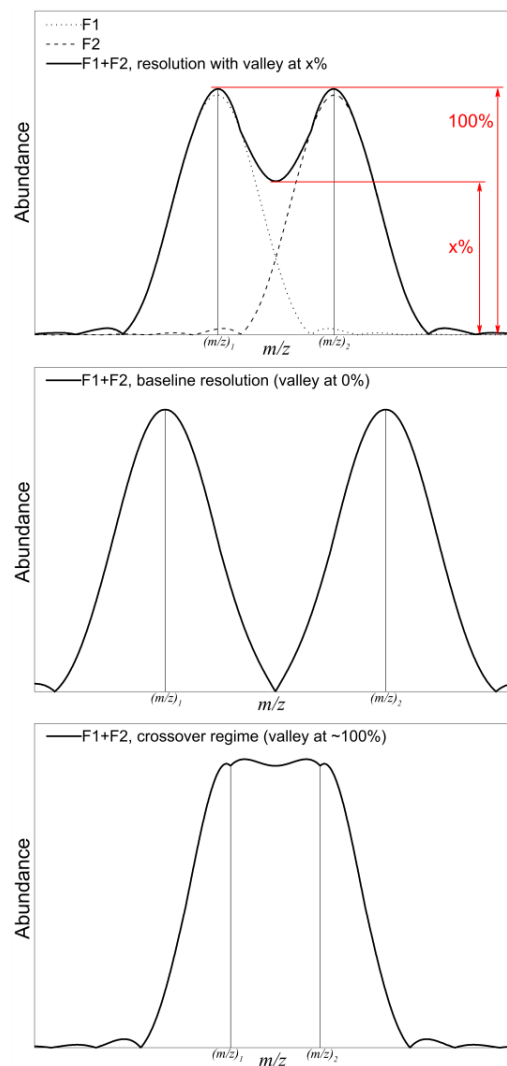


Table S1. Accuracies of the obtained m/z ratios of the most abundant ions in the analysis of hydrated cisplatin by ICR-MS. Results are shown for both FTMS and FDM MS. External calibration of the mass scale was employed. Experiments were performed with standard sample preparation, electrospray ionization, and ICR-MS conditions. The corresponding theoretical isotopic distributions and transient time-domain signals are shown in Figures S1 and S2. The shown accuracies were obtained using the estimations of the true m/z ratios from the simulation of the theoretical isotopic distribution of hydrated cisplatin with resolved isotopic fine structures (doublets) at 283 and 285 nominal m/z ($R = 270k$, see Figure S1, left and Figure S3). For the acquisition time of 100 ms, all abundant peaks were considered since the isotopic fine structures were resolved. For the acquisition times of 2.2 ms and 0.7 ms, the peaks with doublets were not considered due to ambiguity of the interpretation of their m/z accuracy. N/A indicates unresolved peaks.

Acquisition time, ms	Peaks (nominal m/z)	m/z accuracy, ppm	
		FTMS	FDM MS
100	281	1.92	-1.38
	282	-2.20	-1.07
	283 (doublet)	-20.16	-1.05
		7.48	-0.94
	284	-1.73	-1.07
285 (doublet)	-7.07	-1.12	
	9.40	-0.98	
2.2	281	-129.86	-1.80
	282	59.45	-1.48
	284	123.53	-0.65
0.7	281	0.68	1.13
	282	N/A	1.69
	284	N/A	-3.25

Table S2. Results of the deconvolution procedure applied in the analysis of doubly protonated 11 amino acid long peptide substance P. N/A indicates the deconvolution failing due to unresolved peaks.

Acquisition time, ms	Technique	Charge state	Monoisotopic peak, m/z	Monoisotopic mass, Da	Mass accuracy, ppm
667	FTMS	2	674.37176	1346.7289	0.66
	FDM MS	2	674.37218	1346.7298	1.28
29	FTMS	2	674.37995	1346.7453	12.82
	FDM MS	2	674.37171	1346.7288	0.59
11	FTMS	2	674.36618	1346.7178	-7.61
	FDM MS	2	674.37113	1346.7277	-0.26
5	FTMS	N/A	N/A	N/A	N/A
	FDM MS	2	674.37164	1346.7292	0.89

Table S3. Results of the deconvolution procedure applied in the analysis of 21 times protonated protein myoglobin. N/A indicates the deconvolution failing due to unresolved peaks.

Acquisition time, ms	Technique	Charge state	Monoisotopic mass, Da	Mass accuracy, ppm
290	FTMS	21	16940.99	1.62
	FDM MS	21	16940.99	1.62
185	FTMS	N/A	N/A	N/A
	FDM MS	21	16940.99	1.62
155	FTMS	N/A	N/A	N/A
	FDM MS	21	16940.99	1.62

A.6 Paper VI

On the Utility of Isotopic Fine Structure Mass Spectrometry in Protein Identification

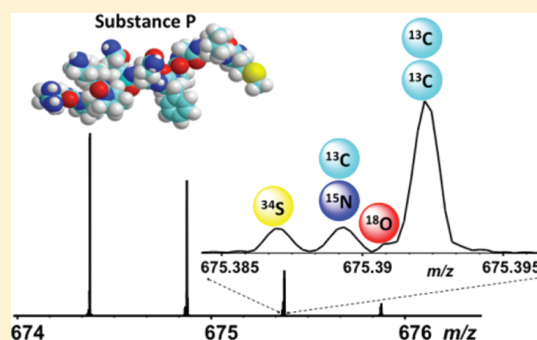
Saša M. Miladinović,[†] Anton N. Kozhinov,[†] Mikhail V. Gorshkov,[‡] and Yury O. Tsybin^{*†}

[†]Biomolecular Mass Spectrometry Laboratory, Ecole Polytechnique Federale de Lausanne, 1015 Lausanne, Switzerland

[‡]Institute for Energy Problems of Chemical Physics, Russian Academy of Sciences, 119334 Moscow, Russian Federation

Supporting Information

ABSTRACT: Modern mass spectrometry (MS)-based protein identification and characterization relies upon accurate mass measurements of the ^{13}C isotopic distributions of the enzymatically produced peptides. Interestingly, obtaining peptide elemental composition information from its isotopic fine structure mass spectrum to increase the confidence in peptide and protein identification has not yet been developed into a bottom-up proteomics-grade analytical approach. Here, we discuss the possible utility and limitations of the isotopic fine structure MS for peptide and protein identification. First, we *in silico* identify the peptides from the *E. coli* tryptic digest and show the increased confidence in peptide identification by consideration of the isotopic fine structures of these peptides as a function of mass and abundance accuracies. In the following, we demonstrate that the state-of-the-art high magnetic field Fourier transform ion cyclotron resonance (FT-ICR) MS allows a routine acquisition of the isotopic fine structure information of a number of isobaric peptide pairs, including a pair of peptides originating from *E. coli*. Finally, we address the practical limitation of the isotopic fine structure MS implementation in the time-constrained experiments by applying an advanced signal processing technique, filter diagonalization method, to the experimental transients to overcome the resolution barrier set by the typically applied Fourier transformation. We thus demonstrate that the isotopic fine structures of peptides may indeed improve the peptide and possibly protein identification, can be produced in a routine experiment by the state-of-the-art high resolution mass spectrometers, and can be potentially obtained on a chromatographic time-scale of a typical bottom-up proteomics experiment. The latter one requires at least an order of magnitude increase in sensitivity of ion detection, which presumably can be realized using high-field Orbitrap FTMS and/or future generation of ultrahigh magnetic field FT-ICR MS equipped with harmonized ICR cells.



Mass spectrometry (MS) is one of the most efficient and versatile analytical tools for the comprehensive analysis of the samples of biological origin.^{1,2} Within the palette of mass spectrometric methods utilized for such an analysis, Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) is a significant contributor to the biological research because of its high mass measurement accuracy, high resolving power, and unique tandem mass spectrometric (MS/MS) capabilities.^{3–5} Since the introduction in 1974,⁶ this type of mass spectrometry passed through a number of breaking developments⁷ demonstrating unmatched resolving power until today. Examples include resolving of the $^3\text{He}^+ / ^3\text{H}^+$ isobaric ions at the peak resolution of 10 million and mass measurement accuracy of ~ 1 ppb⁸ and 200 million peak resolution for $^{40}\text{Ar}^+$ ions and $^{132}\text{Xe}^+$ ions,⁹ as well as 8 million for a protein.¹⁰ In general, resolving powers in an excess of 1 million are routinely attainable for multiply charged peptides and proteins using high magnetic field FT-ICR MS technology. A comparable level of resolving power on other types of mass spectrometers became available recently with the development of high electrostatic field Orbitrap FTMS technology.^{11,12} Specifically, Makarov and

co-workers recently demonstrated resolution of up to 1 million for peptide ions on a chromatographic time scale of 3 s on the Orbitrap FTMS with the advanced signal processing, including phase correction, that corresponds to performance of modern high magnetic field FT-ICR mass spectrometers.¹³

The use of ^{13}C isotopic patterns produced by low resolution MS for improved molecular analysis has been suggested previously for light, <500 Da, molecules.¹⁴ The achieved level of resolving power in FTMS thus provides the opportunity to have a deeper look at the isotopic pattern of detected ion elemental composition in search of a more reliable way of peptide identification. Indeed, Shi et al. and Miura et al. have shown that the correct number of elements, such as S, O, and N, typically present in the elemental compositions of organic molecules, including peptides, can be determined if the resolving power of a mass spectrometer is sufficient enough for obtaining their relative isotopic abundances.^{10,15} These high

Received: December 24, 2011

Accepted: April 2, 2012

Published: April 2, 2012

resolution mass spectra are referred to as the "isotopic fine structures" to distinguish them from the peptide ^{13}C isotopic patterns which are attainable for most of the peptides with the majority of mass analyzers used in proteomics. However, even highly accurate isotopic fine structures cannot be used to distinguish between the peptide isomers. Therefore, information on the relative abundances of the isotopic fine structures should be considered as a source of information complementary to other sources, e.g., accurate mass, fragmentation pattern in MS/MS, retention time in liquid chromatography (LC), etc.

The commercially available FTMS platforms, ICR and Orbitrap, have already demonstrated the feasibility to obtain the isotopic fine structures of natural peptides and small proteins.^{10,13,15–20} Knowing the number of chemical elements in a peptide may significantly reduce the number of possible elemental compositions assigned to an ion with a given mass accuracy.^{21–23} However, despite the advantages of having additional information about the elemental compositions of the peptides, the isotopic fine structure has not been routinely employed in the existing protein identification strategies. There are several factors that limit the routine implementation of isotopic fine structure mass spectrometry. First, the resolving power required is typically in an excess of 1 million for most of the peptides, such as tryptic peptides considered in bottom-up proteomics. The required level of resolving power can be obtained by FTMS instruments but necessarily includes ion trapping in a spatially restricted area of an ion trap. The latter results in a strong influence of Coulombic interactions between the trapped ions on both the relative positions of the peaks in a mass spectrum originating from the ^{15}N , ^{34}S , and ^{18}O elements and the relative abundances of these peaks in the recorded fine structures. Second, to achieve the desired resolving power in the mass spectra, the FT-based ion trap technologies require extremely long experimental times, mostly due to the necessity of acquiring the time-domain signals in the 10 to 30 s range.¹⁸ The above-mentioned limitations dictate the need to develop methods of ion frequency harmonics or multiples detection,²⁴ robust absorption-mode implementations,^{25–28} and novel data processing tools, presumably non-FT based ones, aimed to overcome the resolution limitations of the FT algorithm.

Peptide identification from the *E. coli* tryptic digest has been previously analyzed *in silico* based on accurate mass measurements and isoelectric point calculations.²⁹ Here, we extend this early work and consider, through the *in silico* analysis of tryptic peptides generated from *E. coli* protein database, the utility of isotopic fine structure content in peptide mass spectra for improving the reliability of protein identification based on the bottom-up proteomic strategy. This utility is further investigated with the experimental examples in which we show the feasibility to obtain the isotopic fine structures for a number of pairs of natural and synthetic peptides with close m/z ratios by FT-ICR MS. Importantly, to overcome the experimental time limitations coming from the impractically slow scan rate in these high resolution data acquisition sequences, we considered the possibility to process the time-domain transients with the filter diagonalization method (FDM) instead of the FT-based one.³⁰ The FDM method was recently introduced for FT-ICR mass spectrometry by O'Connor and co-workers for frequency tracing along the time-domain transient signals.^{31,32} Furthermore, O'Connor successfully demonstrated the possibility of employing FDM for obtaining isotopic fine structures from significantly shorter, yet simulated, time-domain signals

compared with a common FT-based approach.³¹ The first application of FDM-based mass spectrometry to the experimental data, termed FDM MS, has recently been shown by our group to resolve ^{13}C isotopic structures of small molecules, peptides, and proteins using shorter transient time-domain signals compared to the FT-based signal processing.³³

EXPERIMENTAL SECTION

Samples and Sample Preparation. Standard peptides, bradykinin ($\text{C}_{50}\text{H}_{73}\text{N}_{15}\text{O}_{11}$, monoisotopic mass 1059.5614 Da) and substance P ($\text{C}_{63}\text{H}_{98}\text{N}_{18}\text{O}_{13}\text{S}$, monoisotopic mass 1346.7281 Da), were purchased from Sigma-Aldrich (Buchs, Switzerland) and used without further purification. Peptide H-VGPPGFSPFVG-OH ($\text{C}_{52}\text{H}_{73}\text{N}_{11}\text{O}_{13}$, bradykinin isobar, monoisotopic mass 1059.5389 Da), isobaric pair H-RVMRGMR-OH ($\text{C}_{35}\text{H}_{68}\text{N}_{16}\text{O}_8\text{S}_2$, monoisotopic mass 904.4847 Da) and H-RSHRGHR-OH ($\text{C}_{35}\text{H}_{60}\text{N}_{20}\text{O}_9$, monoisotopic mass 904.4852 Da), and two isobaric peptides from *E. coli* tryptic digest, annotated in the manuscript as *E. coli* peptide 1 (H-ITNHHDHATGDIQTIGHHFR-OH, $\text{C}_{99}\text{H}_{147}\text{N}_{35}\text{O}_{30}$, monoisotopic mass 2306.1053 Da) and *E. coli* peptide 2 (H-KPIWENQSCDTSNLMVLNSK-OH, $\text{C}_{98}\text{H}_{159}\text{N}_{27}\text{O}_{33}\text{S}_2$, monoisotopic mass 2306.1035 Da), were synthesized in-house by solid-state Fmoc chemistry (Peptide and Protein Synthesis Facility, University of Lausanne, Switzerland). HPLC-MS grade water and acetonitrile were obtained from Fluka (Buchs, Switzerland). Formic acid was obtained from Merck (Zug, Switzerland). Peptides were dissolved in water and acetonitrile (50/50, v/v) with addition of 0.5% of formic acid to make 1–5 μM solutions.

Mass Spectrometry. Peptides were ionized using conventional microelectrospray ionization (ESI) ion source. Mass spectrometric measurements were performed using a hybrid 10 T linear ion trap Fourier transform ion cyclotron resonance mass spectrometer (LTQ FT-ICR MS, Thermo Scientific, Bremen, Germany), described elsewhere.³⁴ The experimental parameters and sequences were controlled by Xcalibur software (Thermo Scientific). Ions in question were first isolated in the LTQ with an isolation window of 5 m/z and then transferred to the capacitively coupled cylindrical ICR cell (Ultra cell, Thermo Scientific) followed by gated trapping at 3 V potentials applied to the trapping electrodes of the ICR cell. The trapping potentials were kept at 3 V during ion excitation event and then reduced to 0.4 V for ion detection.³⁴ The number of accumulated ions was controlled with the automatic gain control (AGC) function, which was set to 5×10^5 charges in the LTQ, unless stated otherwise. An ion relaxation period in the ICR cell of up to 1 s was followed by broadband frequency-sweep excitation to a post-excitation radius corresponding to approximately half of the cell's radius. Final time-domain (transient) signals were obtained by summation of 100 single transients of about 12 s long each. Unless stated otherwise, the MIDAS software package (National High Magnetic Field Laboratory, FL, USA) was employed to process the resulted transient signals in a conventional way with a single zero-filling, Hanning apodization and fast Fourier transform (FFT) to produce magnitude-mode frequency spectrum and convert it into the final mass spectrum using external calibration.³⁵

The mass and abundance measurement accuracies were calculated based on the standard approaches.^{34,36} Briefly, the mass or m/z accuracy (error, in ppm) and the abundance accuracy (error, in %) for a given ion peak were obtained according to the following expressions respectively 171

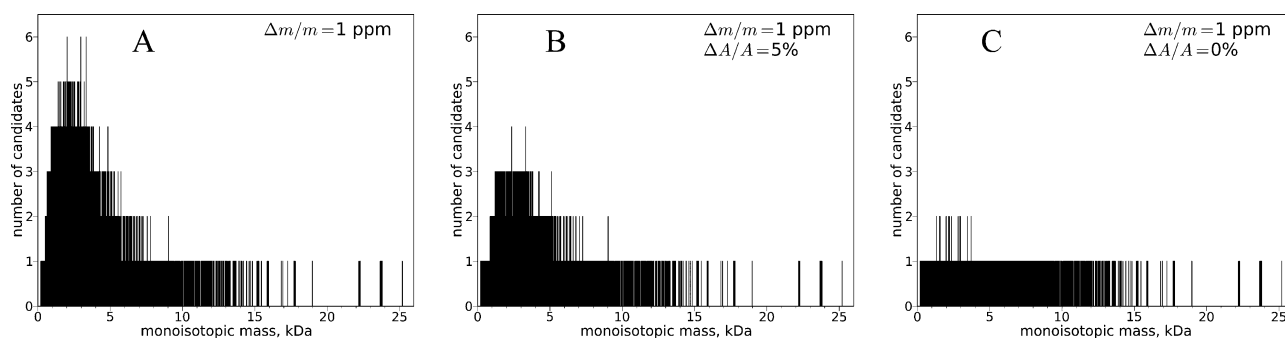


Figure 1. Number of possible candidates for the peptide elemental composition as a function of the peptide monoisotopic mass in identification of *E. coli* tryptic peptides based on the measurement of the monoisotopic masses with 1 ppm mass accuracy alone (A) and with the additional information of the relative abundance of the ^{15}N ion peaks compared to the ^{13}C ones with the 5% (B) and 0% (C) abundance accuracy.

$$\frac{\Delta(m/z)}{m/z} = \frac{(m/z)_{\text{exp}} - (m/z)_{\text{th}}}{(m/z)_{\text{th}}}, \quad \frac{\Delta A}{A} = \frac{A_{\text{exp}} - A_{\text{th}}}{A_{\text{th}}}$$

where $(m/z)_{\text{exp}}$ is the experimental mass-to-charge ratio of the considered peak, A_{exp} is the experimental relative abundance of the peak (compared to either ^{13}C peak for A+1 isotopic fine structure or $^{13}\text{C}_2$ peak for A+2 isotopic structure), and $(m/z)_{\text{th}}$ and A_{th} represent the corresponding theoretical values calculated using masses and natural abundances of isotopes (National Institute of Standards and Technology, 2010).³⁷ Errors, that could be referred to as total errors, of experimental m/z and abundance values include systematic and random errors. In the current work summation of 100 transient signals for each mass spectrum with isotopic fine structure was employed and resulted in random errors of relative abundances to be typically below the systematic ones in the control experiments with bradykinin and its isobaric peptide (see Results and Discussion section) so that the total errors are mainly systematic. See the Supporting Information for details on random error determination procedure. Note, for ^{13}C isotopic distributions the total error of relative abundance measurement was reported to be about 1% in case of summation of 100–200 transients for the employed FT-ICR mass spectrometer.³⁴

Method Development. The in-house developed Python-based library, PyFTMS, was employed to perform *in silico* digestion procedure of the *E. coli* proteins (based on the UniProt repository); select peptides from the obtained digest for given mass and abundance measurement accuracies; calculate theoretical isotopic fine structures for selected peptides; and carry out peptide identification based on matching theoretical and experimental data. The isotopic distribution is hereafter described with the monoisotopic peak denoted as "A", the first isotopic cluster corresponding to a single ^{13}C isotope as "A+1", the second one corresponding to 2 atoms of ^{13}C present as "A+2", etc. Examining the experimental isotopic fine structures could potentially be used to estimate the numbers of atoms of a certain element in the peptide elemental composition. For example, the relative number of nitrogen atoms compared to the number of carbon ones corresponds to the abundance of the ^{15}N peak compared to the ^{13}C one in the A+1 isotopic fine structure cluster.¹⁰

Finally, PyFTMS was used to obtain mass spectra of A+1 isotopic fine structure of selected ions for different lengths of the experimental transient signal using signal processing based on Fourier transform and filter diagonalization method (FDM).

The former was performed in a conventional way. The latter followed a typical FDM routine³⁰ and FDM MS methodology described elsewhere.³³ Briefly, with a chosen number of basis functions as a constraint of the algorithm, FDM calculations were performed for a selected frequency (mass-to-charge) window to find a number of parameters, including frequencies and amplitudes, of signal components from ions of interest. Note, the number of basis functions should exceed the number of peaks of interest in the considered window.³⁰ In the current work, the number of basis functions was selected based on the expected A+1 isotopic fine structure, albeit in general it can be chosen based on other parameters, e.g., stability of the calculations performed for a set of different numbers of basis functions if there is no information on the analyzed sample and experimental conditions. Based on the frequencies and amplitudes, a Lorentzian shape spectrum in the frequency scale is created with user-defined peak width and converted into a mass spectrum using the standard external calibration. Profiling a frequency spectrum is employed only for visual convenience, whereas the reported m/z ratios and abundances are obtained directly from the frequencies and amplitudes using the standard external calibration.³³

RESULTS AND DISCUSSION

Analysis of Peptide Isotopic Fine Structure Utility in Proteomics. The *in silico* identification of *E. coli*-derived tryptic peptides was employed here to evaluate the utility of peptide isotopic fine structure information in the large-scale proteomic studies. Peptide identification based on the measurement of monoisotopic mass and relative abundance of ^{15}N peak compared to ^{13}C one from the A+1 isotopic fine structure was simulated. To characterize the performance of the technique depending on the experimentally provided mass and abundance measurement accuracies, the identification is considered for the sets of mass and abundance measurement accuracies, 0.1, 1, 5, 10 ppm and 0, 1, 2, 5, 10, 20, 50%, respectively. For a given mass accuracy, the potential of the technique was characterized by the abundance accuracy of 0% when the experimental errors in abundance measurements were absent. The distributions shown in Figure 1 and Figure S1 provide a number of possible candidates (vertical axis) for the peptide elemental composition as a function of the monoisotopic mass (horizontal axis) of the actual peptide subjected to the identification for considered mass and abundance measurement accuracies. In particular, Figure 1 shows a result of *in silico* peptide identification for mass accuracy of 1 ppm based on measurement of

Table 1. Analysis of *E. coli* Tryptic Digest Enzymatic Peptide Identification Based on Isotopic Fine Structure As a Function of Mass and Abundance Accuracy^a

abundance accuracy, %	Number of Elemental Compositions Distributed into Mass Bins with <i>N</i> Candidates											
	10 ppm mass accuracy						5 ppm mass accuracy					
	<i>N</i> = 1	<i>N</i> = 2	<i>N</i> = 3	3 < <i>N</i> ≤ 6	6 < <i>N</i> ≤ 14	14 < <i>N</i> ≤ 22	<i>N</i> = 1	<i>N</i> = 2	<i>N</i> = 3	3 < <i>N</i> ≤ 6	6 < <i>N</i> ≤ 14	14 < <i>N</i> ≤ 22
NA	15647	13942	13349	38973	67689	3232	29972	26177	25314	56119	15250	0
50	16217	14704	13769	39667	65493	2982	30766	26894	25776	54999	14397	0
20	23732	21155	18894	45061	43463	527	44104	38878	26814	37923	5113	0
10	42530	34466	24336	40639	10860	1	63434	46776	24878	17238	506	0
5	78900	36462	19742	16868	860	0	101043	38194	11081	2511	3	0
2	116362	27608	7164	1698	0	0	135839	15916	1039	38	0	0
1	133135	17421	2066	210	0	0	148174	4589	68	1	0	0
0	142488	9902	441	1	0	0	152542	288	2	0	0	0
abundance accuracy, %	1 ppm mass accuracy						0.1 ppm mass accuracy					
	<i>N</i> = 1	<i>N</i> = 2	<i>N</i> = 3	3 < <i>N</i> ≤ 6	6 < <i>N</i> ≤ 14	14 < <i>N</i> ≤ 22	<i>N</i> = 1	<i>N</i> = 2	<i>N</i> = 3	3 < <i>N</i> ≤ 6	6 < <i>N</i> ≤ 14	14 < <i>N</i> ≤ 22
	NA	103835	39508	8378	1111	0	0	151358	1468	6	0	0
50	105487	38418	7898	1029	0	0	151366	1460	6	0	0	0
20	121257	27366	3791	418	0	0	152321	511	0	0	0	0
10	137828	14458	526	20	0	0	152757	75	0	0	0	0
5	139991	12569	269	3	0	0	152806	26	0	0	0	0
2	143885	8855	92	0	0	0	152814	18	0	0	0	0
1	150708	2110	14	0	0	0	152818	14	0	0	0	0
0	152804	28	0	0	0	0	152832	0	0	0	0	0

^aA complete mass range of 1-25 kDa, comprising all *in silico* produced enzymatic peptides (152832 elemental compositions), is considered.

monoisotopic mass only (Figure 1 A) and with relative abundances of ¹⁵N to ¹³C peaks measured with abundance accuracy, $\Delta A/A$, of 5% (Figure 1 B) and 0% as the upper limit case (Figure 1 C). Table 1 and Table S1 provide distributions of the total numbers of the elemental compositions over the mass bins in Figure 1 and Figure S1 with different numbers of candidates inside for the considered mass and abundance measurement accuracies. The total number of possible elemental compositions considered for a mass range of 0 to 25 kDa (all tryptic peptides are included) was 152832 (Table 1) and for a mass range of 1 to 7 kDa (truncated distribution of tryptic peptides) was 132289 (Table S1). Therefore, more than 86% of all elemental compositions of tryptic peptides from *E. coli* were within 1 to 7 kDa mass range.²⁹

For both mass regions, the mass measurement accuracy of 1 ppm was not sufficient for the unambiguous *E. coli* peptide elemental composition identification based upon the peptide mass measurement alone.²⁹ Depending on the monoisotopic mass of the peptide, the number of possible candidates varies between 1 and 6. We expect these numbers to be significantly higher for the peptide databases of more complex organisms. The ambiguity of the identification varies as a function of both mass and abundance accuracies and reduces as the abundance accuracy improves for a given mass accuracy, as expected. The presented results allow estimation of the required abundance accuracy threshold for a desired false discovery rate. Finally, all peptide elemental compositions could be uniquely identified only for the mass accuracy of 0.1 ppm and abundance accuracy of 0%, among the presented cases for both mass ranges so that all elemental compositions go to the different bins. This analysis could be further generalized by taking into account relative abundances of other isotopic ions, such as sulfur and oxygen.

The analytical utility of peptide isotopic fine structure in proteomics could be compared to the utility of peptide isotopic

structure that considers the abundance of the unresolved ¹³C isotopic peaks. Fiehn and co-workers demonstrated that relative abundance of isotopic peaks of organic molecules of up to 500 Da acquired on a low-resolution MS can improve molecular identification by providing information on their elemental composition.¹⁴ However, proteolytic peptides considered in the current work possess more complicated isotopic structure. Application of low-resolution MS to quantify the ratio of abundances of monoisotopic and ¹³C-containing isotopic peaks will result in substantially higher error than for the small molecules. Furthermore, to quantitatively compare the utility of isotopic fine structure and ¹³C isotopic structure in peptide identification, we complemented the results presented above by the *in silico* identification of *E. coli*-derived tryptic peptides based on the measurement of the relative abundance of ¹³C peak from the A+1 isotopic cluster compared to the monoisotopic peak, see Figure S2 and Tables S2 and S3. The methodological and computational details are similar to the description of the data presented above for isotopic fine structure calculations. The results demonstrate the overall superiority of isotopic fine structure information to the ¹³C isotope information in peptide elemental composition assignment. Therefore, high resolution MS that resolves peptide isotopic peaks to the level of isotopic fine structure shall (i) improve the performance of the ¹³C isotope-based molecular structure analysis by reducing the abundance errors and (ii) provide complementary information to further reduce the number of possible peptide candidates.

Isotopic Fine Structures of Bradykinin and Its Isobaric Peptide. FT-ICR mass spectra of the two isobaric peptides, bradykinin and H-VGPPGFSPFVG-OH, were obtained with the resolving power of 1,500,000 at *m/z* 1061, Figure 2. Mass difference between the monoisotopic masses (denoted as "A" peaks) of these singly protonated (or neutral) peptides is only 22.47 mDa, Figure 2 A. The obtained resolving power was

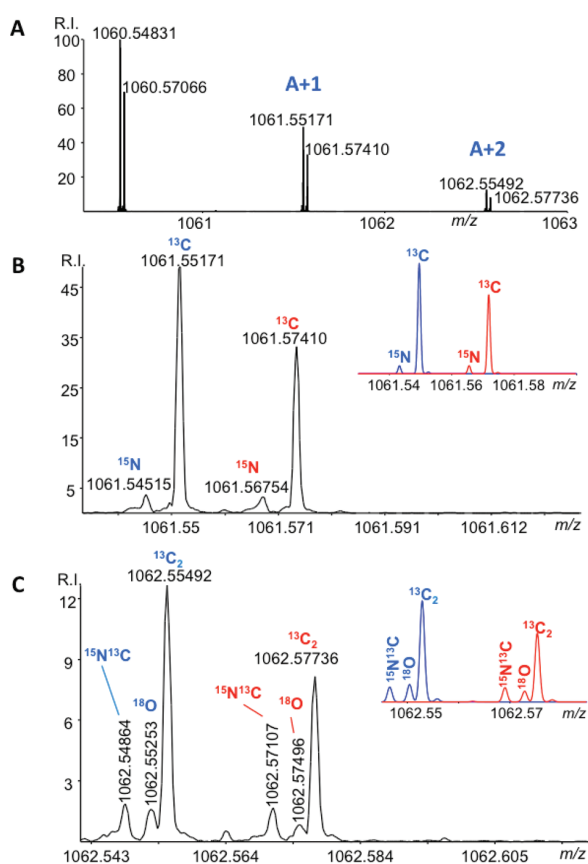


Figure 2. ESI FT-ICR MS of bradykinin and H-VGPPGFSPFVG-OH peptide pair showing (A) the ^{13}C isotopic envelope of the singly charged peptides and the expanded segments of this mass spectrum demonstrating (B) the isotopic fine structure of the A+1 isotopic envelope and (C) the isotopic fine structure of the A+2 isotopic envelope. Ion signals corresponding to bradykinin are shown in red (the heavier peptide) and to the H-VGPPGFSPFVG-OH peptide in blue (the lighter peptide). Abundance accuracies of the experimental values versus the theoretical ones are shown in Table 2, whereas the insets represent the theoretical isotopic fine structure distributions.

sufficient to resolve the isotopic fine structures corresponding to the natural abundances of ^{13}C , ^{15}N , and ^{18}O elements in the A+1 and A+2 isotopic envelopes of these peptides. The A+1 isotopic envelopes of singly protonated peptides reveal the presence of both ^{13}C and ^{15}N isotopic peaks, Figure 2 B, whereas the A+2 isotopic envelopes are composed of ^{13}C , $^{15}\text{N}^{13}\text{C}$, and ^{18}O isotopes, Figure 2 C. The theoretical isotopic fine structures are shown in the insets in Figures 2 B and 2 C, respectively. Theoretically, 15 nitrogen atoms in bradykinin should give the ^{15}N isotopic peak relative abundance of 10.13% compared to the neighboring ^{13}C isotopic peak. At the same time, the synthetic peptide from this pair contains 11 nitrogen atoms providing the ^{15}N isotopic peak relative abundance of 7.15%, Figure 2 B inset. Accordingly, the ratio of $^{15}\text{N}^{13}\text{C}$ to $^{13}\text{C}_2$ ion abundances in these isotopic fine structure envelopes equals to 20.68% for bradykinin ion and 14.6% for the synthetic one, Figure 2 C inset. The ^{18}O signal is present in A+2 isotopic envelope and theoretical abundances versus the corresponding $^{13}\text{C}_2$ are 15.8% and 17.2% for bradykinin and synthetic peptide, respectively.

The accuracies of experimental relative abundances for the peaks in the A+1 and A+2 isotopic fine structure envelopes of bradykinin and its isobaric peptide are summarized in Table 2 (see the Experimental Section and the Supporting Information for more details). The effects that influence the ion signal abundances in FTMS have been described previously.^{21,38–41} In most cases, the $\Delta A/A$ values are relatively small and are in the range of 0.2%–5%, Table 2. However, for the ^{18}O isotopic peaks the abundance errors are 26% and 34% for the synthetic peptide and bradykinin, respectively. The main contribution in such big errors is most probably due to spectral interference⁴² since the ^{18}O isotopic peaks are resolved insufficiently, Figure 2 C. The random errors of relative abundances in a single FT-ICR MS measurement comprising of 100 summed single time-domain transient signals were estimated based on the spectral signal-to-noise ratio (SNR), see the Supporting Information for more details. As expected, for high abundance ^{15}N isotopic peaks the random errors were below 1%, Table 2. For low abundance ^{18}O isotopic peaks, the random errors were higher, up to 3%, since the signal-to-noise ratios of ^{18}O isotopic peaks are lower than the ones for the ^{15}N isotopic peaks.

The obtained accuracies of experimental abundances of ^{15}N isotopic peaks in the considered experiments (Table 2) show the instrumental performance in obtaining accurate abundances of the isotopic peaks in peptide identification based on A+1 isotopic fine structure, Table 1 and Table S1. From the obtained results it follows that FT-ICR MS is able to provide reasonable abundance accuracy of several percents once both systematic and random errors are paid attention to, so that the total error, which includes both systematic and random errors, is kept below a desired level. In particular, (i) at least baseline resolution is required to reduce systematic errors due to a detrimental effect of spectral interference and (ii) appropriate signal-to-noise ratios are needed to keep the random errors below the systematic errors once the latter ones are reduced down to a certain level.

Isotopic Fine Structures of *E. coli* Isobaric Tryptic Peptides. To evaluate the utility of the isotopic fine structure information in identification of peptides from complex mixtures containing isobaric peptides with masses within the 1 ppm window the *E. coli* tryptic digest was selected as a model system.⁴³ Figure 3 shows the high resolution FT-ICR MS-based analysis of the isotopic fine structures for a peptide pair for the triply charged peptides 1 and 2 from the *E. coli* tryptic digest (see the Experimental part for peptide sequences). The monoisotopic mass difference between these peptides in the neutral or singly charged state is 1.81 mDa. The m/z values of the monoisotopic peaks of the selected triply charged *E. coli* tryptic peptides, reported in Figure 3, differ only by 0.62 mTh, or 0.8 ppm in the relative m/z values, and thus represent a challenging case for the mass spectrometric identification. The high-resolution mass spectrometers are capable of resolving the corresponding ion signals, but mass measurement accuracy better than 0.8 ppm should be achieved to identify these two peptides based solely on the m/z values. The high magnetic field FT-ICR MS is able to resolve the isotopic fine structures for both peptides in this pair with the resolving power of 1.5 million, as shown in the expanded segments of the experimental mass spectrum in Figures 3 B and 3 C. It is known that the time-domain transient signal processing algorithms play a significant role in the resolving power and abundance accuracies achieved.⁴⁴ For example, Hanning apodization employed prior to Fourier transform to the experimental

Table 2. Mass Accuracies and Abundance Accuracies of Isotopic Fine Structure Distributions Obtained in the FT-ICR MS Measurements^a

peptide	isotopic peak	$(m/z)_{exp}$	A_{exp} %	$(m/z)_{th}$	A_{th} %	mass accuracy, ppm	abundance accuracy, % (random error, %)
VGPPGFSPFVG	¹⁵ N	1061.54515	7.51	1061.54324	7.15	1.79	5.07 (0.68)
	¹³ C	1061.55171	N/A	1061.54956	N/A	2.03	N/A
	¹⁵ N ¹³ C	1062.54864	14.8	1062.54660	14.6	1.92	1.35 (1.34)
	¹⁸ O	1062.55253	12.7	1062.55045	17.2	1.95	-26.1 (1.56)
	¹³ C ₂	1062.55492	N/A	1062.55292	N/A	1.89	N/A
bradykinin	¹⁵ N	1061.56754	9.93	1061.56571	10.13	1.72	-1.96 (0.76)
	¹³ C	1061.57411	N/A	1061.57203	N/A	1.96	N/A
	¹⁵ N ¹³ C	1062.57108	20.64	1062.56906	20.68	1.89	-0.18 (1.49)
	¹⁸ O	1062.57496	10.4	1062.57292	15.8	1.92	-33.9 (2.95)
	¹³ C ₂	1062.57736	N/A	1062.57538	N/A	1.86	N/A
<i>E. coli</i> peptide 2	¹⁵ N	770.04223	11.6	770.04079	9.31	1.87	24.6
	¹³ C	770.04427	N/A	770.04290	N/A	1.78	N/A
	³⁴ S	770.37511	12.8	770.37371	16.1	1.82	-20.5
	¹⁵ N ¹³ C	770.37668	14.4	770.37524	18.8	1.87	-23.4
	¹³ C ₂	770.37872	N/A	770.37735	N/A	1.78	N/A
<i>E. coli</i> peptide 1	¹⁵ N	770.04301	9.8	770.04139	11.9	2.1	-17.9
	¹³ C ₂	770.0449	N/A	770.04350	N/A	1.82	N/A
	¹⁵ N ¹³ C	770.37762	20.7	770.37584	24.1	2.31	-14.2
	¹³ C ₂	770.37935	N/A	770.37795	N/A	1.82	N/A
RVMRGMR	¹⁵ N	453.74859	20.8	453.74817	15.4	0.93	34.7
	¹³ C	453.75191	N/A	453.75133	N/A	1.29	N/A
	³⁴ S	454.24819	153.3	454.24755	128.6	1.42	19.2
	¹⁵ N ¹³ C	454.25027	49.3	454.24984	31.8	0.94	55.1
	¹⁸ O	454.2524	26.7	454.25177	23.6	1.38	13
RSHRGHR	¹³ C ₂	454.2536	N/A	454.25300	N/A	1.31	N/A
	¹⁵ N	453.74885	20.9	453.7484	19.3	0.99	8.28
	¹³ C	453.75229	N/A	453.75156	N/A	1.61	N/A
	¹⁵ N ¹³ C	454.25054	55	454.25008	39.7	1.02	38.4
	¹⁸ O	454.25272	20	454.25200	26.6	1.57	-24.7
¹³ C ₂	454.25388	N/A	454.25324	N/A	1.41	N/A	

^aThe random abundance errors are provided, in the brackets, only for the experiments with bradykinin and H-VGPPGFSPFVG-OH peptide mixture (see the Experimental Section and the Supporting Information for more details). N/A indicates ¹³C and ¹³C₂ peaks used as a reference in calculations of relative abundances.

transient did not completely resolve the ¹⁸O isotopic peak from the ¹³C₂ one. The application of another type of the apodization function, Hamming apodization, provided better resolution of the A+2 isotopic pattern for the *E. coli* peptide 2, so that even the ¹⁸O isotopic peak was resolved, see Figure S3 in the Supporting Information for more details. However, e.g., for the ¹⁵N¹³C isotopic peak of *E. coli* peptide 2 shown in Figure 3 and Figure S3, the abundance accuracy, $\Delta A/A$, of 37% was achieved when Hamming apodization was applied and -23.4% in the case of Hanning apodization. Therefore, Hanning apodization was used throughout the paper.⁴⁵ The abundance accuracies obtained in this way are reported in Table 2. The achieved abundance accuracies are limited by systematic errors and demonstrate a need for better resolution of the corresponding isotopic ions of the isobaric peptides.

Isotopic Fine Structures of Synthetic Isobaric Peptide Pair. Figure 4 shows the routine application of high magnetic field FT-ICR MS to a rationally designed pair of peptides, H-RVMRGMR-OH and H-RSHRGHR-OH. The monoisotopic mass difference between these peptides in the neutral or singly charged states is only 0.46 mDa. Figure 4 shows that the employed FT-ICR MS was not able to resolve the ¹³C isotopic patterns of these peptides, as the monoisotopic peaks of the

doubly charged ions differ in m/z only by 0.25 mTh (0.5 ppm). Previously, these peptides were baseline resolved with a record resolving power of ~ 3.3 million in a dedicated experiment by Marshall and co-workers and represent a case of the smallest, less than a mass of an electron (0.55 mDa), peptide mass difference resolved up to date.⁴⁶ Here, we routinely achieved the resolving power of ~ 2 million for this peptide pair when the peptides were doubly protonated. The resolving power achieved was sufficient to resolve the isotopic fine structures of both peptides, Figures 4 B and 4 C. Due to the presence of sulfur in one of the peptides in this pair, the ³⁴S peak in the isotopic fine structure can be an obvious marker of this peptide in the mixture. The abundance accuracies, experimental versus theoretical values of ¹³C peak to ¹⁵N peak ratio, for this peptide pair are reported in Table 2. The obtained isotopic fine structure information can potentially be used for peptide identification; however, the achieved abundance accuracies are significantly limited by a systematic error of spectral interference due to insufficient resolution of the same types of isotopic peaks of the isobaric peptides.

Toward Fast Acquisition of High Resolution Isotopic Fine Structures of Peptides. The 10 T FT-ICR MS transient signal length of at least 12 s and more is required to provide the

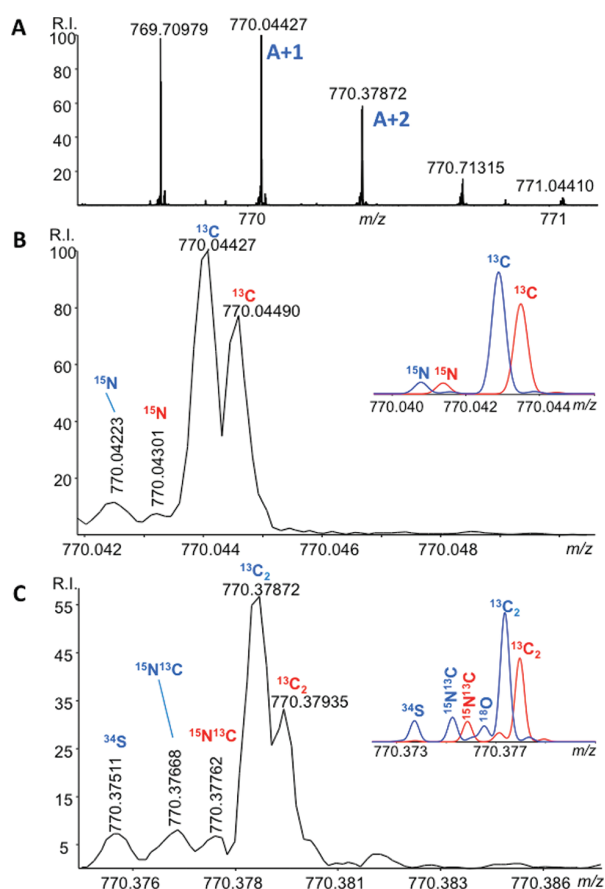


Figure 3. ESI FT-ICR MS of *E. coli* peptide pair showing (A) the ^{13}C isotopic envelope of the triply charged peptides and the expanded segments of this mass spectrum demonstrating (B) the isotopic fine structure of the A+1 isotopic envelope and (C) the isotopic fine structure of the A+2 isotopic envelope. Ion signals corresponding to the *E. coli* peptide 1 are shown in red (the heavier peptide) and to the *E. coli* peptide 2 in blue (the lighter peptide). Abundance accuracies of the experimental values versus the theoretical ones are shown in Table 2, whereas the insets represent the theoretical isotopic fine structure distributions.

high resolution isotopic fine structure information for the isobaric peptide pairs, Figures 2-4. In case of isotopic fine structure measurements for a less complex ion mixture, such as single peptide ion, under the same experimental conditions, shorter acquisition time can be enough, nonetheless the one of several seconds and more is required, Figure S4 (Supporting Information). Even if a single transient provides a sufficient signal-to-noise ratio, such long acquisition times are impractical for the modern macromolecular analysis with the time constraints dictated by the LC-based rapid separation of peptide mixtures in the bottom-up proteomics. Therefore, the peptide isotopic fine structures must be obtained from not longer than ~ 1 – 2 s transients, which would be comparable with the typically obtained transient signals in the bottom-up proteomics and metabolomics by both ICR and Orbitrap FTMS. As introduced above, the filter diagonalization method, FDM, was selected as a method of advanced signal processing to accomplish this task. Figure 5 shows a comparison between FDM-based MS (FDM MS) and FFT (FTMS) processing of a time-domain transient signal of isolated singly charged ions of bradykinin and its isobaric peptide, Figure 2, acquired on a 10

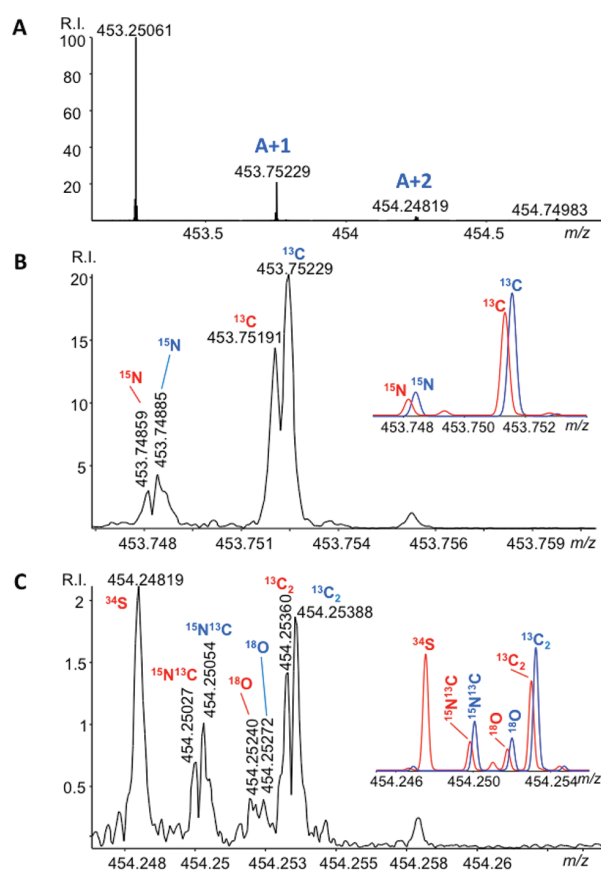


Figure 4. ESI FT-ICR MS of H-RVMRGMR-OH and H-RSHRGHR-OH peptide mixture showing (A) the ^{13}C isotopic envelope of the doubly charged peptides and the expanded segments of this mass spectrum demonstrating (B) the isotopic fine structure of the A+1 isotopic envelope and (C) the isotopic fine structure of the A+2 isotopic envelope. Ion signals corresponding to peptide H-RVMRGMR-OH are shown in red (the lighter peptide) and to H-RSHRGHR-OH peptide in blue (the heavier peptide). Abundance accuracies of the experimental values versus the theoretical ones are shown in Table 2, whereas the insets represent the theoretical isotopic fine structure distributions.

FT-ICR MS for different lengths of the signal. As expected, mass spectra from both FDM and FFT processing of 12.13 s transient signals revealed the isotopic fine structures of the A+1 isotope of bradykinin and its isobaric peptide, Figure 5 top. The FFT of a shorter, 1.66 s, initial portion of the recorded transient produced the mass spectrum without isotopic fine structure resolution. On the other hand, FDM of the same short transient signal revealed the isotopic fine structures, Figure 5 bottom. Even a less than 1 s transient signal can be sufficient for FDM-based signal processing to obtain the required information, see the results for a peptide substance P, Figure S4. Therefore, with FDM as a signal processing tool, the required length of a transient signal to obtain the isotopic fine structures of peptides in FT-ICR MS could be significantly reduced. Note, a signal perturbation is a characteristic of the time-domain transients followed by ion excitation event in FT-ICR MS.^{32,47} The FDM processing of a part of the transient that contains these perturbations demonstrates a reduced method performance. Therefore, in Figure 5 and Figure S4 a time delay of, respectively, 0.16 and 1.0 s between ion

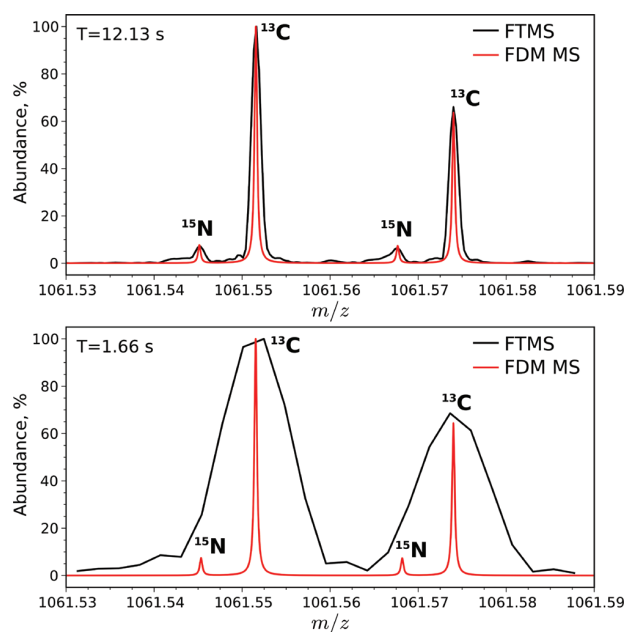


Figure 5. Comparison of super-resolution and Fourier transform signal processing to obtain the A+1 isotopic fine structure of singly protonated bradykinin and its isobaric peptide H-VGPPGFSPFVG-OH from the transient signal of 12.13 s length (top panel) and from its truncated version (1.66 s, bottom panel) using the conventional fast Fourier transform, FFT (shown in black and denoted as FTMS) and filter diagonalization method, FDM (shown in red and denoted as FDM MS). See Table 3 for more details. Bradykinin is a heavier peptide in this isobaric peptide pair.

excitation and detection was employed to exclude a part of the transient with the perturbations. The transient signal perturbation is significantly shorter in Orbitrap FTMS, due to the absence of the RF-excitation events. The bar-plot representation of the FDM MS data from Figure 5 is shown in Figure S5.

The mass accuracies and abundance accuracies for the data reported in Figure 5 are summarized in Table 3. The mass accuracies in both FTMS and FDM MS approaches suggest the systematic m/z error of the mass calibration to be the main contributor to the mass accuracy values, as in all previous experiments, Table 2. The abundance accuracies in the FTMS

approach are consistent with the ones from the control experiments with bradykinin and its isobaric peptide. Note, the abundance measurement in the FDM MS approach is more accurate for shorter transients, Figure 5 bottom and top panels. The obtained abundance accuracies demonstrate the possible numerical effects of the FDM due to the insufficient quality of the transient signals. In general, the quality of transient signals is limited by (i) the space charge effects in the ICR cell,^{32,48,49} (ii) imprecision and inharmonicity of the cylindrical ICR cell,^{32,50} and (iii) insufficient signal-to-noise ratio, though it has been reported that FDM is relatively stable in case of noisy signals.⁵¹ Therefore, the obtained results show the potential of FDM, once the transient signals of the appropriate quality are available, in obtaining isotopic fine structures in ICR MS, and, presumably, high-field Orbitrap FTMS,¹² on the LC-time scale. The computational resources required for FDM MS are substantially more significant than for the FTMS.³³ The FDM MS calculations performed in the current work required up to 30 min per single isotopic fine structure cluster calculations on a quad-core desktop computer. Therefore, application of FDM MS to large-scale LC/MS data sets requires a substantial increase in computational power. Luckily, the current FDM MS implementation allows method parallelization on multicore workstations.

CONCLUSIONS

The isotopic fine structures of the macromolecules, e.g., peptides and heavy metabolites, are the important source of structural information that is not currently being used in the routine MS workflows. Here, we have demonstrated that peptide identification in the bottom-up proteomics might be improved by considering the elemental composition information provided by the isotopic fine structure. Employing the high magnetic field FT-ICR MS for 3 pairs of isobaric peptides that differ in mass, within each pair, by 22.47 mDa, 1.81 mDa, and 0.46 mDa we have further shown that the current state-of-the-art high resolution MS is sufficient to achieve the required performance level and obtain this information in a relatively routine experiment. However, the time constraints of the modern MS experiments drastically limit the application of even the high magnetic field FT-ICR MS due to the long data acquisition time: more than 12 s were needed here to achieve the required resolution for the peptide pairs considered. The upcoming hardware developments, including even higher

Table 3. Mass Accuracies and Abundance Accuracies of Isotopic Fine Structure Distributions Obtained in the ICR MS Measurements of Bradykinin and H-VGPPGFSPFVG-OH Peptide Mixture Using FFT (Denoted as FTMS) and FDM (Denoted as FDM MS) Signal Processing^a

acquisition time, s	peptide	isotopic peak	mass accuracy, ppm		abundance accuracy, %	
			FTMS	FDM MS	FTMS	FDM MS
12.13	H-VGPPGFSPFVG-OH	¹⁵ N	1.91	1.78	-0.1	6.18
		¹³ C	1.94	1.87	N/A	N/A
	bradykinin	¹⁵ N	1.83	1.84	-2.69	13.16
		¹³ C	1.86	1.87	N/A	N/A
1.66	H-VGPPGFSPFVG-OH	¹⁵ N	N/R	1.97	N/R	2.97
		¹³ C	N/R	1.86	N/A	N/A
	bradykinin	¹⁵ N	N/R	2.33	N/R	10.9
		¹³ C	N/R	1.85	N/A	N/A

^aNotations are as follows: N/A – not applicable; N/R – not resolved. N/A indicates ¹³C peaks used as a reference in calculations of relative abundances.

magnetic field FT-ICR MS equipped with harmonized ICR cells, as well as the recent implementation of high-field Orbitrap FTMS, present a strong potential for practical LC/MS by decreasing the experimental time and improving the sensitivity of ion detection at higher resolution. Complementarily, the recent advanced signal processing development accompanied by the substantial increase in the computation power accessible by a typical research laboratory together with the continuous progress in the MS hardware development show a significant promise for applications in large-scale LC/MS-based experiments. Importantly, we demonstrated here that the application of a filter diagonalization method, FDM, to the processing of the experimental FT-ICR MS transients of peptides allows resolving the peptide's isotopic fine structure from a less than 1–2 s transient signals. Thus, the isotopic fine structures of peptides could be potentially revealed by FDM in the common LC/MS analysis of complex peptide mixtures, where it is crucial to have a rapid MS analysis step, and used further to improve protein identification in bottom-up proteomics. We expect to achieve a similar or better level of performance of FDM signal processing for isotopic fine structure analysis on the Orbitrap FTMS transient signals.

■ ASSOCIATED CONTENT

📄 Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +41216939751. E-mail: yury.tsybin@epfl.ch. Corresponding author address: EPFL ISIC LSMB, BCH 4307, 1015 Lausanne, Switzerland.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation (project 200021-125147/1), Joint Research Project of Scientific & Technological Cooperation Program Switzerland-Russia (grant agreement 128357 between EPFL and INEPCP RAS), Russian Basic Science Foundation (project 11-04-00515 to M.V.G.), and European Research Council (ERC Starting Grant 280271 to YOT).

■ REFERENCES

- (1) Domon, B.; Aebersold, R. *Science* **2006**, *312*, 212–217.
- (2) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198–207.
- (3) Liu, T.; Belov, M. E.; Jaitly, N.; Qian, W.-J.; Smith, R. D. *Chem. Rev.* **2007**, *107*, 3621–3653.
- (4) Tsybin, Y. O.; Quinn, J. P.; Tsybin, O. Y.; Hendrickson, C. L.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 762–771.
- (5) Cooper, H. J.; Hakansson, K.; Marshall, A. G. *Mass Spectrom. Rev.* **2005**, *24*, 201–222.
- (6) Comisarow, M. B.; Marshall, A. G. *Chem. Phys. Lett.* **1974**, *25*, 282–283.
- (7) Marshall, A. G. *Int. J. Mass Spectrom.* **2000**, *200*, 331–356.
- (8) Nikolaev, E. N.; Neronov, Y. I.; Gorshkov, M. V.; Talroze, V. L. *JETP Lett.* **1984**, *39*, 534–536.
- (9) Knobler, M.; Wanczek, K. P. In *Proceedings of the 45th ASMS conference on mass spectrometry and allied topics*. Palm Springs, CA, June 1–5, 1997.
- (10) Shi, S. D.-H.; Hendrickson, C. L.; Marshall, A. G. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11532–11537.
- (11) Makarov, A. *Anal. Chem.* **2000**, *72*, 1156–1162.
- (12) Makarov, A.; Denisov, E.; Lange, O. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1391–1396.
- (13) Denisov, E.; Damoc, E.; Makarov, A.; Lange, O. In *Proceedings of the 59th ASMS conference on mass spectrometry and allied topics*. Denver, CO, June 5–9, 2011.
- (14) Kind, T.; Fiehn, O. *BMC Bioinf.* **2006**, *7*, 234.
- (15) Miura, D.; Tsuji, Y.; Takahashi, K.; Wariishi, H.; Saito, K. *Anal. Chem.* **2010**, *82*, 5887–5891.
- (16) Gorshkov, M. V.; Pasa Tolic, L.; Udseth, H. R.; Anderson, G. A.; Bruce, J. E.; Prior, D. C.; Hofstadler, S. A.; Smith, R. D. In *Proceedings of the 46th ASMS conference on mass spectrometry and allied topics*. Orlando, FL, May 31–June 4, 1998.
- (17) Miladinovic, S. M.; Kozhinov, A. N.; Gorshkov, M. V.; Tsybin, Y. O. In *Proceeding of the 59th ASMS conference on mass spectrometry and allied topics*. Denver, CO, June 5–9, 2011.
- (18) Nikolaev, E.; Boldin, I.; Jertz, R.; Baykut, M. G. In *Proceedings of the 59th ASMS conference on mass spectrometry and allied topics*. Denver, CO, June 5–9, 2011.
- (19) Tsybin, Y. O.; Fornelli, L.; Kozhinov, A. N.; Vorobyev, A.; Miladinovic, S. M. *Chimia* **2011**, *65*, 641–645.
- (20) Blake, S. L.; Walker, S. H.; Muddiman, D. C.; Hinks, D.; Beck, K. R. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 2269–2275.
- (21) Stoll, N.; Schmidt, E.; Thurow, K. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1692–1699.
- (22) Kim, S.; Rodgers, R. P.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **2006**, *251*, 260–265.
- (23) Zubarev, R. A.; Hakansson, P.; Sundqvist, B. *Anal. Chem.* **1996**, *68*, 4060–4063.
- (24) Vorobyev, A.; Gorshkov, M. V.; Tsybin, Y. O. *Int. J. Mass Spectrom.* **2011**, *306*, 227–231.
- (25) Beu, S. C.; Blakney, G. T.; Quinn, J. P.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2004**, *76*, 5756–5761.
- (26) Xian, F.; Hendrickson, C. L.; Blakney, G. T.; Beu, S. C.; Marshall, A. G. *Anal. Chem.* **2010**, *82*, 8807–8812.
- (27) Qi, Y. L.; Barrow, M. P.; Van Orden, S. L.; Thompson, C. J.; Li, H.; Perez-Hurtado, P.; O'Connor, P. B. *Anal. Chem.* **2011**, *83*, 8477–8483.
- (28) Qi, Y.; Thompson, C. J.; Van Orden, S. L.; O'Connor, P. B. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 138–147.
- (29) Cargile, B. J.; Stephenson, J. L., Jr. *Anal. Chem.* **2004**, *76*, 267–275.
- (30) Mandelshtam, V. A. *Prog. Nucl. Magn. Reson. Spectrosc.* **2001**, *38*, 159–196.
- (31) Aizikov, K.; O'Connor, P. B. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 836–843.
- (32) Aizikov, K.; Mathur, R.; O'Connor, P. B. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 247–256.
- (33) Kozhinov, A. N.; Tsybin, Y. O. *Anal. Chem.* **2012**, *84*, 2850–2856.
- (34) Ben Hamidane, H.; Vorobyev, A.; Tsybin, Y. O. *Eur. J. Mass Spectrom.* **2011**, *17*, 321–331.
- (35) Blakney, G. T.; Robinson, D. E.; Hgan, V. L.; Kelleher, N. L.; Hendrickson, C. L.; Marshall, A. G. In *Proceedings of the 53rd ASMS conference on mass spectrometry and allied topics*. San Antonio, TX, June 4–9, 2005.
- (36) Blake, S.; Walker, S.; Muddiman, D.; Hinks, D.; Beck, K. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 2269–2275.
- (37) <http://www.nist.gov/pml/data/comp.cfm> (accessed December 3, 2011).
- (38) Gordon, E. F.; Muddiman, D. C. *J. Mass Spectrom.* **2001**, *36*, 195–203.
- (39) Bresson, J. A.; Anderson, G. A.; Bruce, J. E.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 799–804.
- (40) Hofstadler, S. A.; Bruce, J. E.; Rockwood, A. L.; Anderson, G. A.; Winger, B. E.; Smith, R. D. *Int. J. Mass Spectrom. Ion Processes* **1994**, *132*, 109–127.
- (41) Easterling, M. L.; Amster, I. J.; van Rooij, G. J.; Heeren, R. M. A. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1074–1082.

(42) Kozhinov, A. N.; Miladinovic, S. M.; Tsybin, Y. O. In *Proceedings of the 59th ASMS conference on mass spectrometry and allied topics*. Denver, CO, June 5–9, 2011.

(43) Silva, J. C.; Denny, R.; Dorschel, C.; Gorenstein, M. V.; Li, G.-Z.; Richardson, K.; Wall, D.; Geromanos, S. J. *Mol. Cell. Proteomics* **2006**, *5*, 589–607.

(44) Marshall, A. G.; Verdun, F. R. *Fourier transforms in NMR, optical, and mass spectrometry: a user's handbook*. Elsevier: Amsterdam, 1990.

(45) Serreqi, A.; Comisarow, M. B. *Appl. Spectrosc.* **1987**, *41*, 288–295.

(46) He, F.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2001**, *73*, 647–650.

(47) Leach, F. E., III; Kharchenko, A.; Heeren, R. M. A.; Nikolaev, E.; Amster, I. J. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 203–208.

(48) Boldin, I. A.; Nikolaev, E. N. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 3213–3219.

(49) Huang, J.; Tiedemann, P. W.; Land, D. P.; McIver, R. T.; Hemminger, J. C. *Int. J. Mass Spectrom. Ion Processes* **1994**, *134*, 11–21.

(50) Vladimirov, G.; Kostyukevich, Y.; Marshall, A. G.; Hendrickson, C. L.; Blackney, G. T.; Nikolaev, E. In *Proceedings of the 58th ASMS conference on mass spectrometry and allied topics*. Salt Lake City, UT, 2010.

(51) Benko, U.; Juricic, D. *Signal Process* **2008**, *88*, 1733–1746.

Supplementary Information

On the utility of isotopic fine structure mass spectrometry for protein identification

Saša M. Miladinović, Anton N. Kozhinov, Mikhail V. Gorshkov, and Yury O. Tsybin

Estimation of random errors in abundance measurements.

Experimental mass and abundance values are influenced by errors of systematic and random nature.¹ Summation of a number of transient signals is typically performed in FT-ICR MS to reduce the random errors. Obviously, once the random error is dominated by a systematic one in case of summation of a certain number of transients, there is no need to acquire a higher number of transients since the total error is limited by the remaining systematic one. To obtain the required number of transient signals, random errors are needed to be estimated. In general, the estimation is supposed to be performed using sufficiently high number $N \rightarrow \infty$ of measurements. Thus, to obtain the random error in abundance measurements, N spectra are needed, where each spectrum is obtained for a summation of a given number of transient. However, assuming the random errors of abundance measurements to be mainly due to the thermal noise in the detection electronic circuit, for a given peak the inverse value of the conventional spectral signal-to-noise ratio (SNR) is the upper-bound estimation of the relative random error in the abundance measurement in FT-ICR MS,¹ so that a single spectrum provides the relative random error $\sqrt{\text{var } a}/a$ of the experimental abundance a_{exp} :

$$\frac{\sqrt{\text{var } a}}{a} = SNR^{-1} = \frac{\langle a_{noise} \rangle}{a_{exp}},$$

where $\langle a_{noise} \rangle$ is the noise abundance a_{noise} averaged over the frequency domain (or mass-to-charge ratio domain as well, in case of a discrete spectrum):

$$\langle a_{noise} \rangle = \sum_{i=0}^N (a_{noise})_i / N$$

Thus, for example, estimation of the relative random error $\sqrt{\text{var } A}/A$ of the relative abundance $A = a_{15N}/a_{13C}$ of the ¹⁵N peak (with abundance a_{15N}) compared to the ¹³C peak (with abundance a_{13C}) can be obtained from the following relation:

$$\left(\frac{\sqrt{\text{var } A}}{A} \right)^2 = \left(\frac{\langle a_{noise} \rangle}{a_{13C}} \right)^2 + \left(\frac{\langle a_{noise} \rangle}{a_{15N}} \right)^2,$$

which can be simplified if $a_{15N}/a_{13C} \ll 1$, see Table 2:

$$\frac{\sqrt{\text{var } A}}{A} \cong \frac{\langle a_{noise} \rangle}{a_{15N}} = SNR_{15N}^{-1},$$

where SNR_{15N}^{-1} is the inverse SNR of the ¹⁵N peak.

Table S1. Characterization of *E.coli* tryptic peptide identification based on the measurement of monoisotopic mass and relative abundance of ¹⁵N peak compared to ¹³C one from the A+1 isotopic fine structure. A truncated mass range of 1-7 kDa, comprising most of the *in-silico* produced enzymatic peptides (132289 or about 86% of the total number of elemental compositions) is considered.

Number of elemental compositions distributed into mass bins with N candidates												
Abundance accuracy, %	10 ppm mass accuracy						5 ppm mass accuracy					
	N=1	N=2	N=3	3<N≤6	6<N≤14	14<N≤22	N=1	N=2	N=3	3<N≤6	6<N≤14	14<N≤22
NA	11309	11166	10559	31447	64576	3232	22320	21553	21312	51904	15200	0
50	11398	11346	10809	32444	63310	2982	22529	21874	21892	51617	14377	0
20	15220	15593	15279	42265	43405	527	32467	31945	25111	37653	5113	0
10	30649	28006	22563	40210	10860	1	50549	40297	23796	17141	506	0
5	60707	34391	19463	16868	860	0	82201	36525	11049	2511	3	0
2	96121	27306	7164	1698	0	0	115338	15874	1039	38	0	0
1	112722	17291	2066	210	0	0	127657	4563	68	1	0	0
0	121951	9896	441	1	0	0	132003	284	2	0	0	0
Abundance accuracy, %	1 ppm mass accuracy						0.1 ppm mass accuracy					
	N=1	N=2	N=3	3<N≤6	6<N≤14	14<N≤22	N=1	N=2	N=3	3<N≤6	6<N≤14	14<N≤22
NA	86427	36614	8143	1105	0	0	130815	1468	6	0	0	0
50	87330	36204	7730	1025	0	0	130823	1460	6	0	0	0
20	101438	26642	3791	418	0	0	131778	511	0	0	0	0
10	118005	13738	526	20	0	0	132214	75	0	0	0	0
5	120159	11858	269	3	0	0	132263	26	0	0	0	0
2	123346	8851	92	0	0	0	132271	18	0	0	0	0
1	130169	2106	14	0	0	0	132275	14	0	0	0	0
0	132261	28	0	0	0	0	132289	0	0	0	0	0

Table S2. Characterization of *E.coli* tryptic peptide identification based on the measurement of monoisotopic mass and relative abundance of ¹³C peak from the A+1 isotopic cluster compared to monoisotopic peak. A complete mass range of 1-25 kDa, comprising all *in-silico* produced enzymatic peptides (152832 elemental compositions), is considered.

Number of elemental compositions distributed into mass bins with N candidates												
Abundance accuracy, %	10 ppm mass accuracy						5 ppm mass accuracy					
	N=1	N=2	N=3	3<N≤6	6<N≤14	14<N≤22	N=1	N=2	N=3	3<N≤6	6<N≤14	14<N≤22
NA	15647	13942	13349	38973	67689	3232	29972	26177	25314	56119	15250	0
50	15647	13942	13349	38973	67689	3232	29972	26177	25314	56119	15250	0
20	15928	14000	13457	39489	66830	3128	30211	26467	25502	55660	14992	0
10	19846	16722	15816	43275	55621	1552	35038	31720	27371	49037	9666	0
5	31225	25381	21511	48679	25963	73	62918	38120	24948	25190	1656	0
2	67581	34659	24517	24817	1258	0	98496	37245	12881	4184	26	0
1	110856	29785	8557	3576	58	0	132369	17231	2683	549	0	0
0	125125	23538	3767	402	0	0	143698	8984	150	0	0	0
Abundance accuracy, %	1 ppm mass accuracy						0.1 ppm mass accuracy					
	N=1	N=2	N=3	3<N≤6	6<N≤14	14<N≤22	N=1	N=2	N=3	3<N≤6	6<N≤14	14<N≤22
NA	152832	152832	152832	152832	152832	152832	152832	152832	152832	152832	152832	152832

NA	103835	39508	8378	1111	0	0	151358	1468	6	0	0	0
50	103835	39508	8378	1111	0	0	151358	1468	6	0	0	0
20	104792	38675	8271	1094	0	0	151359	1467	6	0	0	0
10	114771	31758	5741	562	0	0	151417	1411	4	0	0	0
5	124291	25254	3166	121	0	0	152695	137	0	0	0	0
2	130368	20382	2034	48	0	0	152824	8	0	0	0	0
1	143868	8789	168	7	0	0	152826	6	0	0	0	0
0	147202	5630	0	0	0	0	152830	2	0	0	0	0

Table S3. Characterization of *E.coli* tryptic peptide identification based on the measurement of monoisotopic mass and relative abundance of ^{13}C peak from the A+1 isotopic cluster compared to monoisotopic peak. A truncated mass range of 1-7 kDa, comprising most of the *in-silico* produced enzymatic peptides (132289 or about 86% of the total number of elemental compositions) is considered.

Number of elemental compositions distributed into mass bins with N candidates												
Abundance accuracy, %	10 ppm mass accuracy						5 ppm mass accuracy					
	N=1	N=2	N=3	3<N≤6	6<N≤14	14<N≤22	N=1	N=2	N=3	3<N≤6	6<N≤14	14<N≤22
NA	11309	11166	10559	31447	64576	3232	22320	21553	21312	51904	15200	0
50	11309	11166	10559	31447	64576	3232	22320	21553	21312	51904	15200	0
20	11321	11191	10585	31746	64318	3128	22353	21658	21444	51878	14956	0
10	12424	12736	12504	37953	55120	1552	24783	25245	24689	47907	9665	0
5	20162	19750	18885	47462	25957	73	46653	34656	24161	25163	1656	0
2	47997	33702	24515	24817	1258	0	78801	36397	12881	4184	26	0
1	91254	28845	8556	3576	58	0	112666	16391	2683	549	0	0
0	105486	22634	3767	402	0	0	123971	8168	150	0	0	0
Abundance accuracy, %	1 ppm mass accuracy						0.1 ppm mass accuracy					
	N=1	N=2	N=3	3<N≤6	6<N≤14	14<N≤22	N=1	N=2	N=3	3<N≤6	6<N≤14	14<N≤22
NA	86427	36614	8143	1105	0	0	130815	1468	6	0	0	0
50	86427	36614	8143	1105	0	0	130815	1468	6	0	0	0
20	86508	36585	8105	1091	0	0	130816	1467	6	0	0	0
10	95663	30428	5636	562	0	0	130874	1411	4	0	0	0
5	105176	23931	3061	121	0	0	132152	137	0	0	0	0
2	110645	19562	2034	48	0	0	132281	8	0	0	0	0
1	124145	7969	168	7	0	0	132283	6	0	0	0	0
0	127471	4818	0	0	0	0	132287	2	0	0	0	0

Figure S1. Analysis of *E.coli* tryptic digest identification based on the measurement of monoisotopic mass and relative abundance of ^{15}N peak compared to ^{13}C one from the A+1 isotopic fine structure as a function of mass and abundance accuracy.

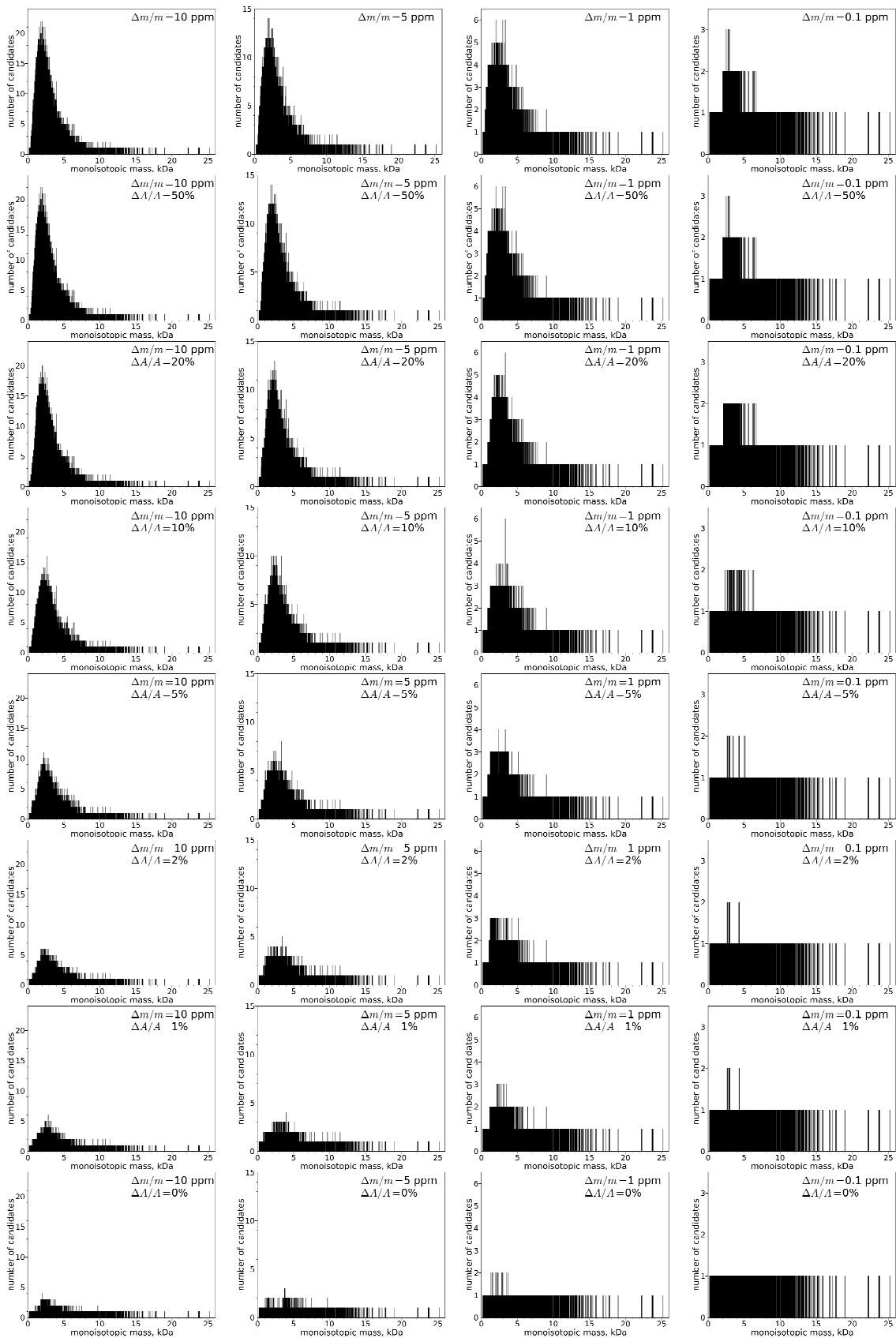


Figure S2. Analysis of *E.coli* tryptic peptide identification based on the measurement of monoisotopic mass and relative abundance of ^{13}C peak from the A+1 isotopic cluster compared to monoisotopic peak as a function of mass and abundance accuracy.

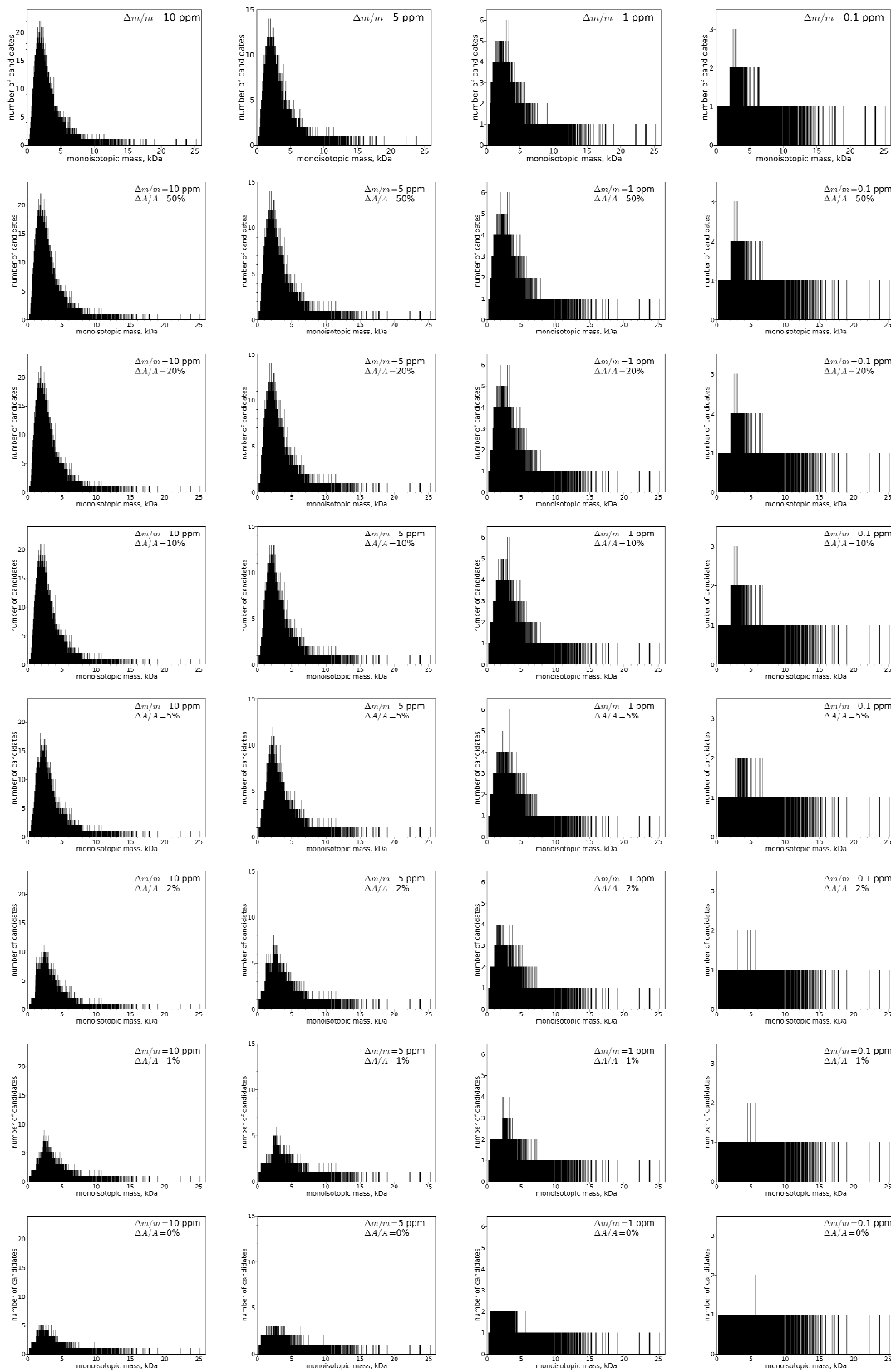


Figure S3. Influence of the apodization function type on the relative abundance error in the A+2 isotopic envelope of isobaric peptides, *E.coli* peptide 1 and *E.coli* peptide 2. Hamming apodization provides higher resolving power but lower accuracy of the relative abundances of isotopic peaks, whereas Hanning apodization slightly reduces the final resolving power obtained, but has a relatively low relative abundance error of the isotopic peaks. Therefore, Hanning apodization was taken as a standard procedure throughout the paper.

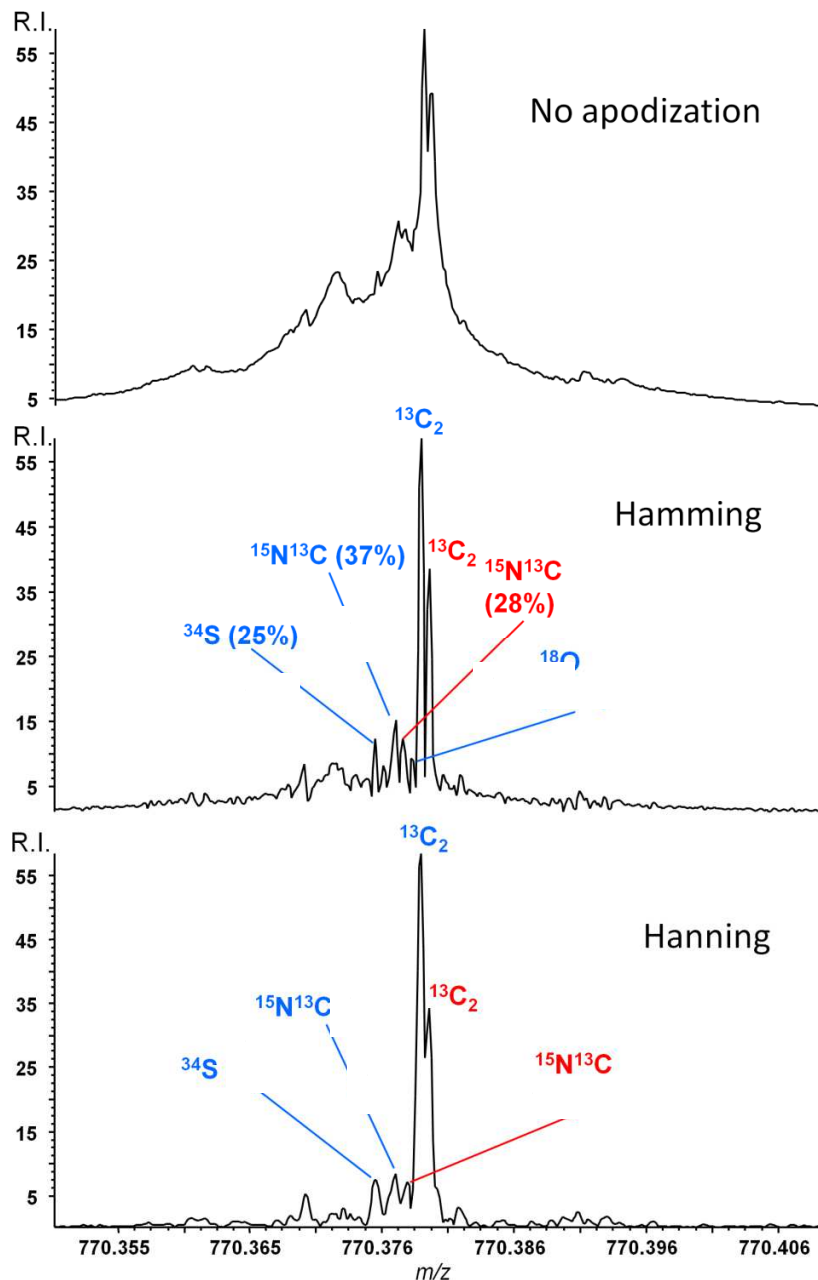


Figure S4. Advantages of FDM versus FT signal processing to obtain the isotopic fine structure shown for (A) the experimental ICR MS transient signal from the isolated doubly protonated substance P and the corresponding mass spectra obtained (B, C) from the full length of the signal (5.144 s) and (D, E) from its truncated version (first 0.84 s) using (B, D) the conventional fast Fourier transform, FTMS (denoted as FTMS) and (C, E) the filter diagonalization method, FDM (denoted as FDM MS). The obtained accuracies (experimental versus theoretical values) of relative abundances of ^{15}N peaks compared to the corresponding ^{13}C peaks are the following:

FTMS: 30% for 5.144 s transient. Not resolved for the 0.84 s transient;

FDM MS: 32% for 5.144 s transient and 20% for the 0.84 s transient.

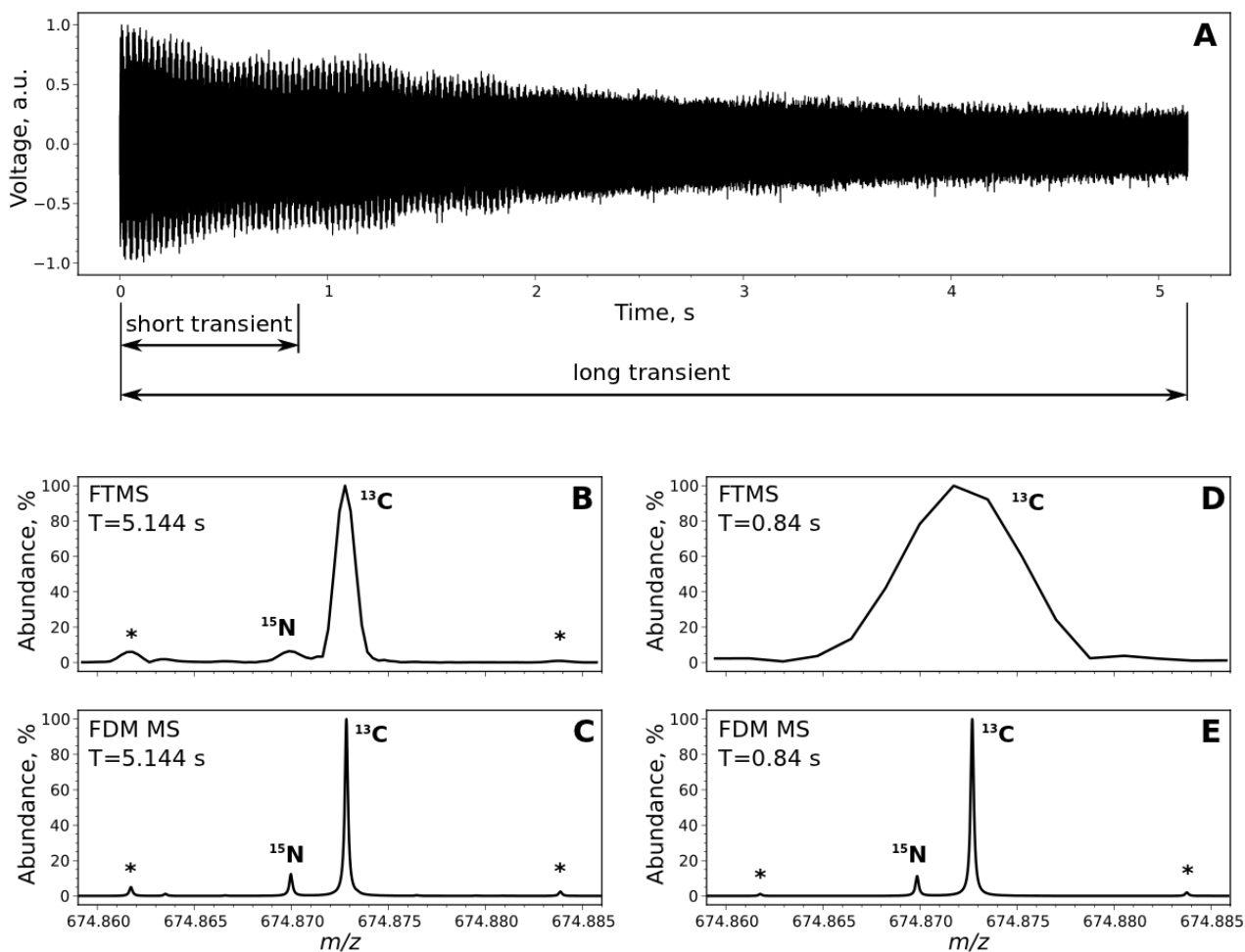
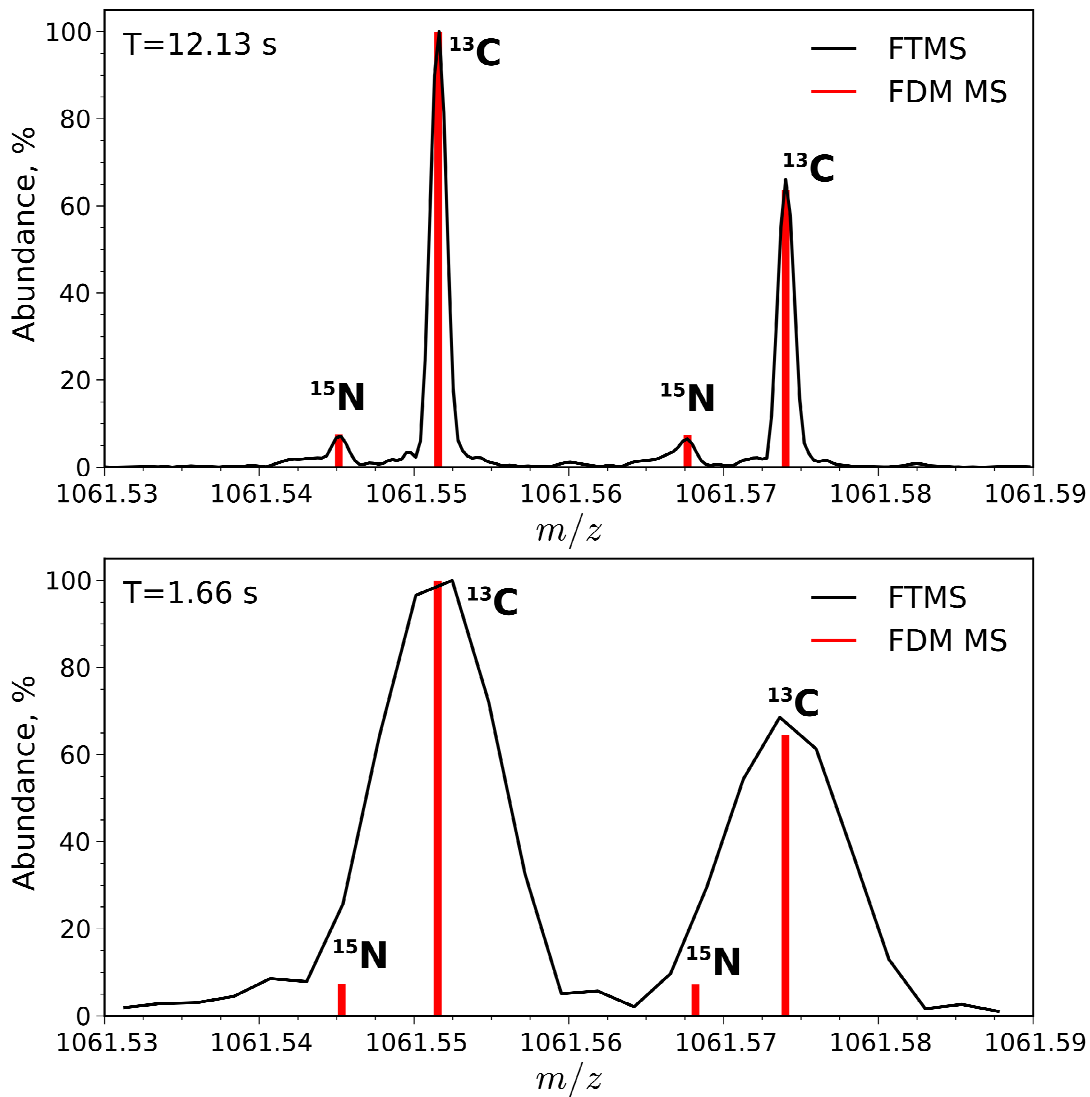


Figure S5. Bar-plot representation of the FDM MS data demonstrated in Figure 5.



References

- (1) Kozhinov, A. N.; Miladinovic, S. M.; Tsybin, Y. O., in *Proceedings of the 59th ASMS conference on mass spectrometry and allied topics*. Denver, CO, June 5-9, 2011.

A.7 Paper VII

High-Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry with Increased Throughput for Biomolecular Analysis

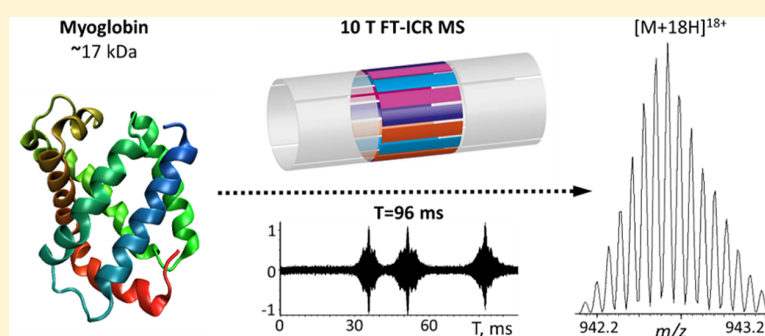
Konstantin O. Nagornov,[†] Mikhail V. Gorshkov,^{‡,§} Anton N. Kozhinov,[†] and Yury O. Tsybin^{*,†}

[†]Biomolecular Mass Spectrometry Laboratory, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

[‡]Institute for Energy Problems of Chemical Physics, Russian Academy of Sciences, 119334 Moscow, Russia

[§]Moscow Institute of Physics and Technology (State University), 141707 Dolgoprudny, Moscow Region, Russia

S Supporting Information



ABSTRACT: A multielectrode ion cyclotron resonance (ICR) cell, herein referred to as the “4X cell”, for signal detection at the quadruple frequency multiple was implemented and characterized on a commercial 10 T Fourier transform ICR mass spectrometer (FT-ICR MS). Notably, with the 4X cell operating at a 10 T magnetic field we achieved a 4-fold increase in MS acquisition rate per unit of resolving power for signal detection periods typically employed in FTMS, viz., shorter than 6 s. Effectively, the obtained resolution performance represents the limit of the standard measurement principle with dipolar signal detection and FT signal processing at an equivalent magnetic field of 40 T. In other words, the achieved resolving powers are 4 times higher than those provided by 10 T FT-ICR MS with a standard ICR cell. For example, resolving powers of 170 000 and 70 000 were obtained in magnitude-mode Fourier spectra of 768 and 192 ms apodized transient signals acquired for a singly charged fluorinated phosphazine (m/z 1422) and a 19-fold charged myoglobin (MW 16.9 kDa), respectively. In peptide analysis, the baseline-resolved isotopic fine structures were obtained with as short as 768 ms transients. In intact protein analysis, the average resolving power of 340 000 across the baseline-resolved ^{13}C isotopic pattern of multiply charged ions of bovine serum albumin was obtained with 1.5 s transients. The dynamic range and the mass measurement accuracy of the 4X cell were found to be comparable to the ones obtained for the standard ICR cell on the same mass spectrometer. Overall, the reported results validate the advantages of signal detection at frequency multiples for increased throughput in FT-ICR MS, essential for numerous applications with time constraints, including proteomics.

High resolving power and mass measurement accuracy are the most crucial characteristics of a mass spectrometer employed for structural analysis of complex molecular mixtures and large biomolecules in a number of applications including metabolomics, proteomics, and petroleomics.^{1–4} Fourier transform mass spectrometry (FTMS) offers the highest resolving power and mass accuracy among all types of MS instruments.^{5–9} However, the price for high performance of FTMS is the slow acquisition rate of mass spectra, i.e., low throughput. As a result, in many applications with time constraints, e.g., in liquid chromatography coupled with mass spectrometry (LC-MS) and protein quantification approaches based on mass spectra counting, the throughput achievable by FTMS limits its coupling with LC. Indeed, high-field Orbitrap FTMS provides resolving power of up to $\sim 300\,000$ at m/z 1000 in eFT mass spectra obtained from 1.5 s time-domain signals (transients).⁴

This is compared with $\sim 100\,000$ resolving power in magnitude-mode mass spectra provided by a 10 T FT-ICR MS for the same acquisition period and m/z range. Further developments of the FTMS technologies enabling high resolving power data acquisition at the increased acquisition rate without a sacrifice in spectral dynamic range are needed to meet the growing challenges of applications and advances in alternative, e.g., time-of-flight (TOF), mass analyzer technologies.¹⁰

What are the options to improve the FTMS throughput? In recent decades, a common approach was to increase the fundamental frequency employed for measurements of ion m/z

Received: April 18, 2014

Accepted: August 20, 2014

Published: August 20, 2014

values via increased magnetic fields in the ICR cell and increased electric field in the orbitrap. However, for the latter, it is getting increasingly difficult to control deviations of the actual electric potential from the exact quadrupole distribution when increasing the DC voltage applied between the orbitrap's electrodes or when reducing the size of the orbitrap.⁹ Moreover, these deviations may result in adverse effects on other instrument performance characteristics, including mass measurement accuracy, dynamic range, and coalescence threshold.¹¹ Finally, this approach also requires improved vacuum conditions in the mass analyzer, whereas these are already near the technological limits.

In the case of FT-ICR MS,^{12–14} apart from economic considerations, the increase in magnetic field has its technological limitations as well. Currently, a number of projects to build 21 T FT-ICR MS are in progress, holding promise for the corresponding improvements in the resolution-per-acquisition rate metrics compared with 10 to 15 T FT-ICR MS instruments commercially available now.⁷ Notably, whereas some of the metrics such as the peak coalescence threshold and the spectral dynamic range improve with magnetic field squared, the resolving power per fixed acquisition period is simply proportional to the magnitude of magnetic field when the standard methods for signal detection and signal processing are employed. Recent innovations in FT-ICR MS cell technologies extend the achievable resolving powers via enabling acquisition of transient signals with up to pressure-limited lifetimes via improved harmonization of the electric field inside the ICR cell.^{15–21} However, these approaches to the ICR cell designs aim at reducing the ions dephasing effect on the transients due to dependencies of the measured frequency components on the amplitude of ion trapping oscillations^{22–25} and do not increase the resolving power per fixed acquisition period. In other words, those developments, while being interesting and impressive, are not getting closer to meeting the emerging demands for higher resolution at higher acquisition rates.

Apart from increased fundamental frequency and improved harmonization, developments of the measurement principle of FTMS, including methods for signal processing and signal detection, are vital for achieving high throughput of high-resolution FTMS. For instance, the step forward in increasing FTMS throughput was reimplementing the adsorption-mode FT spectral representation to FTMS practice, for both Orbitrap and ICR FTMS.^{26–30} The spectral lines in absorption mode have up to twice higher peak resolution in FTMS mass spectra, thus accordingly increasing the high-resolution acquisition rate.³¹ Implementations similar to that for Orbitrap have been realized for other FTMS-based electrostatic ion traps.³² Another signal processing option to increase the throughput for high-resolution FTMS-based analysis is taking advantage of super-resolution (SR) methods for signal processing, whose uncertainty principles for frequency measurements are less strict than that of FT.^{33–35} Particularly, it was shown that the resolving power sufficient for obtaining fine isotopic patterns in peptide mass spectra becomes feasible with the filter-diagonalization method (FDM) applied to as short as 1 s time-domain signals acquired on 10 T FT-ICR MS.³⁶ Similar results have been obtained with SR signal processing based on least-squares fitting (LSF).³⁷

In turn, developments in signal detection are represented by the frequency multiple method that improves the resolution-per-acquisition-rate performance of FTMS.^{38–41} Although this

method for signal detection has been known for almost 30 years, it has been poorly explored. Previously, it demonstrated its capabilities in achieving the highest in FT-ICR MS resolving power of 200 000 000 for the isolated ions of m/z 130 and 7 T magnetic field.⁴² More recently, we demonstrated the envisioned utility of this approach for improving the resolution by a factor of 2 in FT-ICR MS for protein analysis at double frequency multiple.⁴³ Despite these insights, the frequency multiple detection method has not been implemented in the commercial FT-ICR MS instruments designed for proteomics applications. The main reason has been the perception that this method does not improve the resolving power due to the inhomogeneous line broadening mostly caused by the Coulomb interactions between ion clouds, as well as the spatial ion distribution inside the ICR cell.⁴⁴ In addition to acceleration of signal decay rate, these factors would result in measurable frequency drift during ion detection.⁴⁵ The latter would translate into multiplication of the peak width when frequency multiple detection is applied and, thus, lose the only advantage of the method. However, there is a shortage in this historic viewpoint because the performance of FT-ICR MS at frequency multiples was estimated using the model of highly damped transients with exponential decay.

Indeed, ion–ion interactions and variation in the number of coherently rotating ions result in the frequency drift during the acquisition period that exceeds transient decay. The ion clouds with small number of ions are more affected and decay faster compared to their “big brother” counterparts of different m/z due to the destabilization of long-term coherence of their cyclotron motion.^{46,47} On the other hand, the acquisition of relatively short, undamped transients presents significantly better opportunities for the quantitative applications because it provides higher acquisition rate and alleviates discrimination of low abundance ions. In general, it is assumed now that shorter acquisition times will translate into more interrogated peptides and more proteins being identified, as well as improvements in the reproducibility of proteomics experiments.^{48,49} Therefore, implementation of frequency multiple detection of undamped time-domain signals at high acquisition rate could be considered as a complementary approach to the increase in magnetic field strength in FT-ICR MS.

Here we describe implementation and performance of a multielectrode cylindrical ICR cell allowing detection of the ion signal at the quadruple frequency multiple. We further refer to this cell as a “4X cell”.

■ EXPERIMENTAL METHODS

Sample Preparation. MRFA, substance P, myoglobin (horse), and Ultramark were obtained from Sigma-Aldrich (Buchs, Switzerland). Albumin (bovine) was purchased from Fisher Scientific (Schwerte, Germany), and polyphosphoric acid was obtained from Fluka (Buchs, Switzerland). A pair of model peptides was produced by solid phase Fmoc chemistry on an Applied Biosystems 433 A Synthesizer with further purification by liquid chromatography. Model peptide sequences were based on human beta crystalline B2 (P43320) and corresponded to peptide P2, GYEYLLPEGDFR ($C_{68}H_{95}N_{15}O_{21}$, monoisotopic mass 1457.6827 Da), and its deamidated version, peptide P1, GYQYLLPEGDFR ($C_{68}H_{96}N_{16}O_{20}$, monoisotopic mass 1456.6987). LC-MS grade acetonitrile and water were obtained from Fluka (Buchs, Switzerland). Formic acid was obtained from Merck (Zug, Switzerland). The calibration mixture with MRFA

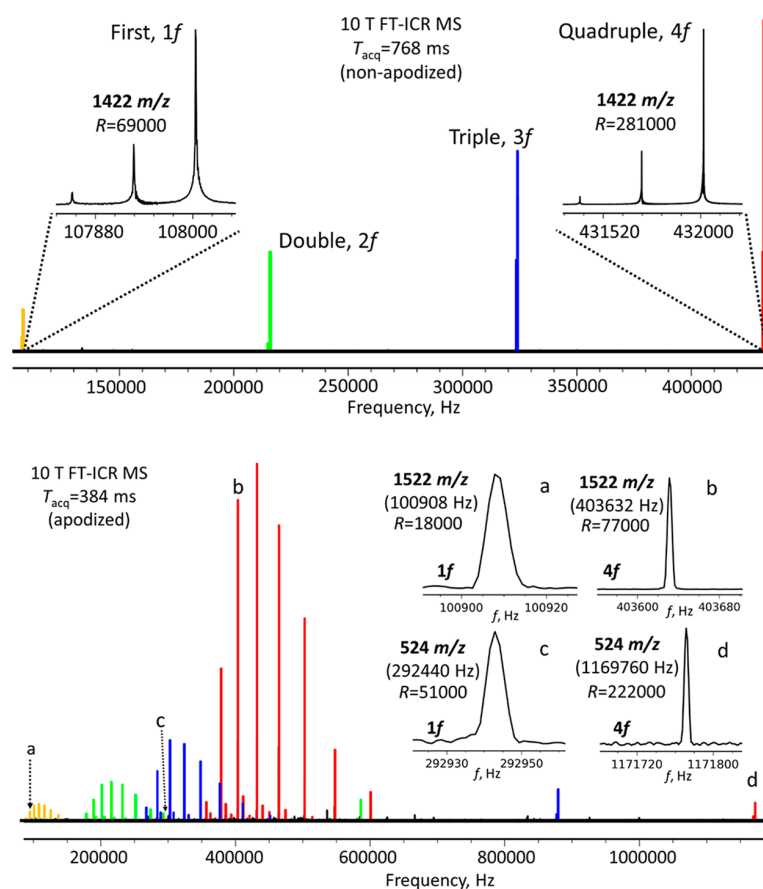


Figure 1. Fourier spectra of transient signals acquired with the 4X cell in the analysis of calibration mixture of MRFA and Ultramark compounds on 10 T FT-ICR MS: (top) spectrum of isolated singly protonated Ultramark ions at m/z 1422 (magnitude-mode FT with no signal apodization; acquisition period 768 ms; isolation width 5 m/z ; trapping potential 0.1 V; $T_{\text{exc1}} = 20$ ms and $V_{\text{exc1}} = 100$ V) and (bottom) broadband frequency spectrum of the calibration mixture (magnitude-mode FT with Hann apodization; acquisition time 384 ms; trapping potential 0.1 V; $T_{\text{exc3}} = 60$ ms and $V_{\text{exc3}} = 100$ V). The corresponding transients are shown in Figure S2 (Supporting Information). Peaks corresponding to the fundamental frequency, as well as its double, triple, and quadruple multiples, are shown in yellow, green, blue, and red, respectively. Insets show expanded views of (top) isotopic distributions and (bottom) monoisotopic ions corresponding to the fundamental frequency and the quadruple frequency multiples of selected analytes.

peptide and Ultramark was prepared in 1:1 (v/v) methanol/water solvent containing 1% of formic acid. Peptide and protein solutions were prepared in 1:1 (v/v) water/acetonitrile solvent mixtures containing 0.1% (v/v) of formic acid. The final peptide and protein concentration was $\sim 2 \mu\text{M}$.

Mass Spectrometry. All experiments were carried out on a hybrid 10 T FT-ICR MS (LTQ FT-ICR MS, Thermo Scientific, Bremen, Germany). The instrument was equipped with the custom-made 4X ICR cell based on the platform of the original standard cell (an open-ended cylindrical Ultra ICR cell, Thermo Scientific). Figure S1 (Supporting Information) shows a schematic diagram and electric wiring for the employed 4X ICR cell. Its overall dimensions, including inner diameter of ~ 56 mm and length of excitation and detection electrodes of ~ 70 mm, correspond to those of Ultra ICR cell. The original four electrodes for ion excitation and detection were replaced with 16 equal-size electrodes. The metal grid employed for ion excitation in the Ultra cell was removed. Thus, capacitive coupling of the RF excitation to the trapping ring regions has not yet been implemented for the 4X ICR cell under evaluation. The configuration of the excitation and detection electrodes of the 4X ICR cell reproduces design suggested

earlier.⁴⁰ Briefly, one electrode set consisting of eight electrodes (four electrodes for excitation 1 and four electrodes for excitation 2, Figure S1, Supporting Information) provided a standard RF dipolar ion excitation. The other electrode set consisting of eight electrodes (four electrodes for detection 1 and four electrodes for 2, Figure S1, Supporting Information) was used for induced ion current detection. Ion confinement was performed using trapping ring electrodes similar to the Ultra cell design. The offset DC potentials applied to the excitation and detection electrodes of the 4X cell were adjusted to increase the signal-to-noise ratio (SNR) and reduce magnetron sidebands.

The samples were ionized using a robotic chip-based nanoelectrospray ionization source (Triversa Nanomate, Advion Biosciences, Ithaca, NY). The ions were then isolated and accumulated in the linear ion trap (LTQ). A predetermined number of charges controlled with the automatic gain control (AGC) function were transferred into the ICR cell through the multipole ion guide system. The potential of 3 V was consequently applied to trapping rings confining ions after trapping during ion relaxation and excitation events, whereas the trapping potential was 19 kV to

the user defined value in the range of 0 to 1 V during ion detection event. Tandem mass spectrometry (MS/MS) experiments, including electron capture dissociation (ECD) and infrared multiphoton dissociation (IRMPD), were performed following standard procedures described elsewhere.⁵⁰

Three ranges of dipolar broadband frequency-sweep excitation, 72–550 kHz, 72–1050 kHz, and 72–2150 kHz, corresponding to the manufacturer settings for 280–2000, 145–2000, and 70–2000 m/z ranges, were employed to excite coherent ion motion (see Supporting Information for details). The amplitude, V_{exc} and duration, T_{exc} of the excitation signal were varied in the range from 0 to 100 V and from 0 to 60 ms, respectively, for each of the three ranges. Standard dipolar detection was utilized to acquire time domain signals. The sampling frequency employed for signal digitization was in the range of 1 to 5 MHz.¹² Transients of various lengths (0.096–25 s) were acquired in MIDAS file format using the advanced user interface capabilities. The transients were apodized with the Hann window, unless stated otherwise, zero-filled once, and Fourier transformed with magnitude-mode spectral representation. Frequency-to- m/z scale conversion of the obtained Fourier spectra was performed using the standard two-parameter calibration equation.^{51,52} Peak picking was carried out using three-point parabolic interpolation. Signal processing, analysis of mass spectra, and data visualization were performed using the data analysis framework pyFTMS developed in-house,⁵³ MIDAS software (NHMFL, Tallahassee, FL), or Xcalibur Qual Browser (Thermo Scientific).

RESULTS AND DISCUSSION

Resolving Power. Figure 1 shows the broadband frequency spectra for the Ultramark calibration mixture (bottom panel) and an isolated Ultramark component at m/z 1422 (top panel). The spectra demonstrate that the resolving power at quadruple ($4f$) frequency multiple exceeds that at single (f), double ($2f$), and triple ($3f$) frequency multiples for the 4X cell described in Experimental Methods and schematically shown in Figure S1 (Supporting Information). In both cases, the 4-fold increase in the peak resolution has been observed at the quadruple frequency compared to that at the fundamental frequency for the same acquisition periods. The frequency spectra for both the calibration mixture and the isolated ions at m/z 1422 were measured for the acquisition periods of 384 and 786 ms, respectively. Figure S2 (Supporting Information) shows the corresponding transient signals. In both cases the lengths of transients were limited by the acquisition periods, T_{acq} . In the case of the standard measurement principle with dipolar signal detection and magnitude-mode FT signal processing, the upper-bound limit of resolving power (at full width of half maximum) achieved for unapodized transients is given by the following equation:¹²

$$\frac{m/z}{\Delta m/z} = \frac{f T_{\text{acq}}}{1.205} \cong 0.132 \frac{qB}{m} T_{\text{acq}} \quad (1)$$

where B is the magnetic field strength and f is the measured frequency, which is approximated with the ion cyclotron frequency here, i.e., the magnetron shift is neglected. Therefore, for signal detection at the quadruple multiple, $n = 4$, of the fundamental frequency, $f_n = nf$, the upper-bound resolving power is

$$\frac{m/z}{\Delta m/z} \cong 0.132 \frac{qB}{m} T_{\text{acq}} n \quad (2)$$

Using eq 2 one can estimate the upper-bound resolving power for peaks in the frequency spectra shown in Figure 1. For example, the estimated resolving power for the peak at the fundamental frequency of the Ultramark component (fluorinated phosphazine) at m/z 1422 shown in Figure 1 top panel is 68 000, which correlates with the experimentally obtained value of 69 000 for the unapodized transient. The measured resolving power is 281 000 for the highest peak in the frequency spectrum shown in Figure 1, top panel, corresponding to the peak at the quadruple frequency multiple. Therefore, according to eq 2, the achieved resolving powers are equivalent to 40 T magnetic field strength.

Figure 2 shows the dependence of resolving power on the acquisition time for the isolated Ultramark component at m/z

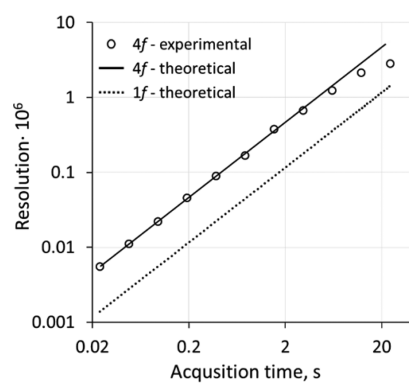


Figure 2. Characterization of the 4X ICR cell installed on a serial 10 T FT-ICR MS. Shown is the dependence of resolving power at the quadruple frequency on the acquisition period obtained with the singly protonated Ultramark component at m/z 1322. Solid and dashed lines show theoretical upper-bound limits of magnitude-mode FT for the quadruple frequency multiple and the fundamental frequency, correspondingly. Other parameters: trapping potential 0.1 V; $T_{\text{exc1}} = 20$ ms and $V_{\text{exc1}} = 100$ V.

1322. This plot demonstrates that the peak resolution at the quadruple frequency $4f$ aspired to the resolution corresponding to the fundamental frequency f for substantially long, up to 6 s and more, acquisition times. Therefore, for the acquisition rates in the range of up to one transient per several seconds, the 4X cell also 4-fold increases the resolving power. On the other hand, many MS applications, including petroleomics and proteomics, typically do not require resolution performance exceeding 1 000 000 and 100 000 at m/z 1000, correspondingly.^{4,36} For instance, given that resolving power of 100 000 at m/z 1000 is a benchmark for shotgun proteomics, the 4X cell installed on the 10 T FT-ICR MS provides this level of resolving power for the magnitude-mode mass spectra at the acquisition rates of up to 5 Hz. This performance is comparable with the state-of-the-art time-of-flight MS, high-field Orbitrap FTMS, or, in terms of resolving power, with virtually the 40 T FT-ICR MS.

Features of the Frequency (mass) Spectra Measured Using 4X Cell. Figure S3 (Supporting Information) demonstrates the comparison of ion abundances for Ultramark compounds obtained with the standard ICR cell and 4X cell under the same experimental parameters. Overall, the distributions of abundances are comparable. However, a

noticeable decrease of ion abundance was observed for peaks on both edges (922 m/z and 1822 m/z) for the 4X ICR cell compared to the standard ICR cell.

The presence of the other than quadruple frequency multiples in the mass spectra narrows the m/z range available for unambiguous measurements for the current 4X cell design and data station configuration, as shown in Figure 1, bottom panel. However, our measurements of the frequency multiples revealed that their position can be determined better than the mass measurement accuracy, *vide infra*, so that the questionable peaks from the lower order multiples can be unambiguously discriminated from the main, 4X ones. Methods of deconvolution of frequency spectra containing harmonics have recently attracted a particular attention in FTMS, and a number of software-based solutions have been proposed.^{54,55} Furthermore, due to the nonlinear dependence of the signal amplitude at the frequency multiples on the postexcitation radius, the negative effect of the presence of different harmonics along with the main one at $4f$ can be reduced by increasing the postexcitation radius.^{39,40} Finally, detailed analysis of the monoisotopic peaks at frequency multiples shows that interharmonics (magnetron sidebands) around each frequency multiple peak can be a major contribution to the intensities of frequency multiples peaks, if unresolved (Figures S4, Supporting Information).⁵⁶ This was the reason for deviations in ratio of different peaks' intensities at the different frequency multiples, which is noticeable for broadband measurements (see Figure 1).

Mass Measurement Accuracy. Mass calibration for the 4X cell was performed for the postexcitation radii that are to be used in the intended application. Indeed, due to magnetic and trapping electric field inhomogeneities, the measured (reduced cyclotron) frequency changes with postexcitation radius. This shift should be larger by a factor of 4 for the 4X cell at the $4f$ frequency. However, the relative frequency changes with postexcitation radius are expected to be the same for the signals at either frequency multiple. The dependence of reduced cyclotron frequencies of isolated doubly protonated substance P (674 m/z) shows the reduced cyclotron frequency shift of 35 ppm with the excitation amplitude (Figure S5, Supporting Information).

To evaluate the mass accuracy performance of the 4X cell, the internal recalibration of the m/z scale with respect to quadruple frequency multiples was performed using a calibration mixture of Ultramark compounds in the range of m/z 280 to 2000. The lowest edge of the available mass range at m/z 280 (assuming the quadruple frequency detection) was defined by the capabilities of the built-in data acquisition system of the instrument employed in this work. The internal calibration mass accuracy did not exceed 0.1 ppm for the number of charges (controlled by AGC) in the range of 5E4–3E6 performed with Ultramark calibration mixture in the range m/z 1000 to 2000 (data not shown).

The mass measurement accuracy (MMA) for the 4X cell was further evaluated for a mixture of polyphosphoric acid poly(H_3PO_4)_{*n*}. The broadband mass spectra were acquired at the $4f$ multiples with external calibration using parameters obtained in the previous experiment with the Ultramark calibration mixture. In this evaluation we varied two parameters related to dynamic range and the throughput of the instrument: a total number of charges loaded into the 4X cell and the acquisition period. The former parameter was controlled by the AGC target value. The results obtained for the MMA as a

function of a number of accumulated charges (AGC) and an acquisition period are shown in Table 1 and Table 2,

Table 1. Mass Accuracy of Phosphoric Acid H_3PO_4 Measurements as a Function of the Number of Charges (AGC)^a

compound	m/z	mass accuracy ($\Delta m/m$, ppm)			
		AGC = 1E5	AGC = 5E5	AGC = 1E6	AGC = 2E6
(H_3PO_4) ₁₁ H ⁺	1078.75312	1.39	0.13	0.20	0.32
(H_3PO_4) ₁₃ H ⁺	1274.70691	–	0.09	0.12	0.15
(H_3PO_4) ₁₆ H ⁺	1568.63760	0.62	0.12	–0.10	–0.17
(H_3PO_4) ₁₈ H ⁺	1764.59140	–0.54	0.32	0.15	–0.07
(H_3PO_4) ₂₀ H ⁺	1960.54518	–1.10	0.44	–0.03	–0.38

^aExternal calibration performed on 30 summed transients with acquisition period 768 ms each. The corresponding mass spectrum and graphic of dependence are shown in Figure S6 (Supporting Information).

Table 2. Mass Accuracy of Phosphoric Acid H_3PO_4 Measurements as a Function of Acquisition Period with an Internal Calibration Performed on 30 Summed Transients with 5E5 Total Number of Charges (AGC) Accumulated for Each Measurement^a

compound	mass accuracy ($\Delta m/m$, ppm)			
	$T_{acq} = 96$ ms	$T_{acq} = 768$ ms	$T_{acq} = 1.562$ s	$T_{acq} = 6.144$ s
(H_3PO_4) ₁₁ H ⁺	–0.29	–0.05	0.01	–
(H_3PO_4) ₁₂ H ⁺	0.49	0.02	0.03	–0.04
(H_3PO_4) ₁₃ H ⁺	–0.19	0.05	–0.01	0.07
(H_3PO_4) ₁₄ H ⁺	0.09	0.02	0.02	0.03
(H_3PO_4) ₁₅ H ⁺	–0.14	–0.07	–0.05	–0.06
(H_3PO_4) ₁₆ H ⁺	0.43	0.12	0.02	–0.01
(H_3PO_4) ₁₇ H ⁺	–0.51	–0.10	–0.06	–0.03
(H_3PO_4) ₁₈ H ⁺	–0.09	–0.01	0.05	0.07
(H_3PO_4) ₁₉ H ⁺	0.17	0.11	–0.08	0.01
(H_3PO_4) ₂₀ H ⁺	0.04	–0.07	0.07	–0.02
RMS value	0.30	0.07	0.05	0.04

^aThis re-calibration was performed using the known masses of polyphosphoric acid poly(H_3PO_4)_{*n*} mixture components.

respectively. For the reference, the example mass spectrum and the corresponding dependencies are presented in Figure S6, Supporting Information. The MMA below 1 ppm in the mass range of m/z 1000 to 2000 for a wide range of AGC values from 1E5 to 2E6 was demonstrated in these experiments utilizing external calibration. Table 1 also shows no visible trend in the MMA related to the number of trapped ions within the selected AGC value range. On the contrary, we found a correlation between the mass measurement accuracy and the acquisition period. Table S1 shows the results of MMA evaluations for the time-domain signals acquired for periods from 96 ms to 6.144 s.

With increasing transient lengths, the accuracy started deteriorating although it stayed within 3 ppm in the whole range of acquisition periods evaluated. We explain this effect by the increased frequency variations during longer acquisitions relative to the frequencies at which the calibration was made. The signal decay times were in the range of 1.5 s and more due to ion cloud dephasing processes. Because of a varying number of synchronously rotating ions in the clouds and the 10³–ion

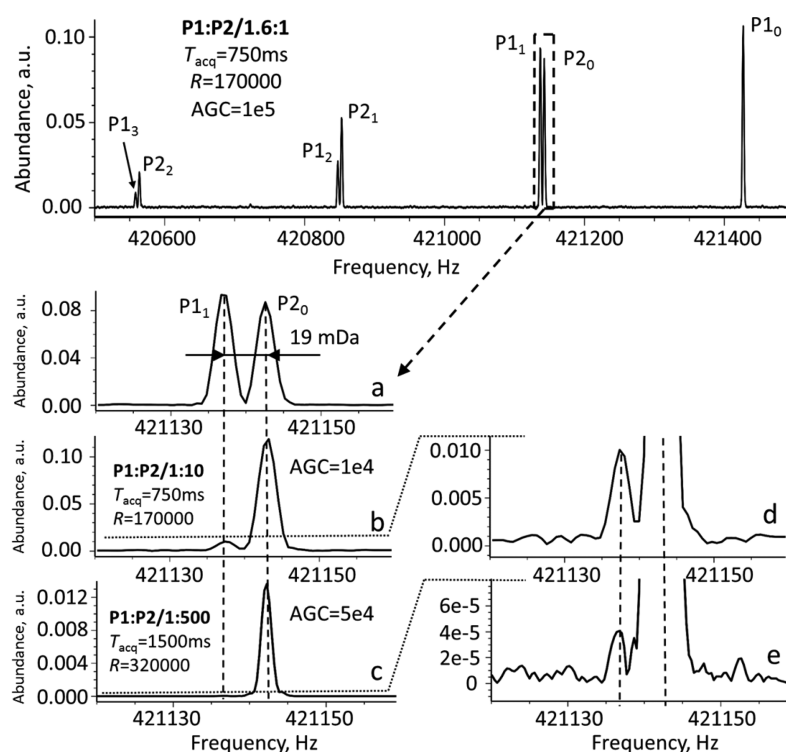


Figure 3. Isobaric peptide mass measurements and spectral dynamic range evaluation at quadruple frequency with the 4X ICR cell on 10 T FT-ICR MS. Peptides: P1, GYQLLEPGDFR; P2, GYEYLLEPGDFR. Other parameters: sum of 10 scans for each measurement; trapping potential 0.11 V; $T_{\text{exc}2} = 60$ ms and $V_{\text{exc}2} = 70$ V.

interactions the measured frequency drifts result in the systematic shifts in the measured masses. These shifts are increasing with the increase in the acquisition time. These results further advocate for the benefits of 4X ICR cell. Indeed, because of the shorter acquisition periods, the cell allows achieving the required resolving power for undamped time-domain transients and, thus, avoiding the above systematic errors.

In a separate analysis we performed internal calibration of the instrument at 4X frequency multiples at each of the acquisition periods, Table 2. We found that internal calibration accuracy of the 4X ICR cell was better than 100 ppb and independent of the acquisition period in a wide m/z range and comparable with one of the standard FT-ICR cells. At short acquisition periods, this accuracy started deteriorating (yet, still well below 1 ppm). The nearest magnetron sidebands become poorly resolved at these acquisition times. Further reduction in the magnetron sidebands should alleviate this problem and allow obtaining mass measurement accuracy with internal calibration in a sub-ppm range for the transients shorter than 100 ms.

Peptide Analysis and Dynamic Range Performance.

Figure 3 shows frequency spectra for a mixture of two singly protonated model peptides, P1 and P2 (see Experimental Methods for details). The difference between peptides is due to deamidation that results in the overlapping of a monoisotopic peak of peptide P2 and ^{13}C isotopic peak of peptide P1 with 19.3 mDa mass difference between them. The resolving power in excess of 300 000 is required to baseline-resolve these peptides at m/z 1459 (corresponds to reduced cyclotron frequency of ~ 105.25 kHz in a 10 T magnetic field). This resolving power was achieved using 4X cell with 1.536 s transients and with abundance ratio between the two peptides

of up to 500 as shown in Figure 3c,e. Note that at the lower peptide concentration ratios of up to 1:10, the isotopic patterns were clearly resolved even for the shorter acquisition time of 768 ms (Figure 3a,b). Indeed, at greater concentration ratios, the signal from the less abundant analyte will be hidden in the shoulder of the larger peak. To avoid this, higher resolution is required.

The other notable observation was that the averaging of only 10 single scans was sufficient to observe the low abundance peaks even for relatively low number of charges loaded into the cell with the AGC settings of $1\text{E}4$ to $5\text{E}4$. In addition, a signal-to-noise ratio (SNR) of 600 was achieved for these two peptides at 1459 m/z with $\text{AGC} = 5\text{E}4$ for a sum of 10 scans using the 4X cell. This SNR value is comparable with the SNR of ~ 700 obtained at 1422 m/z using the standard (Ultra) ICR cell under the same experimental conditions. Moreover, both of these SNR values exceed the minimum threshold SNR value of 100 stated in instrument specifications for this instrument (Ultra ICR cell performance).

To further study the resolving power and dynamic range capabilities of the 4X cell, the fine isotopic structures of the singly, doubly, and triply protonated substance P and isotopic distribution of singly protonated Ultramark compound (1422 m/z) of calibration mixture were analyzed for the isolation mode and broadband measurement at the quadruple frequency and short acquisition times of 3 s, 1.5 s, 768 ms, and 384 ms, respectively (Figures S7 and S8, Supporting Information). These results demonstrated that the 4X cell under evaluation has a spectral dynamic range of at least 3 orders of magnitude and allows obtaining mass spectra of peptide isotope distribution with abundance ratios close to theoretical ones. Finally, peptide structure analysis via tandem mass spectrom-

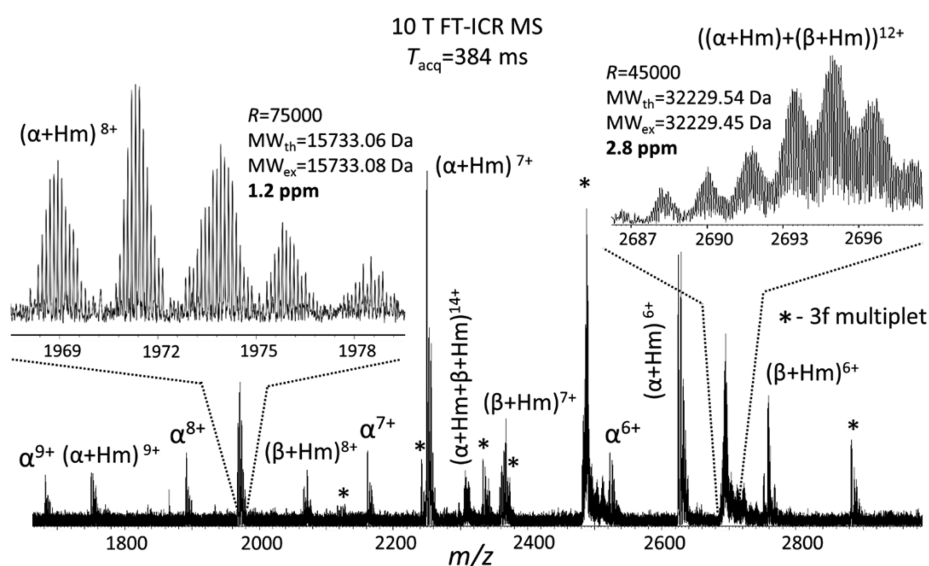


Figure 4. Protein mixture (blood) mass measurements with the 4X ICR cell on 10 T FT-ICR MS. Isotopic-level resolution of hemoglobin monomers and dimers is achieved at quadruple frequency for a single scan with acquisition time of 384 ms. Experimental parameters: 1500–3000 m/z range calibrated at the quadruple frequency; AGC 2E6; trapping potential 0.1 V; $T_{\text{exc1}} = 20$ ms; $V_{\text{exc1}} = 100$ V. Mass spectrum of a similar sample with improved resolution (acquisition time 768 ms, 100 scan averaged) is shown in Figure S11 (Supporting Information). Employed notation: α , reduced hemoglobin subunit alpha (without methionine) (15132 Da); β , reduced hemoglobin subunit beta (without methionine) (15870 Da); Hm, heme B (166 Da); * denotes triple frequency multiple peaks.

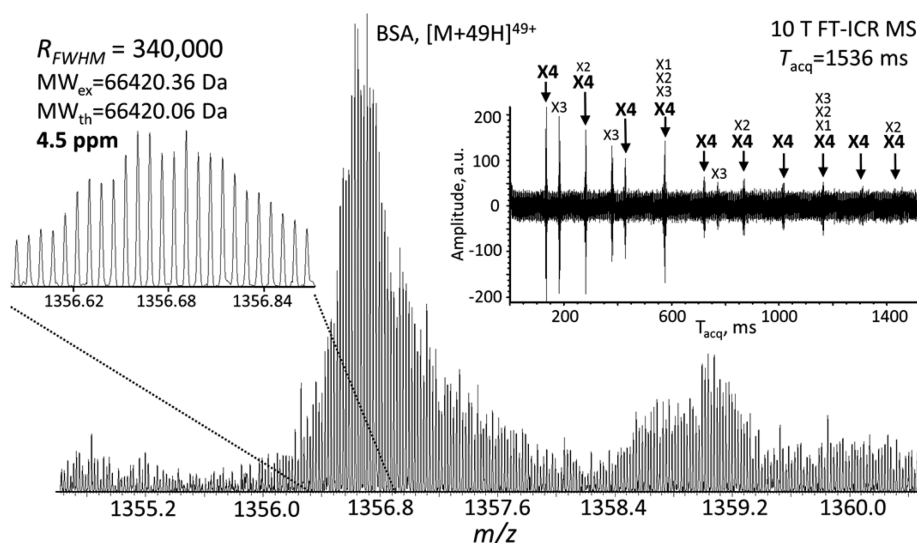


Figure 5. Protein mass measurements with the 4X ICR cell on 10 T FT-ICR MS. Analysis of isolated BSA⁴⁹⁺ cluster at quadruple frequency. Resolution of 340 000 achieved for a single scan with acquisition time of 1.536 s. Left inset shows the baseline resolution of isotopic distribution at the 4f frequency multiple. Detection range includes the corresponding fundamental frequency and its double, triple, and quadruple multiples. Right inset shows the transient of the broadband acquisition with the main beats assigned to the corresponding frequency multiples. The spectrum was obtained using Fourier transform of the whole transient signal including all beats. Other parameters: isolation width 30 m/z ; AGC 2E6; trapping potential 0.1 V, $T_{\text{exc1}} = 20$ ms; $V_{\text{exc1}} = 100$ V.

etry was successfully performed using electron capture dissociation (ECD) of substance P (Figure S9, Supporting Information). The ECD efficiency was comparable to the typical ECD performance on the same instrument with Ultra ICR cell.⁵⁰

Intact Protein Analysis. A typical broadband mass spectrum of ~ 17 kDa horse myoglobin acquired in the mass range of m/z 700 to 1400 shows a baseline ¹³C resolved charge state at the quadruple frequency with acquisition time as short

as 192 ms (Figure S10, Supporting Information). Note that, similar to the above examples with peptides, the resolving power obtained for multiply charged myoglobin ions was four times higher than the “low pressure” theoretical estimate corresponding to 10 T FT-ICR MS. The mass measurement accuracy obtained for the mass spectra shown in Figure S10 (Supporting Information) was better than 2.9 ppm for all myoglobin charge states.

Figure 4 shows a broadband single-scan mass spectrum of an unprocessed and undepleted blood sample measured for the mass range of m/z 1500 to 3000. Expectedly, the major components of the sample contain primarily monomers and dimers of hemoglobin subunit alpha and beta and their complexes with heme B. Here, all these components were identified with mass accuracy of better than 3 ppm (external calibration). Importantly, proteins with molecular weight of 30 kDa were ^{13}C isotope-resolved at quadruple frequency with acquisition time as short as 384 ms. The broadband mass spectrum demonstrates that a majority of m/z range in question is devoid of parasitic peaks at the other frequency multiples. Increasing acquisition times leads to correspondingly improved resolution performance on this complex sample (Figure S11, Supporting Information).

Finally, a mass spectrum of the isolated 49+ charge state bovine serum albumin (66 kDa) at m/z 1356 is shown in Figure 5. The average resolving power across all baseline-resolved ^{13}C isotopic pattern peaks in the magnitude-mode apodized mass spectrum was 340 000 (see Figure 5 left inset) for a single scan with an acquisition period of only 1.536 s. The typical isotopic beats in a transient signal show the difference between measurements at first and quadruple frequencies, see Figure 5 right inset. As expected, the difference is two beats at the first frequency against ten beats recorded at the quadruple frequency. These results compare favorably with the high electric field Orbitrap Elite mass spectrometer employed in the eFT mode. A resolving power of 125 000 was reported for the 40+ charge state BSA ions and 0.768 s transients using that instrument.⁵⁷

CONCLUSIONS

We described the implementation of the 4X ICR cell, a multielectrode cell designed for signal detection at quadruple frequency multiples, and investigated its potential for high-throughput molecular analysis with FT-ICR MS. The levels of resolving power achieved with the 4X cell installed on a serial 10 T FT-ICR MS are comparable with those of the high-field Orbitrap FTMS with eFT signal processing. Under the no-coalescence conditions, the achieved resolving powers correspond to resolution performance that could be provided with a virtual 40 T FT-ICR MS utilizing a conventional measurement principle with dipolar signal detection and FT signal processing. Other analytical characteristics of the 4X cell, including mass accuracy and dynamic range, correspond to the performance levels of a standard 10 T FT-ICR MS.

The obtained results validate the advantages of frequency multiple detection for improved analysis of peptides and proteins, specifically benefiting the high-resolution and throughput-demanding potential applications in middle-down and top-down proteomics. Implementation of absorption mode FT spectral representation for 4X cell-based FT-ICR MS should further elevate its performance. An increase in frequency multiple detection, up to the eighths frequency multiple, is possible with the same 4X cell configuration. However, improved signal processing and data analysis methods are needed for analysis of spectra with overlapping harmonics and frequency multiples. Luckily, certain progress in this area has recently been achieved.^{54,55}

ASSOCIATED CONTENT

Supporting Information

Additional information as noted in the text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: yury.tsybin@epfl.ch.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are grateful to Unige A. Laskay and Luca Fornelli for technical assistance and samples. We express our gratitude to Thermo Fisher Scientific Inc. for providing access under license to selected instrument control parameters. We appreciate the financial support from the Swiss National Science Foundation (SNF projects 200021-147006 and 128357) and the European Research Council (ERC Starting Grant 280271 to Y.O.T.). M.V.G. thanks the European Union Seventh Framework Program (project Prot-HiSPRA #282506) and Russian Science Foundation (grant no. 14-14-00971) for support.

REFERENCES

- (1) Liu, T.; Belov, M. E.; Jaitly, N.; Qian, W.-J.; Smith, R. D. *Chem. Rev.* **2007**, *107*, 3621–3653.
- (2) Mann, M.; Kelleher, N. L. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 18132–18138.
- (3) Hsu, C.-S.; Hendrickson, C. L.; Rodgers, R. P.; McKenna, A. M.; Marshall, A. G. *J. Mass Spectrom.* **2011**, *46*, 337–343.
- (4) Zhurov, K. O.; Kozhinov, A. N.; Tsybin, Y. O. *Energy Fuels* **2013**, *27*, 2974–2983.
- (5) Marshall, A. G.; Hendrickson, C. L. *Annu. Rev. Anal. Chem.* **2008**, *1*, 579–599.
- (6) Xian, F.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2012**, *84*, 708–719.
- (7) Schaub, T. M.; Hendrickson, C. L.; Horning, S.; Quinn, J. P.; Senko, M. W.; Marshall, A. G. *Anal. Chem.* **2008**, *80*, 3985–3990.
- (8) Scigelova, M.; Hornshaw, M.; Giannakopoulos, A.; Makarov, A. *Mol. Cell. Proteomics* **2011**, *10*, doi: 10.1074/mcp.M111.009431.
- (9) Zubarev, R. A.; Makarov, A. *Anal. Chem.* **2013**, *85*, 5288–5296.
- (10) Siek, K.; Binkley, J.; Patrick, J. S. *LCGC Chromatogr. Online* **2013**, 30–37.
- (11) Gorshkov, M. V.; Fornelli, L.; Tsybin, Y. O. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 1711–1717.
- (12) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1–35.
- (13) Marshall, A. G.; Guan, S. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1819–1823.
- (14) Karabacak, N. M.; Easterling, M.; Agar, N. R.; Agar, J. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1218–1222.
- (15) Bruce, J. E.; Anderson, G. A.; Lin, C.-Y.; Gorshkov, M.; Rockwood, A. L.; Smith, R. D. *J. Mass Spectrom.* **2000**, *35*, 85–94.
- (16) Kim, S.; Choi, M. C.; Hur, M.; Kim, H. S.; Yoo, J. S.; Hendrickson, C. L.; Marshall, A. G. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 1423–1429.
- (17) Brustkern, A. M.; Rempel, D. L.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1281–1285.
- (18) Weisbrod, C. R.; Kaiser, N. K.; Skulason, G. E.; Bruce, J. E. *Anal. Chem.* **2008**, *80*, 6545–6553.
- (19) Tolmachev, A. V.; Robinson, E. W.; Wu, S.; Kang, H.; Lourette, N. M.; Paša-Tolić, L.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 586–597.
- (20) Nikolaev, E. N.; Boldin, I. A.; Jertz, R.; Baykut, G. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1125–1133.

- (21) Brustkern, A. M.; Rempel, D. L.; Gross, M. L. *Int. J. Mass Spectrom.* **2011**, *300*, 143–148.
- (22) Hartmann, H.; Chung, K. M.; Baykut, G.; Wanczek, K. P. *J. Chem. Phys.* **1983**, *78*, 424–431.
- (23) Gorshkov, M. V.; Nikolaev, E. N. *Int. J. Mass Spectrom. Ion Processes* **1993**, *125*, 1–8.
- (24) Mitchell, D. W. *Int. J. Mass Spectrom. Ion Processes* **1995**, *142*, 1–22.
- (25) Kaiser, N. K.; Weisbrod, C. R.; Webb, B. N.; Bruce, J. E. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 467–478.
- (26) Lee, J. P.; Comisarow, M. B. *Appl. Spectrosc.* **1989**, *43*, 599–604.
- (27) Qi, Y.; Barrow, M. P.; Li, H.; Meier, J. E.; Van Orden, S. L.; Thompson, C. J.; O'Connor, P. B. *Anal. Chem.* **2012**, *84*, 2923–2929.
- (28) Lange, O.; Damoc, E.; Wieghaus, A.; Makarov, A. Presented at the 59th Conference of American Society for Mass Spectrometry, Denver, CO, June 5–9, 2011.
- (29) Lange, O. Methods and Apparatus for Producing a Mass Spectrum. US Patent 20110240841, 2011.
- (30) Lange, O.; Damoc, E.; Wieghaus, A.; Makarov, A. *Int. J. Mass Spectrom.* **2014**, *369*, 16–22.
- (31) Kilgour, D. P. A.; Wills, R.; Qi, Y. L.; O'Connor, P. B. *Anal. Chem.* **2013**, *85*, 3903–3911.
- (32) Hilger, R. T.; Wyss, P. J.; Santini, R. E.; McLuckey, S. A. *Anal. Chem.* **2013**, *85*, 8075–8079.
- (33) Tsybin, Y. O. *Chimia* **2014**, *68*, 168–174.
- (34) Aizikov, K.; O'Connor, P. B. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 836–843.
- (35) Kozhinov, A. N.; Tsybin, Y. O. *Anal. Chem.* **2012**, *84*, 2850–2856.
- (36) Miladinovic, S. M.; Kozhinov, A. N.; Gorshkov, M. V.; Tsybin, Y. O. *Anal. Chem.* **2012**, *84*, 4042–51.
- (37) Aushev, T.; Kozhinov, A. N.; Tsybin, Y. O. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1263–1273.
- (38) Pan, Y.; Ridge, D. P.; Wronka, J.; Rockwood, A. L.; Marshall, A. G. *Rapid Commun. Mass Spectrom.* **1987**, *1*, 120–121.
- (39) Pan, Y.; Ridge, D. P.; Rockwood, A. L. *Int. J. Mass Spectrom. Ion Processes* **1988**, *84*, 293–304.
- (40) Nikolaev, E. N.; Gorshkov, M. V.; Mordehai, A. V.; Talrose, V. L. *Rapid Commun. Mass Spectrom.* **1990**, *4*, 144–146.
- (41) Misharin, A. S.; Zubarev, R. A. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 3223–3228.
- (42) Knobler, M.; Wanczek, K. P. Proceedings of the 45th ASMS Conference on Mass Spectrometry and Allied Topics, Palm Springs, CA, June 1–5, 1997; American Society for Mass Spectrometry: Santa Fe, NM.
- (43) Vorobyev, A.; Gorshkov, M. V.; Tsybin, Y. O. *Int. J. Mass Spectrom.* **2011**, *306*, 227–231.
- (44) Grosshans, P. B.; Marshall, A. G. *Int. J. Mass Spectrom. Ion Processes* **1991**, *107*, 49–81.
- (45) Guan, S.; Wahl, M. C.; Marshall, A. G. *Anal. Chem.* **1993**, *65*, 3647–3653.
- (46) Peurrung, A. J.; Kouzes, R. T. *Phys. Rev. E* **1994**, *49*, 4362–4368.
- (47) Popov, I. A.; Nagornov, K.; Vladimirov, G.; Kostyukevich, Y. I.; Nikolaev, E. N. *J. Am. Soc. Mass Spectrom.* **2014**, 1–10.
- (48) Nefedov, A. V.; Gilski, M. J.; Sadygov, R. G. *Curr. Proteomics* **2011**, *8*, 125–137.
- (49) Hebert, A. S.; Richards, A. L.; Bailey, D. J.; Ulbrich, A.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. *Mol. Cell. Proteomics* **2014**, *13*, 339–347.
- (50) Ben Hamidane, H.; Vorobyev, A.; Tsybin, Y. O. *Eur. J. Mass Spectrom.* **2011**, *17*, 321–331.
- (51) Ledford, E. B.; Rempel, D. L.; Gross, M. L. *Anal. Chem.* **1984**, *56*, 2744–2748.
- (52) Shi, S. D. H.; Drader, J. J.; Freitas, M. A.; Hendrickson, C. L.; Marshall, A. G. *Int. J. Mass Spectrom.* **2000**, *195–196*, S91–S98.
- (53) Kozhinov, A. N.; Zhurov, K. O.; Tsybin, Y. O. *Anal. Chem.* **2013**, *85*, 6437–6445.
- (54) Misharin, A. S.; Novoselov, K.; Doroshenko, V. M. Spectral deconvolution in ion cyclotron resonance mass spectrometry. US 20110251801 A1, 2011.
- (55) Ding, L.; Badheka, R. A method of processing image charge/current signals. EP 20130159402, 2013.
- (56) Miladinović, S. M.; Kozhinov, A. N.; Tsybin, O. Y.; Tsybin, Y. O. *Int. J. Mass Spectrom.* **2012**, *325–327*, 10–18.
- (57) Damoc, N.; Denisov, E.; Lange, O.; Moehring, T.; Makarov, A. Proceedings of the 59th ASMS conference on mass spectrometry and allied topics, Denver, CO, June 5–9, 2011; American Society for Mass Spectrometry: Santa Fe, NM.

Supplementary Information

High-Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry with Increased Throughput for Biomolecular Analysis

Konstantin O. Nagornov,¹ Mikhail V. Gorshkov,^{2,3} Anton N. Kozhinov,¹ and Yury O. Tsybin^{1,*}

¹Biomolecular Mass Spectrometry Laboratory, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

²Institute for Energy Problems of Chemical Physics, Russian Academy of Sciences, 119334 Moscow, Russia

³Moscow Institute of Physics and Technology (State University), 141707 Dolgoprudny, Moscow region, Russia

Experimental methods details.

Three ranges of dipolar broadband frequency-sweep excitation (varied from 72 kHz to 2150 kHz) were employed to excite ion cyclotron motion: first excitation frequency range of 72 – 550 kHz (m/z range 280-2000); second range of 72 – 1050 kHz (m/z range 145-2000) and third range of 72 – 2150 kHz (m/z range 70 – 2000). The amplitude, $V_{\text{exc}(n)}$, and duration, $T_{\text{exc}(n)}$, of the excitation voltage were varied in the range from 0 to 100 V and from 0 to 60 ms, respectively, where, n – number of excitation frequency range. Preliminary studies demonstrated that to excite the same m/z ions to the same post-excitation radius by using wider excitation frequency range the higher excitation energy is required. For example, ($T_{\text{exc1}} = 60$ ms, $V_{\text{exc1}} = 35$ V) \approx ($T_{\text{exc2}} = 60$ ms, $V_{\text{exc2}} = 70$ V) \approx ($T_{\text{exc3}} = 60$ ms, $V_{\text{exc3}} = 100$ V). Also, ($T_{\text{exc1}} = 60$ ms, $V_{\text{exc1}} = 35$ V) \approx ($T_{\text{exc1}} = 20$ ms, $V_{\text{exc1}} = 100$ V). In addition, the different operational conditions (during ion injection and excitation events) was the reason of the difference between excitation energy required for MS/MS (see Figure S9) and typical MS experiments (see Figures 1, S8, and S10).

Figure S1. Schematics of the 4X ICR cell aligned along magnetic field lines: (left) perspective schematic and (right) cross section with a wiring diagram. RF excitation is provided through two pairs of four electrodes each. Detection is organized through other two pairs of four electrodes each. Frequency $f = \omega/2\pi$ of the cyclotron mode of ion motion yields an ion signal at the quadruple frequency multiple, $4f$. The offset DC potentials applied to the excitation electrodes of the 4X cell were adjusted to increase the signal-to-noise ratio (SNR) and reduce magnetron sidebands. The offset DC potentials applied to detection electrodes did not significantly affect signal strength nor the magnetron sidebands due to their symmetrical placement in the 4X cell and offset potentials wiring (offset potentials applied to detection 1 and 2 electrode sets, but not to

each detection electrode separately) in the present configuration of the 4X cell, Figure S1. Contrary to the detection electrodes, the two sets of excitation electrodes (excitation 1 and 2) are positioned in front of each other and independent application of the offset potentials to the sets of excitation electrodes influences ion motion to a larger extent.

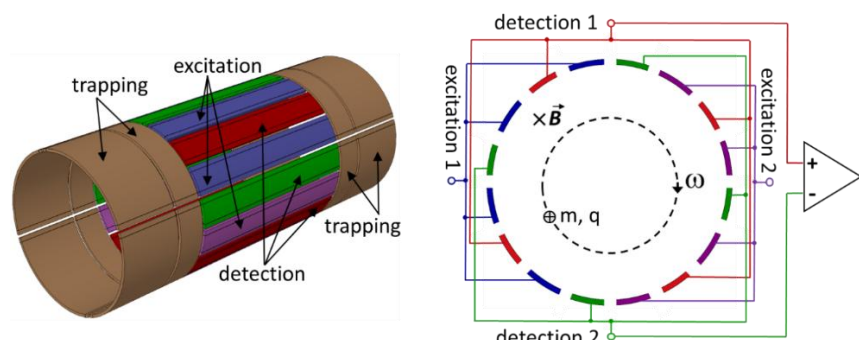


Figure S2. 4X cell 10 T FT-ICR MS time domain signals (transients) of (top) the isolated singly protonated Ultramark component at m/z 1422 and (bottom) the calibration mixture containing MRFA and Ultramark components. Trapping potential 0.1 V.

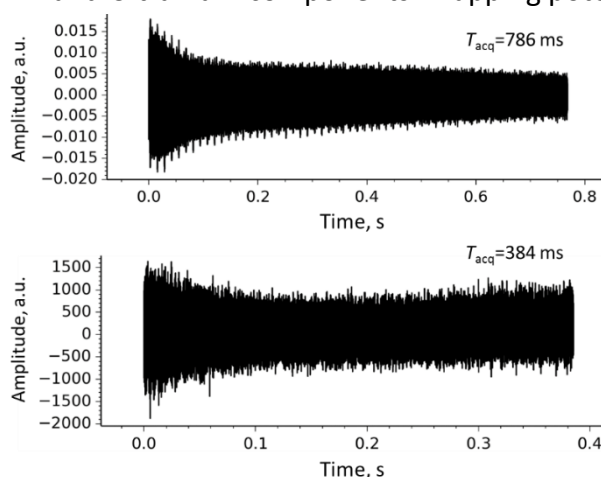


Figure S3. Broadband mass spectra of Ultramark with standard (Ultra) ICR cell, all peaks corresponding to single (f) frequency multiple (top) and 4X cell, peaks corresponding to quadruple ($4f$) frequency multiple (in red, bottom,) were obtained under the same experimental conditions: acquisition period of 384 ms; AGC=2E6; $T_{exc1} = 20$ ms and $V_{exc1} = 100$ V.

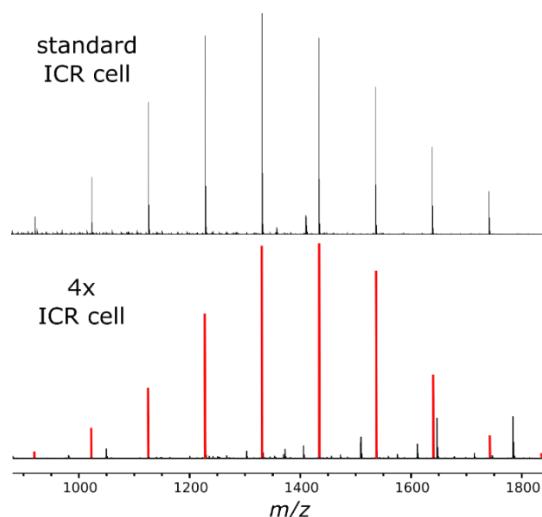
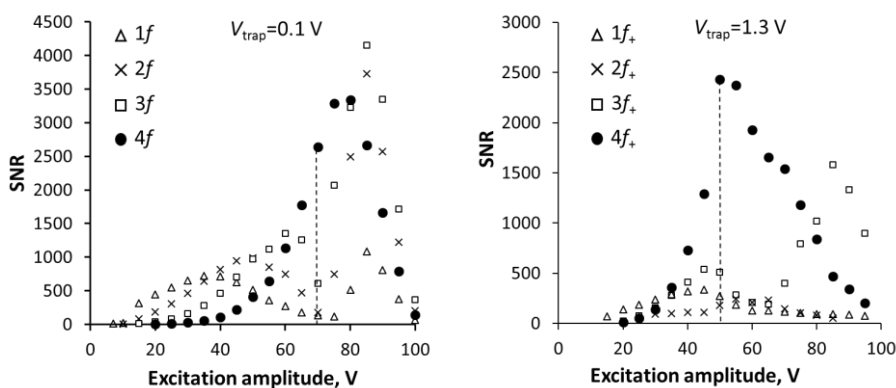


Figure S4. Characteristics of 10 T FT-ICR MS with 4X cell of the isolated doubly protonated substance P (674 m/z). Depending on the excitation amplitude the intensities of peaks at f , $2f$ and $3f$ frequency multiples can be reduced to a level of less than 20% of the main $4f$ component (top left) for low trapping potential of 0.1 V. Here, detection frequency range includes the corresponding single, double, triple, and quadruple frequency multiples. Expanded views of all frequency multiples show spectral composition around the monoisotopic peak (bottom left). Expanded views from spectrum acquired with higher trapping potential of 1.3 V, as expected⁴⁵, due to higher magnetron frequency ω_- , reveal that the spectral composition contains peak f_+ corresponding to the reduced cyclotron frequency ω_+ at the corresponding frequencies and their interharmonics (magnetron sidebands: $nf_+ \pm mf_-$, $n=1,2,\dots$, $m=1,2,\dots$, where f_- corresponds to the magnetron frequency ω_- , bottom right).^{45, 54} Depending on the excitation amplitude the intensities of peaks corresponding to the reduced cyclotron frequencies f_+ , $2f_+$, and $3f_+$ can be reduced to a level of less than 5% of the main $4f_+$ component (top right). The intensities of the corresponding interharmonics can be reduced to the lowest possible level by optimizing the cell's DC offset potentials. In the current instrument configuration, the adjustment of ICR cell offset DC potentials allows efficient reduction of the interharmonics. However, due to the specifics of the current 4X cell design we were limited to one degree of freedom in controlling the ion cloud position via variation of electrode offset potentials (electrode sets E1 and E2, see Figure S1, right panel). Experimental results show that the intensity of double reduced cyclotron frequency multiple, $2f_+$, is less than 5% w.r.t. the $4f_+$ peak for the excitation amplitude of 50 V, whereas its interharmonics reach 30% intensity of the $4f_+$ peak (bottom right). All experimental data were obtained with isolated doubly protonated substance P, m/z 674.4; acquisition period 384 ms and $T_{\text{exc}2} = 60$ ms.



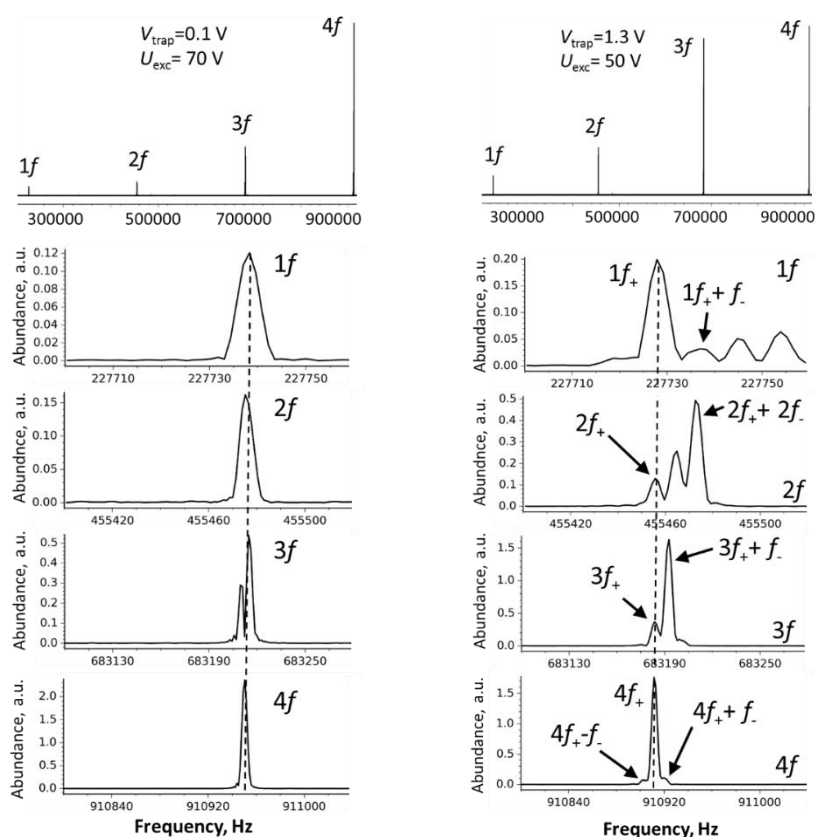


Figure S5. Variation of reduced cyclotron frequencies, f_+ , of the (top) quadruple and (bottom) first frequency multiples of isolated doubly protonated substance P (m/z 674) with the excitation amplitude. Experimental parameters: 4X cell in 10 T FT-ICR MS; acquisition period 384 ms; AGC=3E5; trapping potential 1 V; $T_{\text{exc}2} = 60$ ms. Variation of excitation voltage duration was achieved through the use of advanced user interface, the added-on utility kindly provided by Thermo Scientific.

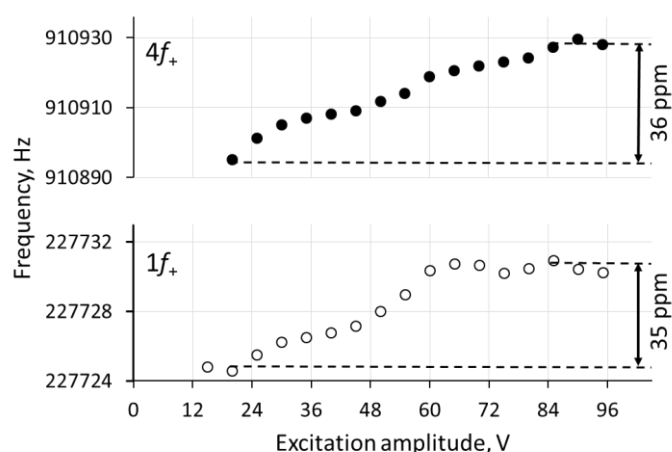


Figure S6. Characterization of mass accuracy performance of the 4X cell in 10 T FT-ICR MS. (Top) broadband mass spectrum of phosphoric acid. (Bottom) dependence of mass accuracy for broadband mass measurements of phosphoric acid H_3PO_4 with external calibration as a function of (left) number of charges and (right) acquisition period ($T_{\text{exc}1} = 20$ ms and $V_{\text{exc}1} = 100$ V). Numerical values are shown in Tables 1 and S1.

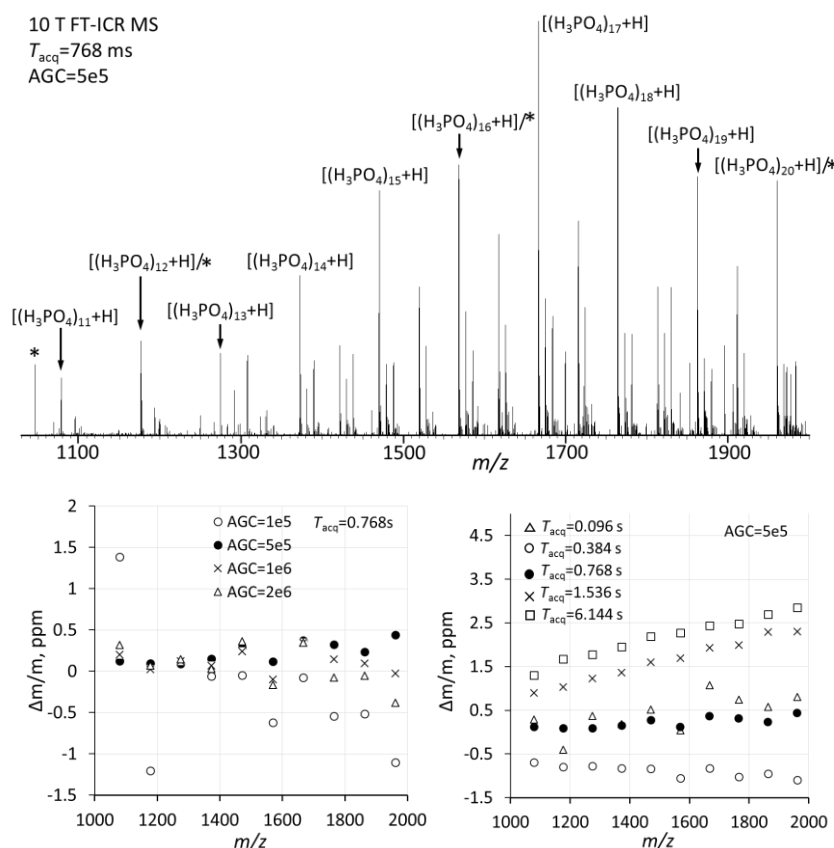


Figure S7. Characterization of dynamic range performance of the 4X cell in 10 T FT-ICR MS. Mass spectrum of isolated singly protonated substance P (1348 m/z) calibrated at the quadruple frequency for a sum of 30 scans with transient duration of 3 s each (top). Broadband frequency spectrum of a calibration mixture (see Figure 1) (bottom). Insets in $^{13}\text{C}_5$ isotopologue of substance P (top) and $^{13}\text{C}_3$ isotopologue of Ultramark component (1422 m/z) (bottom) demonstrate the dynamic range of 3 orders at least both for isolation mode and for broadband mode measurements. Isotope distribution envelope for the corresponding peptides was calculated theoretically and is indicated in light blue color bars.

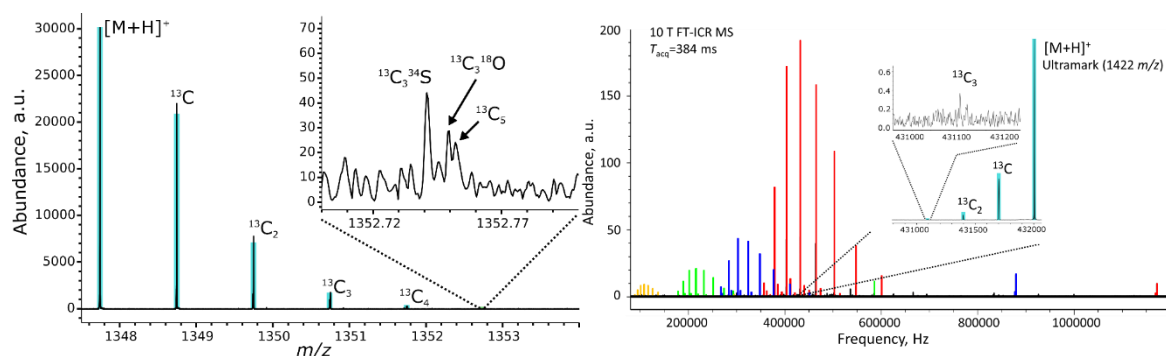


Figure S8. High-resolution performance of 4X cell 10 T FT-ICR MS at quadruple frequency of isolated (top) triply, (middle) doubly (middle), and (bottom) singly protonated substance P. The corresponding mass spectra are shown at the left and the transients at the right. The fine isotopic structure distributions are shown in the insets. Experimental parameters: trapping potential 0.1 V; AGC=3E5; $T_{\text{exc}2} = 60$ ms and $V_{\text{exc}2} = 70$ V for singly

and doubly protonated ions; $T_{\text{exc}3} = 60$ ms and $V_{\text{exc}3} = 95$ V for triply protonated ions. The purpose of this study was to show that the 4X cell does not induce any additional factors that may result in the loss of resolving power capabilities in spite of the increase in the individual peak resolution.

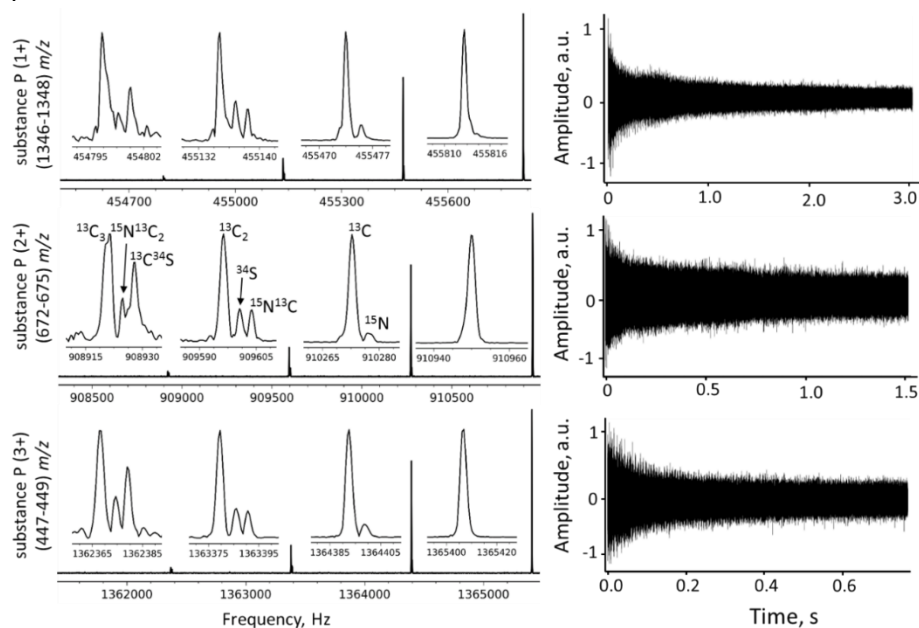


Figure S9. 4X cell 10 T FT-ICR MS/MS: ECD mass spectrum of doubly protonated substance P (m/z 674). Although readily feasible, distinguishing peak overlaps between the 4X peaks of the product ions and the non-4X peaks (different frequency multiples) requires high-resolution performance. Experimental parameters: m/z 320-1500 externally calibrated at the quadruple frequency multiple, acquisition period 384 ms; AGC=5E5; trapping potential 0.11 V; $T_{\text{exc}3} = 60$ ms and $V_{\text{exc}3} = 90$ V; 200 scans averaged; ECD parameters: electron energy 1 eV, electron injection duration 200ms.

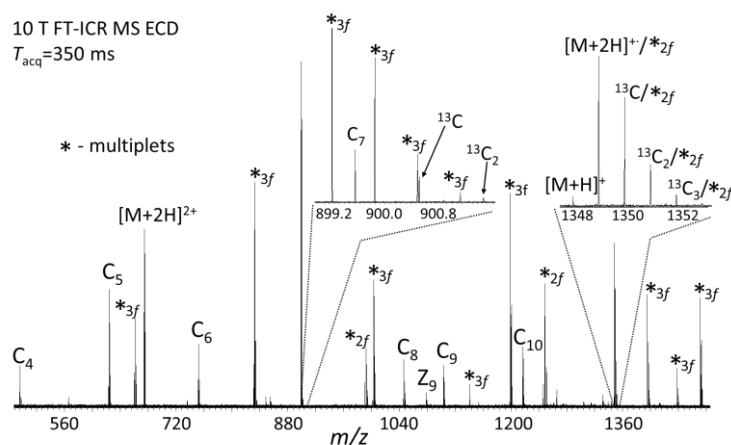


Figure S10. 4X cell 10 T FT-ICR MS broadband mass spectrum of myoglobin, ~17 kDa. Experimental parameters: acquisition period 192 ms; m/z 700 - 1500 range calibrated at the quadruple frequency; AGC=1E5; trapping potential 0.1 V; $T_{\text{exc}3} = 60$ ms and $V_{\text{exc}3} = 95$ V.

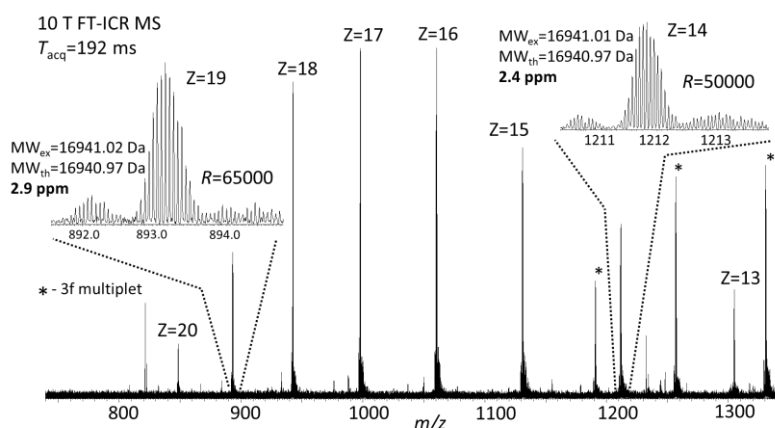


Figure S11. Protein high-resolution mass measurements with 4X cell 10 T FT-ICR MS. Analysis of blood sample (acquisition period 768 ms; 100 scans averaged). Insets show resolved isotopic distributions for all proteins detected. Nomenclature: α – reduced hemoglobin subunit alpha (without methionine), 15132 Da; β – reduced hemoglobin subunit beta (without methionine), 15870 Da; Hm – Heme B, 166 Da; * - denotes triple frequency multiple peaks.

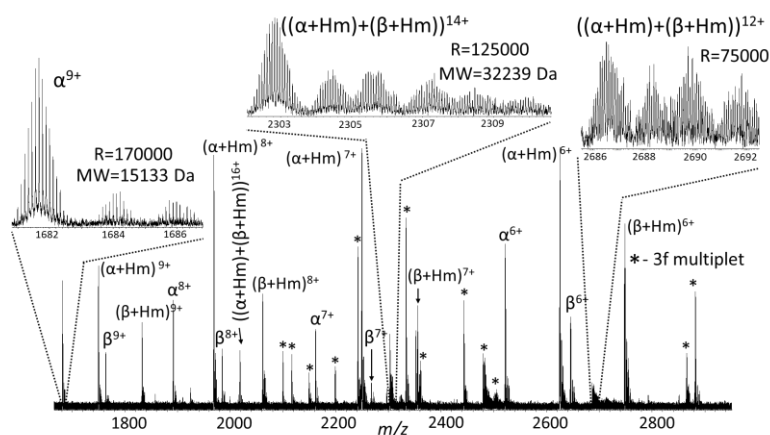


Table S1. Mass accuracy of phosphoric acid H_3PO_4 measurements as a function of acquisition period. External calibration performed on 30 summed transients with 5E5 total number of charges (AGC) accumulated for each measurement. The corresponding mass spectrum and graphic of dependence are shown in Figure S6 (Supplementary Information).

Compound	Mass accuracy ($\Delta m/m$, ppm)			
	$T_{\text{acq}}=96$ ms	$T_{\text{acq}}=768$ ms	$T_{\text{acq}}=1562$ ms	$T_{\text{acq}}=6144$ ms
$(\text{H}_3\text{PO}_4)_{11}\text{H}^+$	0.29	0.13	0.90	1.30
$(\text{H}_3\text{PO}_4)_{13}\text{H}^+$	0.37	0.09	1.23	1.78
$(\text{H}_3\text{PO}_4)_{16}\text{H}^+$	0.04	0.12	1.69	2.28
$(\text{H}_3\text{PO}_4)_{18}\text{H}^+$	0.74	0.32	1.99	2.48
$(\text{H}_3\text{PO}_4)_{20}\text{H}^+$	0.80	0.44	2.30	2.85

A.8 Paper VIII

Paradigm shift in ion trap design and capabilities for Fourier transform ion cyclotron resonance mass spectrometry

Konstantin O. Nagornov, Anton N. Kozhinov, and Yury O. Tsybin*

Biomolecular Mass Spectrometry Laboratory, Ecole Polytechnique Fédérale
de Lausanne, 1015 Lausanne, Switzerland

* Correspondence should be addressed to Prof. Yury O. Tsybin, EPFL ISIC
LSMB, BCH 4307, 1015 Lausanne, Switzerland. E-mail: yury.tsybin@epfl.ch

Running title: ICR cell with narrow aperture detection electrodes for FTMS

Submitted to: *JASMS*

Current manuscript date:

21 August 2014

Keywords: Fourier transform; FT; Fourier transform mass spectrometry;
FTMS; ion cyclotron resonance; ICR; transient signal; ion trap; proteomics;

Abstract

The current paradigm in ion trap (cell) design for Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) is the ion detection with wide aperture detection electrodes. Specifically, excitation and detection electrodes are typically 90 degrees wide and positioned radially at a similar distance from the ICR cell axis. Here, we demonstrate that ion detection with narrow aperture detection electrodes (NADEL) positioned radially inward the cell's axis is feasible and advantageous for FT-ICR MS. We describe design details and performance characteristics of a 10 T FT-ICR MS equipped with a NADEL ICR cell having a pair of narrow aperture (flat) detection electrodes and a pair of standard 90 degrees excite electrodes. Despite a smaller surface area of detection electrodes, the sensitivity of the NADEL ICR cell is not reduced thanks to improved excitation field distribution, detection electrodes reduced capacitance and closer detector positioning to the orbits of the excited ions. The performance characteristics of the NADEL ICR cell are comparable with the state-of-the-art FT-ICR MS implementations for small molecule, peptide, protein, and petroleomics analyses. In addition, the NADEL ICR cell design improves the flexibility of ICR cells and allows for potential implementation of advanced capabilities, e.g., quadrupolar or quadrature ion excitation as well as quadrupolar ion detection for improved mainstream applications. It also creates an intriguing opportunity for addressing the major bottleneck in FTMS – increasing its throughput via implementation of a parallel multi-transient acquisition approach or via transient shape transformation offering high order harmonics content of mass spectra.

Introduction

Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) provides the highest resolving power and mass accuracy for molecular structural analysis [1-6]. Top-down proteomics and petroleomics, due to their extreme requirements for analytical performance, are the primary application areas of FT-ICR MS nowadays [3, 7]. Over the past decade, other application areas of interest became primarily addressed using Orbitrap FTMS and high-resolution time-of-flight (TOF) MS [4, 8-11]. Nevertheless, the extreme analytical challenges in molecular structural analysis require further increase of FTMS performance. Due to fundamental power of FT-ICR MS, further advancing its performance is feasible and may allow bridging the gap between the analytical requirements and MS capabilities. For that, a corresponding increase in high-performance analysis throughput is required. For example, even at the current data acquisition rate the exceptional FT-ICR MS capabilities trigger development of novel approaches of molecular structural analysis [12]. Among those, the uses of isotopic fine structure information in qualitative bottom-up proteomics and mass defect information in quantitative bottom-up proteomics appear particularly attractive [13-15]. Increased, for example 2...10-fold, data acquisition rate should allow interrogation of more peptides at the required structural level and provide improved qualitative and quantitative protein analysis. Development of novel concepts for ion conditioning and manipulation in ICR cells based on improved understanding of ion motion fundamentals has thus become a target of recent innovations in the field [6, 16].

Since the inception of FT-ICR MS by Marshall and Comisarow, the design of ICR cells has been primarily based on the use of wide aperture detection electrodes and narrowband (relatively low sampling frequency range, 1-5 MHz sampling rate) ion detection [12, 17]. Since at least a decade, the most common ICR cell configuration is a cylindrical cell with a pair of 90 degrees wide excitation electrodes and a pair of similar 90 degrees wide detection electrodes [4, 18]. The detected image charge ion signal is averaged during ion motion along the detection electrode surface, requiring low frequency (as determined from Nyquist criteria) sampling. The drawbacks of this configuration include a certain limitation on the acquisition rate, substantial perturbation of excitation radiofrequency (RF) field by detection electrodes, and a low degree of design flexibility. The latter aspect, for instance, requires the use of external electronic switching when the same pair of wide aperture electrodes is to be employed for both excitation and detection [19]. For example, Marshall and co-workers have recently demonstrated by simulations that 120 degrees wide excitation electrodes are optimal for coherent ion dipolar excitation and elimination of the third harmonic [20]. However, implementation of such electrodes for ion excitation would reduce the efficiency of ion detection using the currently employed ICR cell design concept. A feasible, although technically challenging, solution to this problem is the use of external switching electronic circuitry allowing for ion excitation and detection on the same pair of 120 degrees wide electrodes. Quadrupolar and quadrature modes of ion excitation are other attractive approaches for improving the coherence of excited ion motion [21-24]. ICR cell operation in these modes also suffers from the same low

flexibility of the current ICR cell design, when wide aperture detection electrodes compete with the excite electrodes for space.

Improved coherence of ion motion in ICR cells has been efficiently addressed by harmonizing the trapping electric fields by a number of groups, including those of Tolmachev [25-26], Gross [27], Bruce [28-29], and Nikolaev [30-31]. The obtained improvements in FT-ICR MS performance are quite dramatic. Specifically, the transient lifetime has been substantially, about 10-100 fold, increased. The dynamically harmonized ICR cell, following the original idea of Boldin and Nikolaev [30], is particularly impressive, with demonstrated resolution of up to 40 million on a peptide (single peak) and a routinely achieved fine isotopic structure resolution on peptide ions even in moderate, 7 T, magnetic fields [32]. The main drawback of the harmonized cells is that they do not improve the throughput (or the acquisition rate) of FT-ICR MS.

Detection at multiples of ion (reduced) cyclotron frequency, which is a feasible approach for increasing the throughput [33-37], is also limited by the design of these cells which still employ relatively wide (e. g., ~22 degrees in 16-electrode ICR cell) aperture excite and detect electrodes. A number of attempts have been made to reduce the aperture of detection electrodes. For example, Nikolaev and co-workers suggested an ICR cell with four detection electrodes made of conducting wires [33]. However, the experimental implementation of the suggested ICR cell with detect conducting wires was deferred. One of the possible reasons for the failure was the positioning of the wire detect electrodes directly below the excite electrodes in the areas

with strong excite RF field, leading to excite field distortion and pre-amplifier saturation prior to ion detection.

Here, we describe a concept, design, and implementation of ICR cells with narrow aperture (flat) detection electrodes (NADEL) enabling both narrowband and broadband ion detection and addressing the limitations of FT-ICR MS design and capabilities specified above. The analytical characteristics of thus introduced NADEL ICR cells are presented and their particular advantages are outlined.

Experimental methods

Sample preparation. Substance P, MRFA, Ultramark, and ubiquitin were purchased from Sigma-Aldrich (Buchs, Switzerland). ESI calibration mixture pre-formulated at low concentration was purchased from Agilent (Basel, Switzerland). Serum albumin (bovine) was purchased from Fisher Scientific (Schwerte, Germany). Polyphosphoric acid was obtained from Fluka (Buchs, Switzerland). Peptides P1, P2, and P3: P1 – EESTR (619.29255 Da), P2 – PMMSR (619.29342 Da), and P3 – HDGHR (619.29389 Da) were synthesized in-house with solid state peptide synthesis. All peptide and protein samples were used without further purification. LC-MS grade acetonitrile and water were obtained from Fluka (Buchs, Switzerland). Formic acid was obtained from Merck (Zug, Switzerland). Peptide and protein solutions were prepared in 1:1 (v/v) water/acetonitrile solvent mixtures containing 0.1% (v/v) of formic acid. The final sample concentration for all samples was ~1 μ M, except ~10 μ M for albumin. The polyphosphoric calibration mixture was prepared by adding 1% (v/v) of polyphosphoric acid to 1 mL of water.

Maltene fraction of Venezuelan crude oil for petroleomics-type measurements was dissolved in dichloromethane, followed by 1:23 dilution in acetonitrile with the addition of 2% of formic acid (by volume), as described previously [9].

NADEL ICR cell design and mass spectrometry. The design of the employed here ICR cell with narrow aperture detection electrodes (NADEL) was developed based on the open-ended cylindrical ICR cell (Ultra Cell, Thermo Scientific, Bremen, Germany). The commercially available Ultra cell contains two pairs, outer and inner, of segmented trapping ring electrodes and excitation grids, placed over the entire length of the cell, in addition to the conventional 90° excitation and detection electrodes [4]. Ultra cell's inner diameter is of ~56 mm and length of excitation and detection electrodes is of ~70 mm. In NADEL ICR cell, the standard 90° detection electrodes were substituted with narrow aperture (flat) detection electrodes with a thickness of 1 mm. Importantly, the new flat detection electrodes were inserted radially closer to the center of the cell, **Figure 1** and Figure S1 (Supporting Information). In the evaluated here NADEL ICR cell implementation the inscribed radius was 60% of the initial ICR cell radius.

The NADEL as well as a standard Ultra (90° detection electrodes) ICR cells were mounted, one at a time, onto a hybrid linear ion trap Fourier transform ion cyclotron resonance mass spectrometer (LTQ FT-ICR MS, Thermo Scientific, Bremen, Germany) equipped with a 10 T actively shielded superconducting magnet (Oxford Nanoscience, Oxon, UK), described elsewhere. Instrument control was performed by the standard data

acquisition control system and software (Xcalibur, Thermo Scientific). The ions were formed with robotic chip-based nano-electrospray ionization source (Triversa Nanomate, Advion Biosciences, Ithaca, NY) and transported to the linear ion trap (LTQ) through an inlet set of multipole ion guides. A pre-determined number of charges to be transferred into the ICR cell using another set of multipole ion guides were controlled by the automatic gain control (AGC) function of the LTQ. For petroleomics measurements the AGC function was disabled and accumulation time in the LTQ was set to 50 ms, while the potentials of the inlet transfer system were further optimized manually. Vacuum conditions in FT-ICR MS with NADEL or Ultra ICR cells were identical ($2\text{E}-10$ Torr).

Ion trapping in ICR cell was achieved with two pairs of trapping ring electrodes. All sections of outer trapping rings were grounded throughout the complete experimental sequence. A potential of 3 V was applied to the grid-free sections of inner trapping rings to confine transferred ions inside the ICR cell during ion trapping and relaxation events. The same level of trapping potential was kept during ion excitation event, whereas it was user-defined in the range of 0-10 V during the ion detection event. Approximately 4.6-fold higher potential was applied to the sections of inner rings covered with grid. The shape of the detection electrodes in NADEL ICR cell was designed to form trapping potential distribution closer to quadratic. SIMION calculations confirmed that a certain curvature of the detection electrodes is thus required, Figure S1 and Figure 1. For the NADEL ICR cell reported here a curvature radius of 125 mm was employed.

Ion excitation with NADEL ICR cell was technically performed in the same way as with Ultra ICR cell [4]. Coherent motion of ions was excited by standard dipolar frequency-sweep excitation in the range of 72 kHz – 2150 kHz applied during 10 ms to the excitation grids. The peak-to-peak amplitude of the excitation RF voltage was varied from 0 V to 100 V. The particular advantage of the NADEL ICR cell compared to other ICR cell designs, including the Ultra cell, is in the improved excite field distribution, Figure S1, bottom left (Supporting Information). Due to the positioning of the detection electrodes in the central plane, symmetrically between the top and bottom excite electrodes, narrow aperture detection electrodes disturb the RF excite field significantly less than the standard 90° detect electrodes, as employed in Ultra cell. Therefore, ion excitation in the NADEL cell provides improved coherence of ion excitation, especially to larger radii. Overall, the maximum excitation radius is limited by the location of the detection electrodes, which may reduce the potentially useful volume of the ICR cell. However, to the best of our knowledge ion excitation to large orbits, > 0.7 radius, is not employed in modern FT-ICR MS [23, 25, 38]. Moreover, according to the estimation by the authors, most of the ICR cells function with ion excitation to 0.3-0.4 cell radius due to the increasing of the inhomogeneity of magnetic field at higher radii, although the exact data could not be found in publications. Therefore, the detection electrodes were placed at the corresponding positions, allowing ion excitation to 0.6 cell radius, and thus not reducing the effective ion volume. Note, a finite initial magnetron radius would correspondingly reduce the maximum achievable excitation radius. Figure S1, bottom right demonstrates the difference in the

ion detection conditions between NADEL and Ultra cells. Compared to the standard 90° detection electrodes, the narrow aperture detection electrodes increase the non-linear character of ion detection as a function of a post-excitation radius. Importantly for NADEL cell sensitivity level, detection electrode capacitance of the installed NADEL cells (measured together with the connecting wires and feedthrough) was reduced four-fold compared to the Ultra ICR cell, as expected.

Ion detection was performed using the commercial ion detection configuration of Ultra ICR cell. Briefly, standard dipolar differential detection was employed to acquire time domain signals (transients). The transients of variable length in the range of 92 ms – 25 s were recorded in MIDAS format at 1–5 MHz sampling frequency (narrowband ion detection) using the advanced software interface of the built-in data acquisition system (Thermo Scientific). A given number, e. g., 10–100, of transients were averaged to obtain the final transients, which were Hann-apodized and zero-filled once before fast Fourier transformation (FFT) to yield Fourier spectra which were further calibrated to mass spectra using standard data analysis software (Xcalibur, Thermo Scientific) or the framework pyFTMS developed in-house.

Finally, the offset DC potentials in the range of -100...+100 mV were applied to the NADEL and Ultra ICR cell detection and excitation electrodes independently relative to each other to vary the position of the ion cloud prior to ion excitation and during ion detection events. The influence of offset potentials on ion motion was monitored with diverse quality attributes of mass spectra, e.g., peak shape and magnetron sidebands [39]. Overall, it was possible to optimize the offset potentials to effectively decrease the

magnetron sideband intensities below 1 % relative to the corresponding reduced cyclotron frequency peak for each experiment.

Results and Discussion

NADEL ICR cell initial performance tests: MS and MS/MS data acquisition. A typical broadband mass spectrum obtained with NADEL FT-ICR MS demonstrates efficient simultaneous confinement and detection of ions during 3 s in mass range of m/z 300 – 3300, **Figure 2**. Similarly, Figure S2 (Supporting Information) shows a broadband mass spectrum of a calibration mixture revealing simultaneous detection of ions spanning a broad range of m/z 150 – 1900 with acquisition period of 0.384 s. Direct comparison of NADEL FT-ICR mass spectra with those obtained with a linear ion trap (LTQ) on the same instrument demonstrates that for all constituent signals the relative peak intensity ratios are similar (data not shown). The decrease of intensity was observed only for ions below 300 m/z , due to an expected m/z -dependent time-of-flight ion discrimination during their transfer between LTQ and ICR cell [40]. The insets in Figure 2 and Figure S2 demonstrate the typical peak shape for monoisotopic peaks of selected compounds.

In-cell tandem mass spectrometry (MS/MS) experiments with NADEL ICR cell demonstrate efficiency comparable to Ultra cell. Typical electron capture dissociation (ECD) [41], infrared multiphoton dissociation (IRMPD) [42], and electron induced dissociation (EID) [43-44] FT-ICR mass spectra are shown in Figure S3 (Supporting Information). MS/MS broadband mass spectra contain analyte peaks with a magnitude spreading over 3–4 orders.

Particularly, the spectral dynamic range (ratio of the highest to the lowest

abundance peaks within a single mass spectrum) [45] of 4.4 orders was achieved for EID MS/MS measurements. Thus, NADEL ICR cell provides dipolar excitation and detection of ions, either injected from an external ion source or produced inside of the cell, in the wide mass range without the discrimination of different m/z ions and at a dynamic range comparable to the state-of-the-art FT-ICR MS performance. Note, the employed ion excitation for standard NADEL ICR cell operation (narrowband ion detection regime) aims to excite ions to moderate radii, when ions do not yet approach the detection electrodes. The resulting transient components for these regimes are primarily sinusoidal, with a very low (< 5%) high-order harmonic content.

NADEL ICR cell performance: sensitivity. For the estimation of fundamental characteristics of the NADEL ICR cell, in particular magnitude of magnetron frequency, ω_- , and optimal post-excitation ion radii (optimal excitation parameters), the reduced cyclotron frequency ω_+ and the signal-to-noise ratio (SNR) of peaks corresponding to ω_+ frequency and its interharmonics ($n\omega_+$, $n\omega_+ \pm k\omega_-$, where $n, k = 1, 2, 3$) as the function of excitation energy were measured for isolated singly protonated ions of MRFA (m/z 524.2) at the different trapping potentials applied during ion detection event, **Figure 3**. As expected, the SNR of the peak corresponding to the reduced cyclotron frequency ω_+ linearly increases with excitation amplitude (cyclotron radius). However, a significant decrease of the SNR was observed for higher than ~40 V excitation amplitudes. On the other hand, the intensity ratio of the interharmonic $3\omega_+$ to that of the ω_+ peak increased with the excitation

amplitude, Figure 3, meaning that the post-excitation cyclotron radius continues to increase with the corresponding growth of the high-order harmonics. The SNR decline for the fundamental frequency, ω_+ , is presumably due to the reduced harmonicity of the trapping electric potential and homogeneity of the excitation electric field for high excitation amplitudes, which cause decoherence of ion clouds leading to a decrease of ion signal abundance. Additionally, we observed even harmonics of ω_+ and ω_c in mass spectra. The presence of even harmonics in mass spectra testifies for a non-zero magnetron radius and an initial radial shift of an ion cloud from the axis of the ICR cell. Notably, even harmonics' intensities were reduced to a less than 2% level by the fine tuning of the offset DC potentials applied to the excitation and detection electrodes independently for each trapping potential.

The sensitivity levels of the NADEL ICR cell and the standard Ultra cell with 90° detection electrodes were compared for an isolated singly protonated MRFA peptide (m/z 524.2) under identical optimal instrumental parameters for both cells. As expected, a signal to noise ratio (SNR) of peak corresponding to the reduced cyclotron frequency ω_+ linearly increased with the number of charges (ion abundance) for both NADEL and Ultra ICR cells, Figure 3. The SNR values reached a plateau at around $(1-2)E5$ number of charges in the isolation mode for both cells as well. This is likely due to the limitations of ion accumulation in LTQ and transport from LTQ to ICR cell of ions with the same nominal mass. Importantly, Figure 3 shows the increase of the peak SNR almost 2 times for the NADEL cell compared to the standard Ultra cell. The experimental results demonstrate the higher amount of

induced current generated in the detection circuit of the NADEL ICR cell compared to the corresponding signal in the standard 90° electrode Ultra ICR cell, despite the significantly reduced surface area of detection electrodes of the NADEL cell. This result can be explained by the substantially lower capacitance of the flat detect electrodes and shorter distance between the orbit of excited ions and detection electrodes. It is thus not surprising that only a few charges, approximately 10 ions (AGC=30) per cell load, of a triply protonated substance P (m/z 449.9) ions were detected with acquisition period of 384 ms and 5 transients averaging, Figure S4 (Supporting Information). This result correlates well with the known sensitivity performance level of the state-of-the-art FT-ICR MS with standard ICR cells, which is reported to be at the level of 50 charges.

NADEL ICR cell performance: mass accuracy. The reduced cyclotron frequency ω_+ effectively remained constant within 1 ppm range for low trapping potentials (<1 V) over the entire range of excitation amplitudes, where ion signal can be detected, except the region of low excitation amplitudes, Figure 3, bottom right. However, frequency reduced with an increase in the trapping potential for high excitation amplitudes due to increased unharmonicity of the trapping electric potential. Thus, the electric field distribution of the NADEL ICR cell is uniform in the wide range of excitation amplitudes (post excitation radii) at low trapping potentials. Note, the DC offset potentials were adjusted for each trapping potential independently when constructing these dependencies. Furthermore, at low excitation amplitudes (small post excitation radii) for all trapping potentials a

space charge of sufficient ion abundance (AGC=5E4) is the reason for the observed frequency shift, as reported earlier [46]. A magnetron frequency of 2.7 Hz/V was experimentally determined, Figure 3, bottom right. Experimental results allowed us to define optimal excitation amplitude of 30 – 40 V with a duration of 20 ms for optimal trapping potentials 0.8 – 1.0 V. Additionally, the stability of the reduced cyclotron frequency measured for isolated singly protonated ions of MRFA 524.4 m/z with trapping potential of 1.0 V and excitation amplitude of 30 V indicates the absence of space charge effect in the wide range of number of charges, AGC = 1E2 – 1E5, for the NADEL ICR cell, Figure 3, bottom left.

The internal calibration of the FT-ICR mass spectrometer equipped with the NADEL ICR cell was performed by standard two parametric calibration formula with nine monoisotopic ions of ESI-L low concentration tuning mixture (Agilent). Calibration was performed at two AGC values (5E5 and 2E6) and two acquisition times (0.578 s and 1.336 s) in the mass range of m/z 300 – 1200 with four calibrants and in the mass range of m/z 300 – 3000 with nine calibrants resulting in eight calibration parameter sets, **Table 1**. The root-mean-square (RMS) mass error of internal calibration remained within 250 ppb for the calibration in the mass range up to 1200 m/z , whereas increased up to 1.4 ppm for wider mass range. Further, the external mass calibration was evaluated by considering the mass measurement accuracies achieved for 28 analyte peaks of polyphosphoric acid $H_{n+2}P_nO_{3n+1}$ present in the same mass ranges, Table 1. The corresponding dependencies in the mass ranges of m/z 300 – 3000 are given in Figure S5 (Supporting information). The RMS mass error was within 250

ppb for the calibration in the mass range up to 1200 m/z and increased for the higher mass range up to 1.5 ppm.

Additionally, internal mass calibration was evaluated for polyphosphoric acid $H_{n+2}P_nO_{3n+1}$ in the wide mass range of m/z 300 – 3100 and number of charges $AGC=(3 - 60)E5$ with acquisition period of 1336 ms, **Table 2**. Figure S5 (Supporting Information) demonstrates the SNR of analyte peaks linearly increasing with an increase of the AGC value in the current range. The RMS mass accuracy of 60 ppb was achieved for eight peaks of polyphosphoric acid with an internal calibration in the range of 300 - 1000 m/z and $AGC=7E5$. Furthermore, the RMS mass error remained within 200 ppb for the number of charges up to $AGC=2E6$ for low mass calibration range up to 1000 m/z , Table 2 and Figure S5. However, the mass error became more pronounced for higher AGC values, $3E6$ and more, especially for the highest m/z peaks, as can be expected due to an increased space-charge field [46-47].

Thus, the mass accuracy performance with NADEL ICR cell is comparable to those obtained with conventional and harmonized cells [25, 48]. This conclusion will be further confirmed below for NADEL ICR MS application to crude oil fraction analysis. Finally, application of absorption mode FT signal processing allows improving RMS mass accuracy values two times, as expected (data not shown) [49-50].

NADEL ICR cell performance: resolving power. The resolving power performance of the NADEL ICR cell was evaluated for analysis of peptide and protein samples with maximum detection period of 25.576 s for the

employed hardware configuration. Acquisition of longer transients was not possible due to the technical limitations imposed by the on-board LTQ FT computer memory, whereas the lifetime of excited and coherent ion motion apparently surpasses this temporal limitation, for example see Figure S6 (Supporting Information).

Figure 4 shows a mass spectrum of three isobaric singly protonated peptides P1, P2, and P3 (all positioned at m/z 620) with mass differences of 0.9 mDa and 0.5 mDa between the pairs of peptides. Notably, mass of 0.5 mDa is close to the mass of an electron. Expanded segments of the mass spectrum show baseline-resolved monoisotopic and ^{13}C peaks for the three peptides of interest with the resolving power achieving 3'300'000 for a single scan. To avoid the peak coalescence, a relatively low number of ions was employed (AGC setting of 3E4). The particular importance of the result reported in Figure 4 is in the long time of ion cloud coherence after excitation, which allows recording of 25 s and longer transients. The reported separation of this peptide triplet can be compared to the notable example of a peptide doublet separation (mass difference close to the mass of an electron) by Marshall and co-workers on a 9.4 T FT-ICR MS.

Another notable example of resolving power performance provided by NADEL cell for peptide and protein analysis is the acquisition of the fine isotopic structure distribution of triply protonated substance P (449.9 m/z) and major isotopologues of 11+ charge state of bovine ubiquitin (779.0 m/z), **Figure 5**. The mass spectrum of substance P is the average of 10 transients and the corresponding averaged transient is shown in Figure S6. Mass spectrum of bovine ubiquitin was obtained from sum of 20 single time

domain signals. Fine isotopic structure components of substance P and bovine ubiquitin are resolved with resolution of 4'400'000 and 2'900'000 respectively.

To evaluate NADEL cell performance for heavy protein analysis, protein mass measurements were performed for a bovine serum albumin (BSA) with direct infusion electrospray ionization. Broadband mass spectrum of BSA obtained by averaging of 50 single scans of 1.536 s shows the charge state distribution in the range from 35+ up to 72+, Figure S7 (Supporting Information). Furthermore, the resolving power of 320'000 was achieved for an isolated BSA⁴⁸⁺ charge state with a single scan and acquisition period of 6.144 s, Figure S8 (Supporting Information).

Finally, petroleomic sample was chosen to evaluate the NADEL cell performance for complex mixture analysis. A positive ESI broadband mass spectrum of a fraction of South American crude oil was acquired with an acquisition period of 3.072 s, **Figure 6**. The insets demonstrate mass scale expansions at 451 m/z nominal mass, revealing typical representative spectral compositions for a petroleum sample with resolving power of 450'000. Mass assignments are given in **Table 3**. Additionally, mass accuracy values for the 2765 monoisotopic peaks identified in the broadband mass spectrum of a crude oil fraction in the mass range of 300 – 700 m/z , were measured as a function of m/z and SNR by using standard two-parametric equation and internal calibration based on 29 monoisotopic peaks of a prominent homologous series, Figure S9 (Supporting information). Overall, 5635 peaks were detected with the intensity higher than 2% relative to the highest magnitude peak in mass spectrum. RMS mass accuracy and

mean values were 238 ppb and 7 ppb respectively. As expected, mass error increases with decreasing SNR and increasing m/z value.

To conclude, the combined benefits of trapping and excitation field configuration in the NADEL cell provide comparable performance of standard ICR cells in routine analysis in regard to resolving power performance and approach conditions of the most sophisticated ICR cells with complex approaches to electric field harmonization, either statically or dynamically [25, 30, 48]. The described NADEL ICR cell has been employed to perform routine mass analysis in our laboratory and demonstrated an ability to achieve the required analytical objectives of molecular structural analysis in a number of applications, for example in supramolecular complexes analysis [51].

NADEL ICR cell concept advantages. The particular benefits of the NADEL ICR cell concept for current and envisioned FT-ICR MS configurations include: (i) enhanced dipolar ion excitation conditions due to improved configuration of excitation electric field. The latter is achieved by reduced excite field disturbance by radially inserted detection electrodes compared to the standard 90 degrees wide detection electrodes. Further improvement of excite electric field distribution can be presumably obtained by utilization of wide aperture excitation electrodes with an optimized angle, which can be varied up to 180°. Interestingly, flat excitation electrodes (capacitor configuration) may be considered to provide an optimum excite field configuration; (ii) simultaneous detection of a number of transients or detection at frequency multiples when several NADEL pairs are employed, for

instance leading to an optimized design of the recently implemented quadruple frequency multiple detection ICR cell [37]; (iii) enabling implementation of efficient quadrature ion excitation and quadrupolar ion detection schemes when two pairs of 90° wide excite electrodes and two NADEL pairs are employed without the need for a high-frequency switch between excitation and detection modes. Quadrupolar ion detection should further benefit high mass accuracy and high dynamic range demanding applications by providing measurements of trapping potential-insensitive unperturbed ion cyclotron frequency [24, 52]; (iv) creation of special ion motion conditions in ICR cell even with dipolar ion excitation leading to ion detection at unperturbed cyclotron frequency [53]; (v) facilitating and improving the efficiency of fluorescence-based ion spectroscopy due to substantially increased optical access to and from the cell [54-55]; and, last but not least, (vi) enabling high throughput high-resolution mass spectrometry by accelerating ions to high post-excitation radii, when significantly, an order of magnitude, higher resolving power can be obtained in the same ion detection period. The latter regime requires the development of the matching signal processing methods, capable of efficient analysis of transients with high order harmonics components [56].

Conclusions

We designed and implemented an ICR cell with a pair of narrow aperture detection electrodes (NADEL) for improving the analytical capabilities and increasing the mass analyzer design flexibility of FT-ICR MS. Importantly, the obtained results demonstrate that the narrow aperture, flat, detection

electrodes provide comparable or superior performance to the standard ICR cells in modern FT-ICR MS. The proven here ability to replace the currently employed in all ICR cell designs wide aperture (azimuthally large) detection electrodes with the narrow aperture ones opens new avenues for ICR cell development and FT-ICR MS capabilities. A number of advantages and envisioned benefits summarized above validate the importance of the described ICR cell development. Based on the initial results reported here, a true paradigm shift in ICR cells design and capabilities is expected to drive the FT-ICR MS development and applications in the near future. However, rationalizing the underlying ion physics phenomena leading to these capabilities, e. g., unperturbed ion cyclotron frequency detection with dipolar ion detection, require in-depth theoretical analysis of the described NADEL ICR cell and its further modifications.

Acknowledgments

We are grateful to Oleg Yu. Tsybin and Konstantin O. Zhurov for technical assistance and discussions. We also express our gratitude to Thermo Fisher Scientific Inc. for providing access under license to LTQ FT transient signals and advanced user interface. We thank financial support through the Swiss National Science Foundation (SNF project 200021-125147/1) and the European Research Council (ERC Starting Grant 280271).

References

- [1] Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S.: Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Reviews* 17, 1-35 (1998)
- [2] Amster, I. J.: Fourier transform mass spectrometry. *Journal of Mass Spectrometry* 31, 1325-1337 (1996)
- [3] Marshall, A. G.; Hendrickson, C. L.: High-resolution mass spectrometers. In *Annual Review of Analytical Chemistry*. 2008, vol. 1, pp 579-599.
- [4] Scigelova, M.; Hornshaw, M.; Giannakopoulos, A.; Makarov, A.: Fourier transform mass spectrometry. *Molecular & Cellular Proteomics* 10, M111.009431 (2011)
- [5] Xian, F.; Hendrickson, C. L.; Marshall, A. G.: High resolution mass spectrometry. *Analytical Chemistry* 84, 708-719 (2012)
- [6] Tsybin, Y. O.: Structural Analysis of Complex Molecular Systems by High-Resolution and Tandem Mass Spectrometry. In *Discovering the Future of Molecular Sciences*, ed. Pignataro, B. Wiley-VCH Verlag GmbH & Co. KGaA, 2014.
- [7] Hsu, C.-S.; Hendrickson, C. L.; Rodgers, R. P.; McKenna, A. M.; Marshall, A. G.: Petroleomics: Advanced molecular probe for petroleum heavy ends. *Journal of Mass Spectrometry* 46, 337-343 (2011)
- [8] Zubarev, R. A.; Makarov, A.: Orbitrap Mass Spectrometry. *Analytical Chemistry* 85, 5288-5296 (2013)

- [9] 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)
- [10] Fornelli, L.; Damoc, E.; Thomas, P. M.; Kelleher, N. L.; Aizikov, K.; Denisov, E.; Makarov, A.; Tsybin, Y. O.: Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS. *Molecular & cellular proteomics : MCP* 11, 1758-67 (2012)
- [11] Tsybin, Y. O.; Fornelli, L.; Stoermer, C.; Luebeck, M.; Parra, J.; Nallet, S.; Wurm, F. M.; Hartmer, R.: Structural Analysis of Intact Monoclonal Antibodies by Electron Transfer Dissociation Mass Spectrometry. *Analytical Chemistry* 83, 8919-8927 (2011)
- [12] Marshall, A. G.; Hendrickson, C. L.: Fourier transform ion cyclotron resonance detection: principles and experimental configurations. *International Journal of Mass Spectrometry* 215, 59-75 (2002)
- [13] Miladinovic, S. M.; Kozhinov, A. N.; Gorshkov, M. V.; Tsybin, Y. O.: On the utility of isotopic fine structure mass spectrometry in protein identification. *Analytical chemistry* 84, 4042-51 (2012)
- [14] Hebert, A. S.; Merrill, A. E.; Bailey, D. J.; Still, A. J.; Westphall, M. S.; Strieter, E. R.; Pagliarini, D. J.; Coon, J. J.: Neutron-encoded mass signatures for multiplexed proteome quantification. *Nat Meth* 10, 332-334 (2013)
- [15] McAlister, G. C.; Huttlin, E. L.; Haas, W., *et al.*: Increasing the Multiplexing Capacity of TMTs Using Reporter Ion Isotopologues with Isobaric Masses. *Analytical Chemistry* 84, 7469-7478 (2012)

- [16] Nikolaev, E. N.; Kostyukevich, Y. I.; Vladimirov, G. N.: Fourier transform ion cyclotron resonance (FT ICR) mass spectrometry: Theory and simulations. *Mass Spectrometry Reviews* n/a-n/a (2014)
- [17] Marshall, A. G.: Milestones in Fourier transform ion cyclotron resonance mass spectrometry technique development. *International Journal of Mass Spectrometry* 200, 331-356 (2000)
- [18] Guan, S.; Marshall, A. G.: Ion traps for Fourier transform ion cyclotron resonance mass spectrometry: principles and design of geometric and electric configurations. *International Journal of Mass Spectrometry and Ion Processes* 146-147, 261-296 (1995)
- [19] 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. *Analytical Chemistry* 76, 5756-5761 (2004)
- [20] Kaiser, N.; Weisbrod, C.; Quinn, J.; Blakney, G. T.; Beu, S.; Chen, T.; Hendrickson, C. L.; Marshall, A. G.: Development of an FT-ICR mass spectrometer in preparation for 21 tesla. *Proceedings of the 62nd American Society for Mass Spectrometry Conference on Mass Spectrometry and Allied Topics*, Baltimore, MD, USA, June 15-19 2014,
- [21] Guan, S.; Gorshkov, M. V.; Marshall, A. G.: Circularly polarized quadrature excitation for Fourier-transform ion cyclotron resonance mass spectrometry. *Chemical Physics Letters* 198, 143-148 (1992)
- [22] Guan, S.; Wahl, M. C.; Marshall, A. G.: Elimination of frequency drift from Fourier transform ion cyclotron resonance mass spectra by

- digital quadrature heterodyning: ultrahigh mass resolving power for laser-desorbed molecules. *Analytical Chemistry* 65, 3647-3653 (1993)
- [23] Schweikhard, L.; Marshall, A. G.: Excitation modes for Fourier transform ion cyclotron resonance mass spectrometry. *Journal of the American Society for Mass Spectrometry* 4, 433-452 (1993)
- [24] Heck, M.; Blaum, K.; Cakirli, R. B.; Rodríguez, D.; Schweikhard, L.; Stahl, S.; Ubieto-Díaz, M.: Dipolar and quadrupolar detection using an FT-ICR MS setup at the MPIK Heidelberg. *Hyperfine Interact* 199, 347-355 (2011)
- [25] Tolmachev, A. V.; Robinson, E. W.; Wu, S.; Kang, H.; Lourette, N. M.; Pasa-Tolic, L.; Smith, R. D.: Trapped-ion cell with improved DC potential harmonicity for FT-ICR MS. *Journal of the American Society for Mass Spectrometry* 19, 586-97 (2008)
- [26] Tolmachev, A. V.; Robinson, E. W.; Wu, S.; Smith, R. D.; Futrell, J. H.; Pasa-Tolic, L.: Angular averaged profiling of the radial electric field in compensated FTICR Cells. *Journal of the American Society for Mass Spectrometry* 23, 1169-72 (2012)
- [27] Brustkern, A. M.; Rempel, D. L.; Gross, M. L.: An Electrically Compensated Trap Designed to Eighth Order for FT-ICR Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* 19, 1281-1285 (2008)
- [28] Bruce, J. E.; Anderson, G. A.; Lin, C.-Y.; Gorshkov, M.; Rockwood, A. L.; Smith, R. D.: A novel high-performance Fourier transform ion cyclotron resonance cell for improved biopolymer characterization. *Journal of Mass Spectrometry* 35, 85-94 (2000)

- [29] Weisbrod, C. R.; Kaiser, N. K.; Skulason, G. E.; Bruce, J. E.: Trapping Ring Electrode Cell: A FTICR Mass Spectrometer Cell for Improved Signal-to-Noise and Resolving Power. *Analytical Chemistry* 80, 6545-6553 (2008)
- [30] Boldin, I. A.; Nikolaev, E. N.: Fourier transform ion cyclotron resonance cell with dynamic harmonization of the electric field in the whole volume by shaping of the excitation and detection electrode assembly. *Rapid communications in mass spectrometry: RCM* 25, 122-6 (2011)
- [31] Popov, I. A.; Nagornov, K.; Vladimirov, G.; Kostyukevich, Y. I.; Nikolaev, E. N.: Twelve Million Resolving Power on 4.7 T Fourier Transform Ion Cyclotron Resonance Instrument with Dynamically Harmonized Cell—Observation of Fine Structure in Peptide Mass Spectra. *Journal of The American Society for Mass Spectrometry* 1-10 (2014)
- [32] Nikolaev, E. N.; Jertz, R.; Grigoryev, A.; Baykut, G.: Fine Structure in Isotopic Peak Distributions Measured Using a Dynamically Harmonized Fourier Transform Ion Cyclotron Resonance Cell at 7 T. *Analytical Chemistry* 84, 2275-2283 (2012)
- [33] Nikolaev, E. N.; Gorshkov, M. V.; Mordehai, A. V.; Talrose, V. L.: Ion cyclotron resonance signal-detection at multiples of the cyclotron frequency. *Rapid Communications in Mass Spectrometry* 4, 144-146 (1990)

- [34] Grosshans, P. B.; Marshall, A. G.: General theory of excitation in ion cyclotron resonance mass spectrometry. *Analytical Chemistry* 63, 2057-2061 (1991)
- [35] Knobler, M.; Wanczek, K. P.: Detection of Harmonics and Multiples of the Fundamental Frequency in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Proceedings of the 45th ASMS Conference on Mass Spectrometry and Allied Topics*, Palm Springs, CA, USA, June 1-5
- [36] Nikolaev, E. N.; Rakov, V. S.; Futrell, J. H.: Analysis of harmonics for an elongated FTMS cell with multiple electrode detection. *International Journal of Mass Spectrometry and Ion Processes* 157, 215-232 (1996)
- [37] Nagornov, K. O.; Gorshkov, M. V.; Kozhinov, A. N.; Tsybin, Y. O.: High-Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry with Increased Throughput for Biomolecular Analysis. *Analytical Chemistry* 86, 9020-9028 (2014)
- [38] Gorshkov, M. V.; Nikolaev, E. N.: Optimal cyclotron radius for high resolution FT-ICR spectrometry. *International Journal of Mass Spectrometry and Ion Processes* 125, 1-8 (1993)
- [39] Miladinovic, S. M.; Kozhinov, A. N.; Tsybin, O. Y.; Tsybin, Y. O.: Sidebands in Fourier transform ion cyclotron resonance mass spectra. *International Journal of Mass Spectrometry* 325, 10-18 (2012)
- [40] Kaiser, N.; Quinn, J.; Blakney, G.; Hendrickson, C.; Marshall, A.: A Novel 9.4 Tesla FTICR Mass Spectrometer with Improved Sensitivity, Mass Resolution, and Mass Range. *Journal of The American Society for Mass Spectrometry* 22, 1343-1351 (2011)

- [41] Zhurov, K. O.; Fornelli, L.; Wodrich, M. D.; Laskay, U. A.; Tsybin, Y. O.: Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. *Chemical Society Reviews* 42, 5014-5030 (2013)
- [42] Polfer, N. C.: Infrared multiple photon dissociation spectroscopy of trapped ions. *Chemical Society Reviews* 40, 2211-2221 (2011)
- [43] Fung, Y. M. E.; Adams, C. M.; Zubarev, R. A.: Electron Ionization Dissociation of Singly and Multiply Charged Peptides. *Journal of the American Chemical Society* 131, 9977-9985 (2009)
- [44] Zubarev, R. A.; Yang, H.: Multiple Soft Ionization of Gas-Phase Proteins and Swift Backbone Dissociation in Collisions with ≤ 99 eV Electrons. *Angewandte Chemie International Edition* 49, 1439-1441 (2010)
- [45] Kaiser, N. K.; McKenna, A. M.; Savory, J. J.; Hendrickson, C. L.; Marshall, A. G.: Tailored Ion Radius Distribution for Increased Dynamic Range in FT-ICR Mass Analysis of Complex Mixtures. *Analytical Chemistry* 85, 265-272 (2012)
- [46] Easterling, M. L.; Mize, T. H.; Amster, I. J.: Routine part-per-million mass accuracy for high-mass ions: Space-charge effects in MALDI FT-ICR. *Analytical Chemistry* 71, 624-632 (1999)
- [47] Masselon, C.; Tolmachev, A. V.; Anderson, G. A.; Harkewicz, R.; Smith, R. D.: Mass measurement errors caused by "local" frequency perturbations in FTICR mass spectrometry. *Journal of the American Society for Mass Spectrometry* 13, 99-106 (2002)

- [48] Schaub, T. M.; Hendrickson, C. L.; Horning, S.; Quinn, J. P.; Senko, M. W.; Marshall, A. G.: High-Performance Mass Spectrometry: Fourier Transform Ion Cyclotron Resonance at 14.5 Tesla. *Analytical Chemistry* 80, 3985-3990 (2008)
- [49] 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. *Analytical Chemistry* 85, 3903-3911 (2013)
- [50] 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. *Analytical Chemistry* 82, 8807-8812 (2010)
- [51] Schouwey, C.; Holstein, J. J.; Scopelliti, R.; Zhurov, K. O.; Nagornov, K. O.; Tsybin, Y. O.; Smart, O. S.; Bricogne, G.; Severin, K.: Self-Assembly of a Giant Molecular Solomon Link from 30 Subcomponents. *Angewandte Chemie Int. Ed.* (DOI: 10.1002/anie.201407144)
- [52] Schweikhard, L.; Lindinger, M.; Kluge, H. J.: Quadrupole-detection FT-ICR mass spectrometry. *International Journal of Mass Spectrometry and Ion Processes* 98, 25-33 (1990)
- [53] Nagornov, K. O.; Kozhinov, A. N.; Zhurov, K. O.; Tsybin, Y. O.: Mapping ion cloud dynamics reveals optimal regimes in FT-ICR MS. *Proceedings of the 62nd American Society for Mass Spectrometry Conference on Mass Spectrometry and Allied Topics*, Baltimore, MD, USA, June 15-19 2014,
- [54] Chingin, K.; Chen, H.; Gamez, G.; Zenobi, R.: Exploring Fluorescence and Fragmentation of Ions Produced by Electrospray Ionization in

- Ultrahigh Vacuum. *Journal of the American Society for Mass Spectrometry* 20, 1731-1738 (2009)
- [55] Frankevich, V.; Chagovets, V.; Widjaja, F.; Barylyuk, K.; Yang, Z.; Zenobi, R.: Fluorescence resonance energy transfer of gas-phase ions under ultra high vacuum and ambient conditions. *Physical Chemistry Chemical Physics* 16, 8911-8920 (2014)
- [56] Kozhinov, A. N.; Nagornov, K. O.; Ayoub, D.; Tsybin, Y. O.: Unexplored reserves of resolution in Fourier transform mass spectrometry. *Proceedings of the 62nd American Society for Mass Spectrometry Conference on Mass Spectrometry and Allied Topics*, Baltimor, MD, USA, June 15-19 2014

Figure captions.

Figure 1. Schematic representation of the narrow aperture detection electrodes (NADEL) ICR cell: (top panel) a 3D view and (bottom panel) its unrolled surface. The NADEL cell contains two pairs, outer and inner, of trapping ring electrodes, excitation grids, conventional 90° excitation and narrow aperture (flat) detection electrodes with a thickness of 2 mm and curvature $R = 125$ mm.

Figure 2. Broadband mass spectrum of a poly-phosphoric acid, acquired with 10 T FT-ICR MS equipped with NADEL ICR cell. The mass spectrum was obtained from a sum of 30 transients recorded in the mass range of 300 – 3300 m/z with the acquisition period 3.072 s and number of charges $1E6$ (AGC).

Figure 3. Characterization of the NADEL ICR cell in 10 T FT-ICR MS. (Top left): relative (w.r.t. intensity of ω_+) intensity of peaks corresponding to the interharmonic $3\omega_+$ and absolute intensity of peak corresponding to the reduced cyclotron frequency ω_+ , as a function of the excitation amplitude. (Top right) dependence of the absolute abundance of peak corresponding to the reduced cyclotron frequency ω_+ on the defined number of charges (logarithmic scales) were obtained under identical instrumental parameters for both the standard (open circles) and NADEL (solid circles) ICR cells. (Bottom left) dependence of a reduced cyclotron frequency (triangles) and of the abundance of corresponding peak (circles) on the user defined value of

number of charges (AGC) in the wide range $2E1 - 2E6$. (Bottom right) dependence of a reduced cyclotron frequency ω_+ on the excitation amplitude for different trapping potentials. All experimental data were acquired for isolated singly protonated MRFA peptide (m/z 524.2) with the acquisition period of 3.072 s and 10 ms of excitation duration.

Figure 4. Isobaric peptide mixture measurements with NADEL cell in 10 T FT-ICR MS. The resolving power 3'300'000 was achieved for a single scan with acquisition period of 25.576 s and number of charges (AGC) of 30'000 for isolated ions of triplet of singly protonated peptides P1, P2 and P3.

Figure 5. Peptide (substance P, top panel) and protein (ubiquitin, bottom panel) fine structure isotopic distribution measurements with NADEL cell in 10 T FT-ICR MS. Mass spectrum of isolated triply protonated substance P 449.9 m/z was obtained from a sum of 10 transients of 25.576 s each. The corresponding transient is shown in Figure S6 (Supporting Information). Mass spectrum of bovine ubiquitin was obtained from sum of 20 transients of 25.576 s acquired for isolated 11+ ions.

Figure 6. Electrospray ionization broadband mass spectrum of a crude oil fraction obtained with NADEL ICR cell in 10 T FT-ICR MS from the averaging of 600 transients of 3.072 s each in the positive-ion mode. Insets show expanded regions around (top) 451 m/z and (bottom) 445 - 453 m/z for the displayed broadband mass spectrum. Annotated peaks are listed in Table 3.

Table 1. RMS mass accuracy of Agilent ESI-L low concentration tuning mixture (internal calibration) and phosphoric acid (external calibration, based on the ESI-L mixture) obtained with two values of a number of charges (AGC equal to 5E5 or 2E6) and two acquisition periods (0.578 s and 1.336 s). Mass calibration procedures were performed in the mass ranges of m/z 300 – 1200 and m/z 300 – 3000, using 30 summed transients for each of the ranges. The corresponding dependencies for external calibration are shown in Figure S5 (Supporting Information).

Table 2. RMS mass accuracy of poly-phosphoric acid (internal calibration) as a function of number of charges. Mass calibration was performed in the mass ranges of m/z 300 – 1000, m/z 300 – 2000, and m/z 300 – 3100. For each of the ranges, 50 transients with acquisition period of 1336 ms were summed. The corresponding dependence is graphically shown in Figure S5 (Supporting Information).

Table 3. Assignments of peaks and mass errors in the segments around 451 m/z of the positive electrospray broadband mass spectrum of a crude oil fraction. Corresponding inset is shown in Figure 6.

Figure 1.

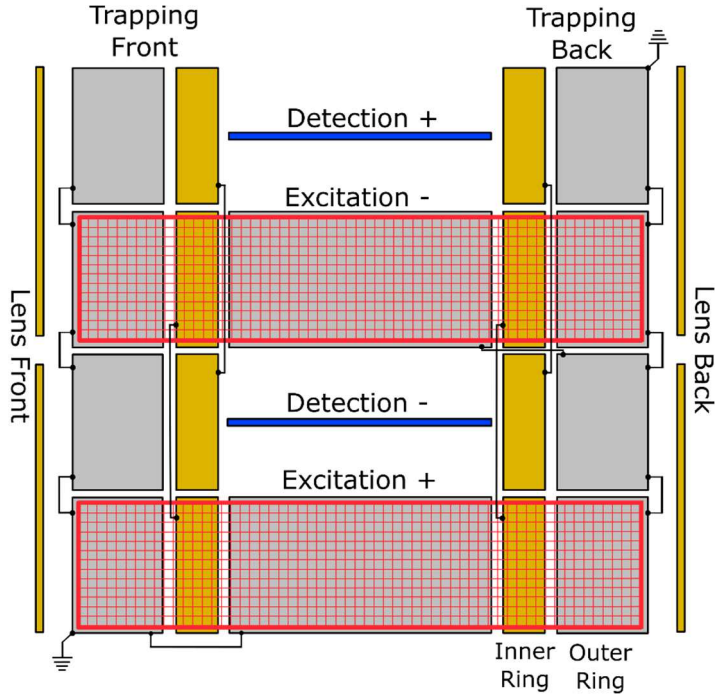
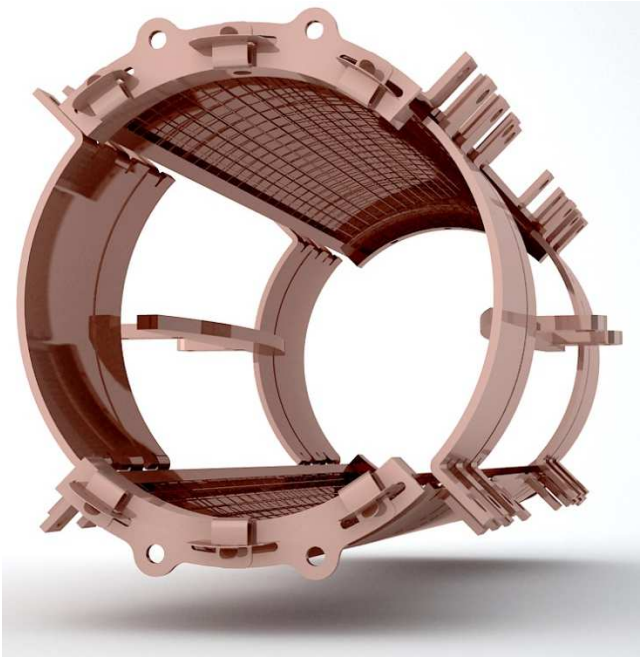


Figure 2.

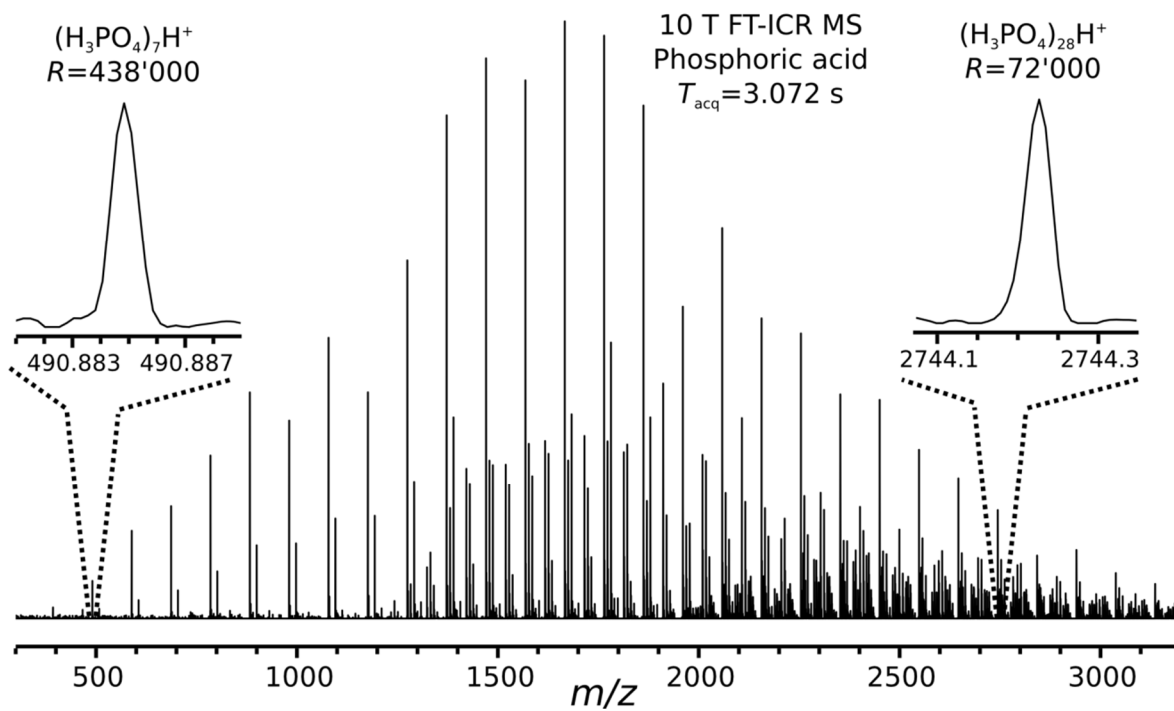


Figure 3.

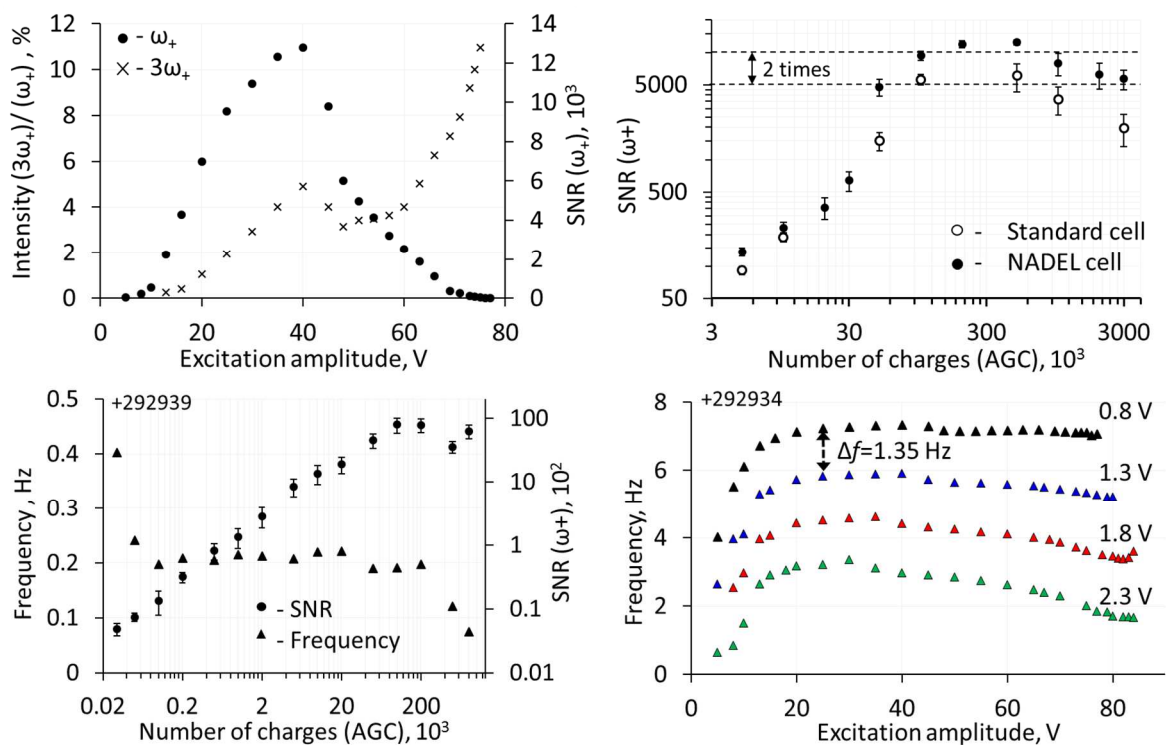


Figure 4.

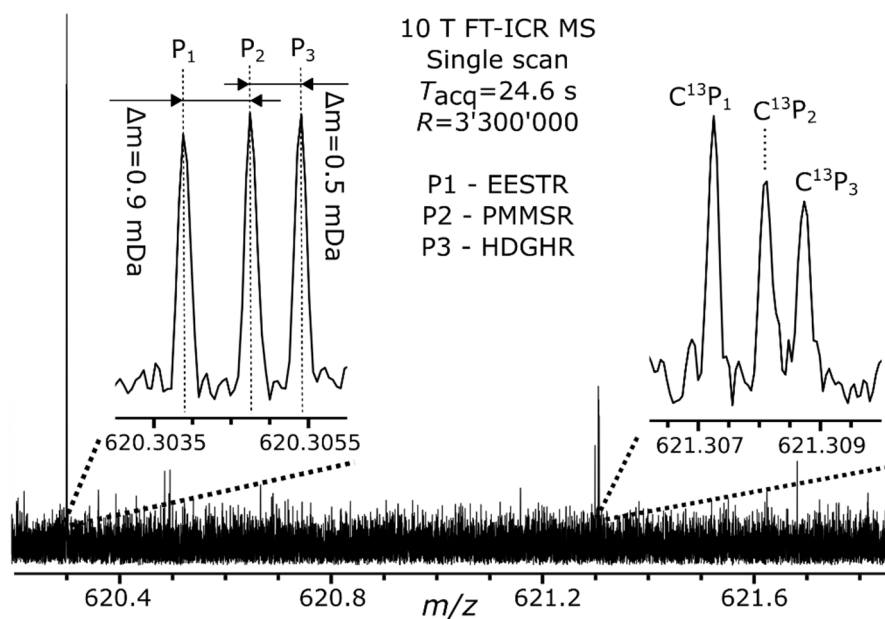


Figure 5.

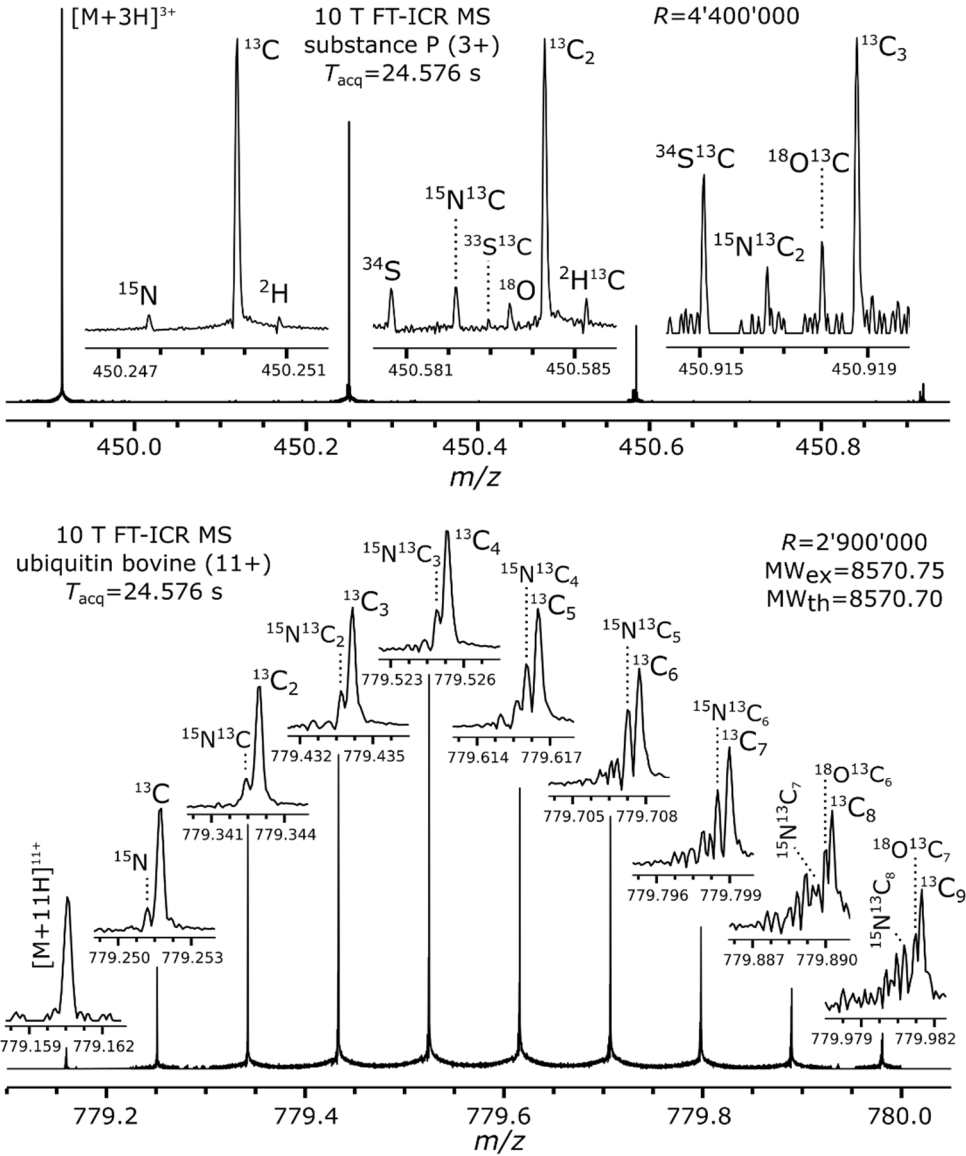


Figure 6.

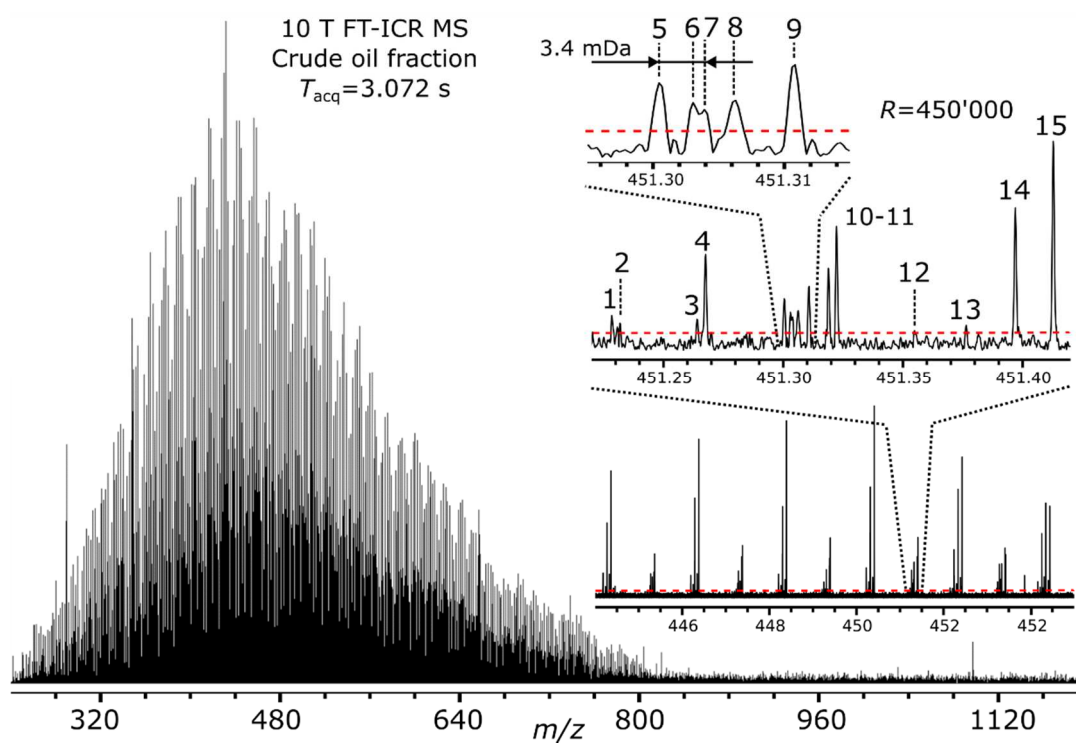


Table 1.

AGC	Mass range, m/z	RMS mass error ($\Delta m/m$, ppm)			
		Internal calibration (ESI-L Agilent)		External calibration Poly-(H ₃ PO ₄) _n	
		$T_{acq}=578$ ms	$T_{acq}=1336$ ms	$T_{acq}=578$ ms	$T_{acq}=1336$ ms
5E5	300-1200	0.08	0.13	0.20	0.18
	300-3000	0.49	0.79	0.76	0.83
2E6	300-1200	0.15	0.16	0.22	0.23
	300-3000	0.56	1.35	1.19	1.49

Table 2.

Mass range, <i>m/z</i>	RMS mass error ($\Delta m/m$, ppm), internal calibration				
	AGC=3E5	AGC=7E5	AGC=2E6	AGC=4E6	AGC=6E6
300 – 1000 (8 peaks)	0.11	0.06	0.18	0.37	0.57
300 – 2000 (19 peaks)	0.69	0.44	0.98	1.67	2.21
300 – 3100 (29 peaks)	0.92	0.88	1.43	2.54	3.35

Table 3.

peak no.	composition	<i>m/z</i>	$\Delta m/m$, ppm	class
1	$^{13}\text{C}_1\text{C}_{31}\text{H}_{32}\text{NS}$	451.228355	-0.135	NS
2	$^{13}\text{C}_1\text{C}_{28}\text{H}_{36}\text{NS}_2$	451.231726	-0.098	NS ₂
3	$\text{C}_{27}\text{H}_{40}\text{O}_2\text{SNa}$	451.264122	0.119	O ₂ S
4	$\text{C}_{24}\text{H}_{44}\text{O}_2\text{S}_2\text{Na}$	451.267496	0.059	O ₂ S ₂
5	$\text{C}_{28}\text{H}_{44}\text{OSNa}$	451.300510	-0.069	OS
6	$\text{C}_{30}\text{H}_{43}\text{OS}$	451.302913	0.295	OS
7	$\text{C}_{25}\text{H}_{48}\text{OS}_2\text{Na}$	451.303879	-0.337	OS ₂
8	$\text{C}_{27}\text{H}_{47}\text{OS}_2$	451.306285	0.120	OS ₂
9	$\text{C}_{32}\text{H}_{39}\text{N}_2$	451.310778	-0.152	N ₂
10	$^{13}\text{C}_1\text{C}_{33}\text{H}_{40}\text{N}$	451.318881	-0.209	N
11	$^{13}\text{C}_1\text{C}_{30}\text{H}_{44}\text{NS}$	451.322255	-0.035	NS
12	$\text{C}_{29}\text{H}_{48}\text{O}_2\text{Na}$	451.354651	0.202	O ₂
13	$\text{C}_{26}\text{H}_{53}\text{NOSNa}$	451.375783	0.078	NOS
14	$\text{C}_{29}\text{H}_{55}\text{OS}$	451.396816	0.155	OS
15	$^{13}\text{C}_1\text{C}_{32}\text{H}_{52}\text{N}$	451.412784	0.034	N

Supporting Information.

Paradigm shift in ion trap design and capabilities for Fourier transform ion cyclotron resonance mass spectrometry

Konstantin O. Nagornov, Anton N. Kozhinov, and Yury O. Tsybin

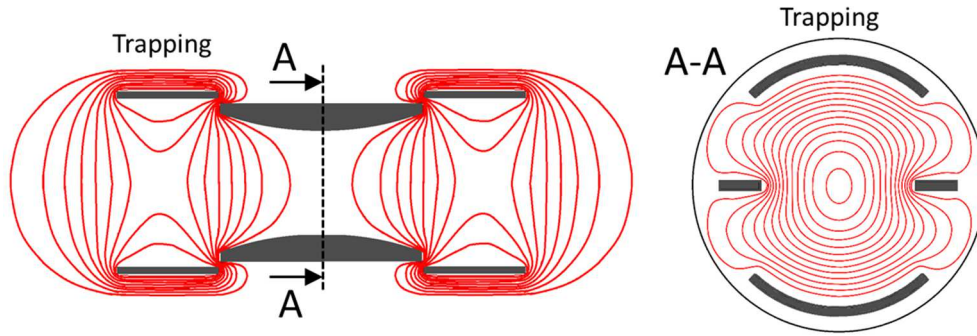
Biomolecular Mass Spectrometry Laboratory, Ecole Polytechnique Fédérale de Lausanne,
1015 Lausanne, Switzerland

Table of SI Contents:

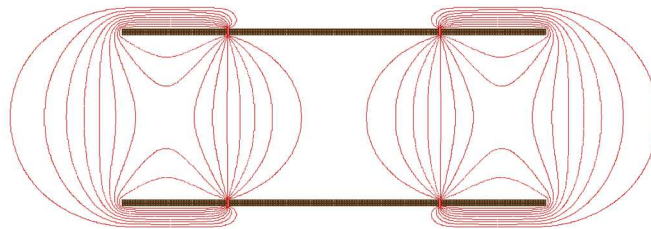
1. Figure S1. Electric potential and field distributions in NADEL ICR cell.
2. Figure S2. Broadband mass spectrum of a calibration mixture.
3. Figure S3. MS/MS: ECD, IRMPD, and EID of substance P dication.
4. Figure S4. Limit of detection evaluation of NADEL ICR cell.
5. Figure S5. Characterization of mass accuracy performance.
6. Figure S6. Time domain signal (transient), 25 s long.
7. Figure S7. Bovine serum albumin MS: charge state distribution.
8. Figure S8. Bovine serum albumin MS: +48 charge state isolation.
9. Figure S9. Mass accuracy of oil crude fraction measurements.

Figure S1. SIMION-modeled electric potential distributions in the ICR cells with narrow aperture (curved or rectangular) detection and standard 90° excitation electrodes, as well as with standard 90° detection/excitation electrodes. Top panel: the trapping potential distribution in the plane (left) along magnetic field and (right) perpendicular to magnetic field. Bottom panel: the excitation field lines and the detection electrode potential contours. The employed simulation parameters: excitation electrodes sustained at -1 V and 1 V, detection electrodes at 0 V and 1 V, trapping electrodes at 1 V; all contours correspond to the range of 0.1 V to 0.9 V with 0.1 V increments.

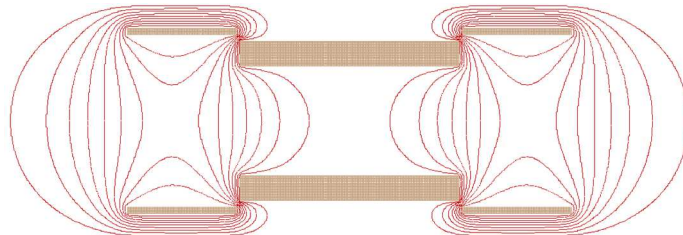
NADEL cell with curved detection electrodes:



Standard cylindrical ICR cell:



NADEL cell with rectangular detection electrodes:



NADEL ICR cell with curved detection electrodes vs. standard cylindrical ICR cell: excitation and detection

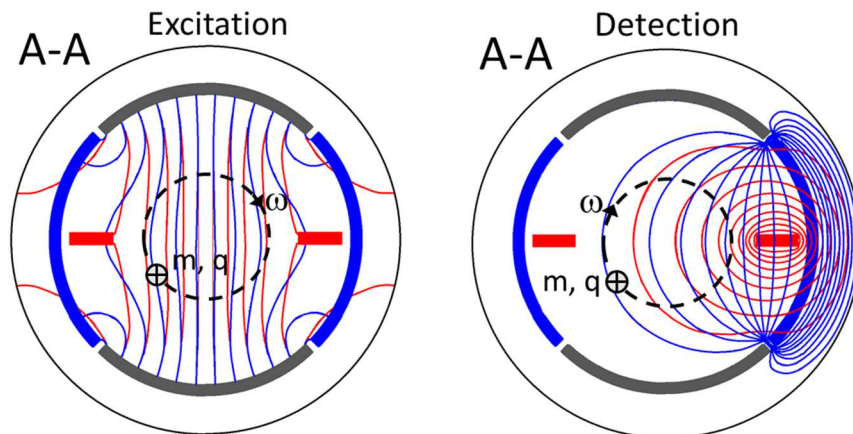


Figure S2. NADEL ICR cell 10 T FT-ICR MS broadband mass spectrum of a calibration mixture containing MRFA, caffeine, and Ultramark compounds. The mass spectrum was obtained from a sum of 5 transients recorded in the mass range of m/z 150 – 1900 with acquisition period 0.384 s and with number of charges 5E5 (AGC).

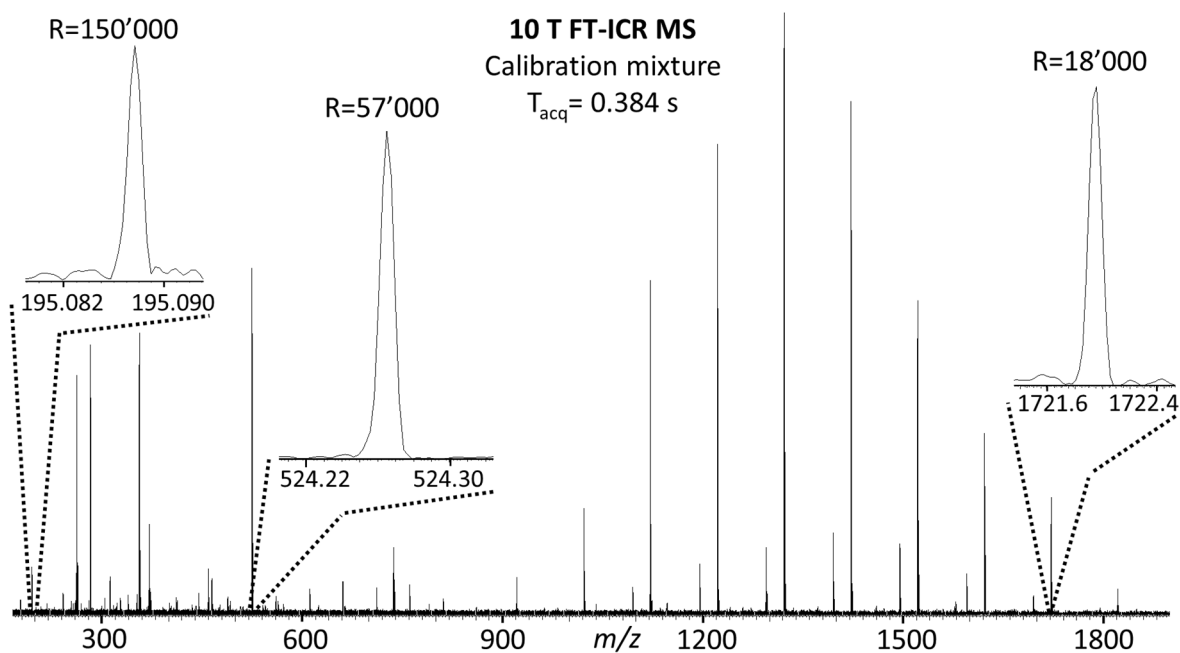


Figure S3. NADEL cell 10 T FT-ICR MS/MS: (top left) ECD, (top right) IRMPD and (bottom) EID mass spectra of doubly protonated substance P (674 m/z). Experimental ECD parameters: acquisition period 384 ms; AGC=5E5; 20 scans averaged; electron energy 5 eV, electron injection duration 80 ms. IRMPD parameters: acquisition period 384 ms; AGC=5E5; 20 scans averaged; laser energy 80 %, photon injection duration 35 ms. EID parameters: acquisition period 1.536 s; AGC=1E6; 50 scans averaged; electron energy 70 eV, electron injection duration 5 ms.

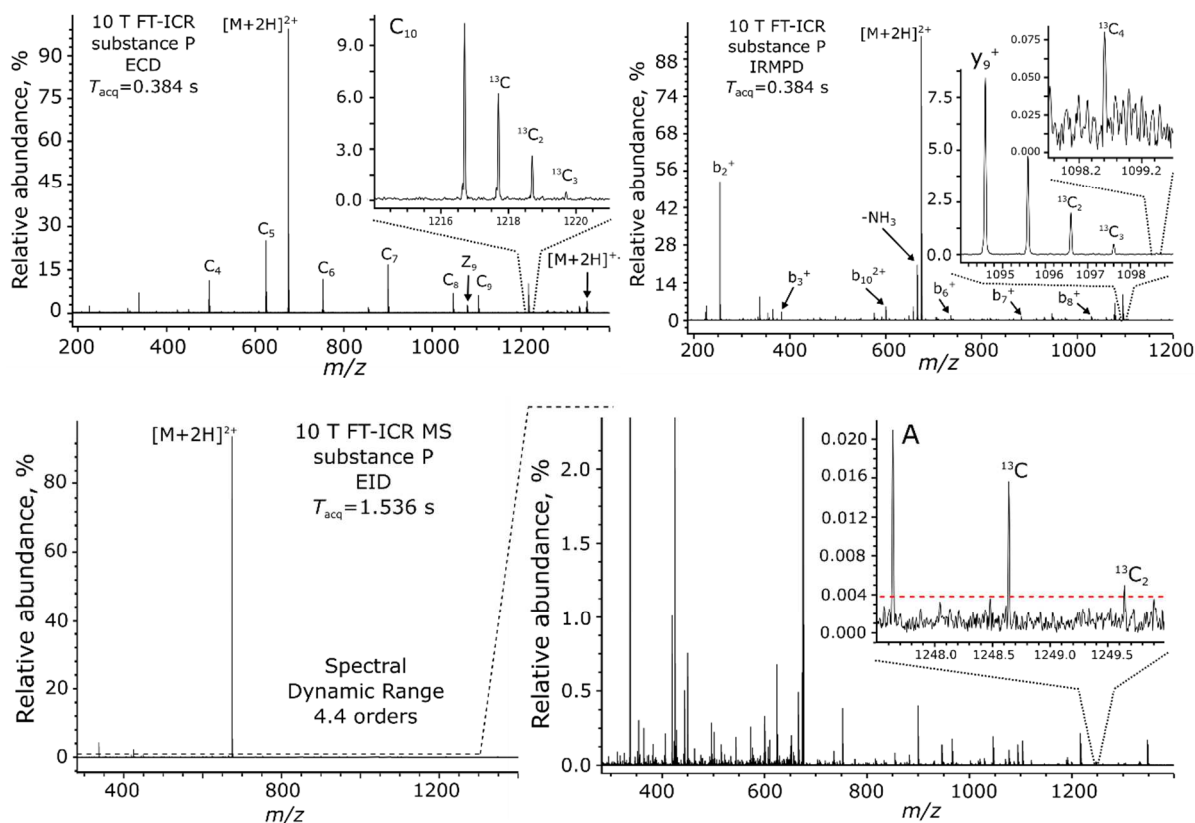


Figure S4. Limit of detection evaluation of NADEL ICR cell. Left panel: mass spectrum of triply protonated substance P 449.9 m/z acquired with NADEL cell in 10 T FT-ICR MS for a low number of ions (AGC=30) with acquisition period of 0.384 s. Right panel: mass spectrum of doubly protonated substance P 674.4 m/z acquired with NADEL ICR cell in 10 T FT-ICR MS for a low number of ions (AGC=10) with acquisition period of 0.768 s.

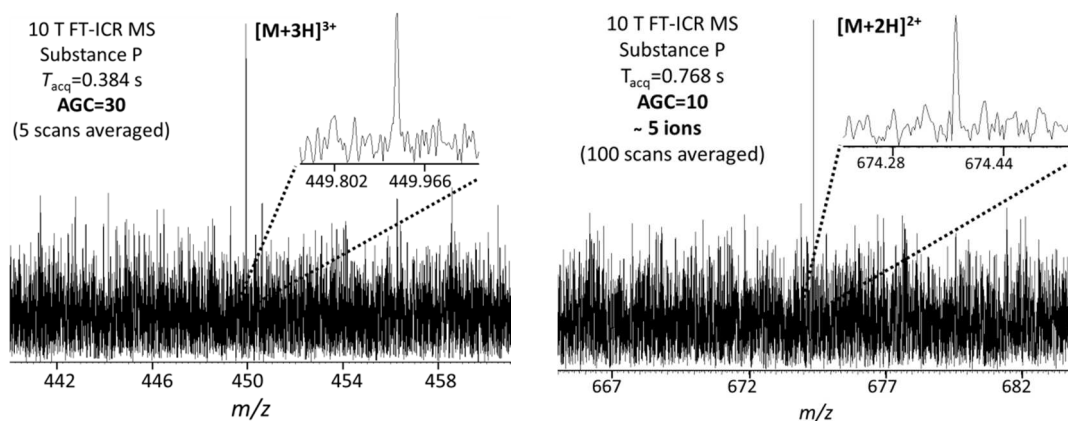


Figure S5. Characterization of mass accuracy performance of the NADEL ICR cell in 10 T FT-ICR MS. The mass accuracy for broadband mass measurements of phosphoric acid H_3PO_4 with external calibration was calculated with acquisition period (top left) of 578 ms and (top right) 1336 ms with number of charges AGC=5e5 (empty circles) and AGC =2e6 (full circles). Numerical values are given in Table 1 (main text). (Bottom left) total SNR measured in each scan for analyte peaks from broadband mass spectrum of mixture of poly-phosphoric acid poly- $(\text{H}_3\text{PO}_4)_n$ in the mass range of 250 - 3100 m/z as a function of number of charges AGC. Only peaks with abundance higher than $3 \cdot \sigma$ were considered for the calculation of total SNR in each mass spectrum. (Bottom right) the dependencies of mass accuracy were measured with an internal calibration as a function of total SNR for broadband mass measurements (50 scans averaged) with 8 analyte peaks of phosphoric acid H_3PO_4 in the range of 250 - 1000 m/z and 29 analyte peaks in 250 - 3100 m/z .

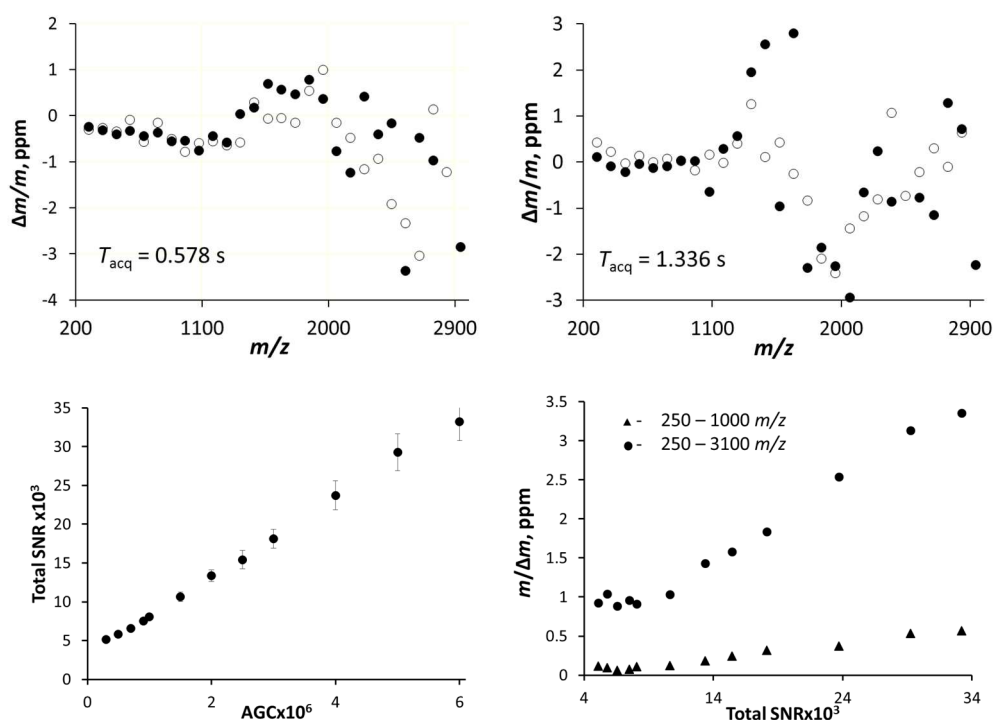


Figure S6. NADEL ICR cell 10 T FT-ICR MS time domain signal (transient) of the isolated triply protonated substance P, m/z 449.9.

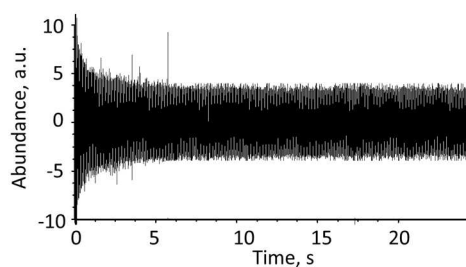


Figure S7. Heavy protein mass measurements with NADEL ICR cell in 10 T FT-ICR MS. Broadband mass spectrum of bovine serum albumin (BSA) obtained by averaging of 50 transients of 1.536 s shows the distribution of charge states in the range from 35+ up to 72+.

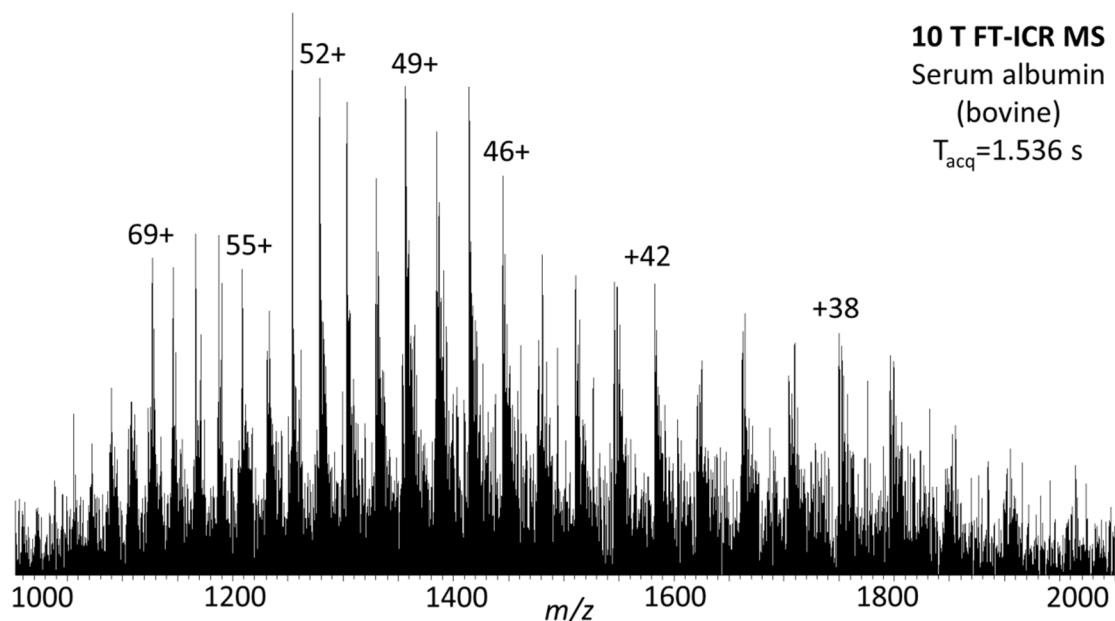


Figure S8. Heavy protein mass measurements with NADEL ICR cell in 10 T FT-ICR MS. Isolated charge state BSA⁴⁸⁺ was resolved for a single scan with acquisition period of 6.144 s. Left inset shows the baseline resolution of isotopic distribution. Right inset shows time domain signal with the main beats.

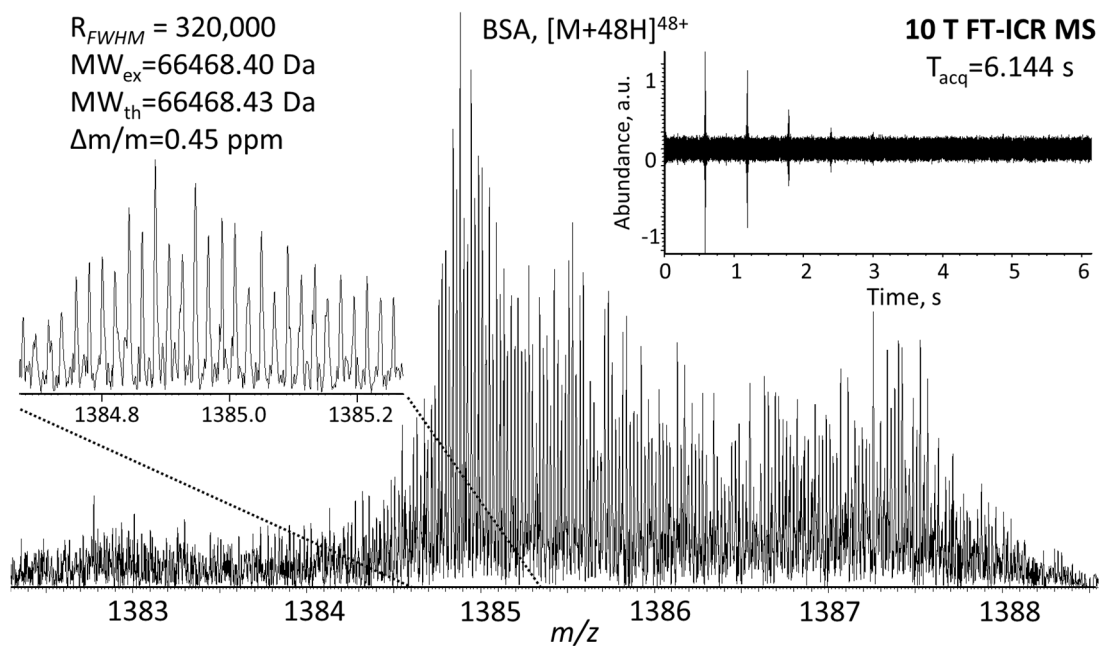
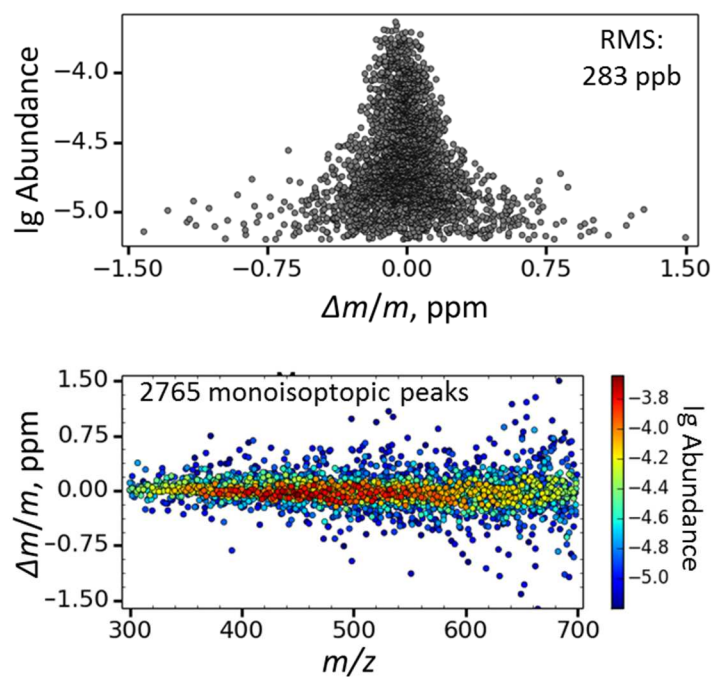


Figure S9. Mass accuracy of monoisotopic peaks identified in the broadband mass spectrum of crude oil fraction in the range 300 – 700 m/z . (Top) decimal logarithm of abundance vs. mass error and (bottom) mass error vs. m/z of analytes for the corresponding broadband mass spectra. Corresponding mass spectrum is shown in Figure 6.



Bibliography

1. J. J. Thomson, "Cathode rays," *Philosophical Magazine*, vol. 293, 1897.
2. J. J. Thomson, "Further experiments on positive rays," *Philosophical Magazine*, vol. 24, no. 140, pp. 209–253, 1912.
3. J. J. Thomson, "Rays of positive electricity," *Proceedings of the Royal Society A*, vol. 89, no. 140, pp. 1–20, 1913.
4. A. J. Dempster, "A new method of positive ray analysis," *Phys. Rev.*, vol. 11, pp. 316–325, 1918.
5. F. W. Aston *Philos. Mag.*, vol. 38, p. 707, 1919.
6. M. B. Comisarow and A. G. Marshall, "Fourier transform ion cyclotron resonance spectroscopy," *Chemical Physics Letters*, vol. 25, no. 2, pp. 282–283, 1974.
7. M. B. Comisarow and A. G. Marshall, "Frequency-sweep Fourier transform ion cyclotron resonance spectroscopy," *Chemical Physics Letters*, vol. 26, no. 4, pp. 489–490, 1974.
8. A. Makarov, "Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis," *Analytical Chemistry*, vol. 72, no. 6, pp. 1156–1162, 2000.
9. "NIST: Atomic weights and isotopic compositions of the elements." <http://www.nist.gov/pml/data/comp.cfm>. [Accessed 3 December 2011].
10. J. R. de Laeter, J. K. Böhlke, P. De Bièvre, H. Hidaka, H. S. Peiser, K. J. R. Rosman, and P. D. P. Taylor, "Atomic weights of the elements. Review 2000 (IUPAC Technical Report)," *Pure and Applied Chemistry*, vol. 75, no. 6, pp. 683–800, 2003.
11. J. A. Yergey, "A general approach to calculating isotopic distributions for mass spectrometry," *International Journal of Mass Spectrometry and Ion Physics*, vol. 52, no. 2–3, pp. 337–349, 1983.
12. J. F. J. Todd, "Recommendations for nomenclature and symbolism for mass spectroscopy (including an appendix of terms used in vacuum technology). (Recommendations 1991)," *Pure and Applied Chemistry*, vol. 63, no. 10, pp. 1541–1566, 1991.

Bibliography

13. K. T. Bainbridge, "The isotopic weight of H₂," *Physical Review*, vol. 42, pp. 1–10, 1932.
14. K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida, and T. Matsuo, "Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry," *Rapid Communications in Mass Spectrometry*, vol. 2, no. 8, pp. 151–153, 1988.
15. J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse, "Electrospray ionization for mass spectrometry of large biomolecules," *Science*, vol. 246, no. 4926, pp. 64–71, 1989.
16. W. Paul and H. S. Steinwedel, "Ein neues massenspektrometer ohne magnetfeld," *Z. Naturforsch. A*, vol. 8, no. 7, pp. 448–450, 1953.
17. B. A. Mamyurin, V. I. Karataev, D. V. Shmikk, and V. A. Zagulin, "The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution," *Sov. Phys.-Tech. Phys.*, vol. 16, p. 1177, 1972.
18. A. G. Marshall, C. L. Hendrickson, and G. S. Jackson, "Fourier transform ion cyclotron resonance mass spectrometry: A primer," *Mass Spectrometry Reviews*, vol. 17, no. 1, pp. 1–35, 1998.
19. I. J. Amster, "Fourier transform mass spectrometry," *Journal of Mass Spectrometry*, vol. 31, no. 12, pp. 1325–1337, 1996.
20. A. G. Marshall, "Milestones in Fourier transform ion cyclotron resonance mass spectrometry technique development," *International Journal of Mass Spectrometry*, vol. 200, no. 1-3, pp. 331–356, 2000. Volume 200: The state of the field as we move into a new millenium.
21. M. Scigelova, M. Hornshaw, A. Giannakopoulos, and A. Makarov, "Fourier transform mass spectrometry," *Molecular & Cellular Proteomics*, vol. 10, no. 7, 2011.
22. A. G. Marshall and C. L. Hendrickson, "High-resolution mass spectrometers," *Annual Review of Analytical Chemistry*, vol. 1, no. 1, pp. 579–599, 2008.
23. A. G. Marshall and T. Chen, "40 years of Fourier transform ion cyclotron resonance mass spectrometry," *International Journal of Mass Spectrometry*, 2014. DOI:10.1016/j.ijms.2014.06.034.
24. W. Heisenberg, "Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik," *Zeitschrift für Physik*, vol. 43, no. 3-4, pp. 172–198, 1927.
25. M. Benedicks, "On Fourier transforms of functions supported on sets of finite Lebesgue measure," *Journal of Mathematical Analysis and Applications*, vol. 106, no. 1, pp. 180–183, 1985.
26. A. G. Bendat, J. S.; Piersol, *Random Data: Analysis and Measurement Procedures*. 10, Hoboken, NJ: Wiley, 4 ed., 2010.

27. D. A. Dahl, "SIMION for the personal computer in reflection," *International Journal of Mass Spectrometry*, vol. 200, no. 1–3, pp. 3–25, 2000. Volume 200: The state of the field as we move into a new millenium.
28. S. Guan and A. G. Marshall, "Ion traps for Fourier transform ion cyclotron resonance mass spectrometry: principles and design of geometric and electric configurations," *International Journal of Mass Spectrometry and Ion Processes*, vol. 146–147, pp. 261–296, 1995. Honour Biography.
29. H. Hartmann, K. Chung, G. Baykut, and K. Wanczek, "Dependence of ion cyclotron frequency on electric field inhomogeneity," *The Journal of Chemical Physics*, vol. 78, no. 1, 1983.
30. M. V. Gorshkov and E. N. Nikolaev, "Optimal cyclotron radius for high resolution FT-ICR spectrometry," *International Journal of Mass Spectrometry and Ion Processes*, vol. 125, no. 1, pp. 1 – 8, 1993.
31. D. W. Mitchell, "Theory of trapped ion motion in the non-quadrupolar electrostatic potential of a cubic ion cyclotron resonance cell," *International Journal of Mass Spectrometry and Ion Processes*, vol. 142, no. 1–2, pp. 1–22, 1995.
32. N. Kaiser, C. Weisbrod, B. Webb, and J. Bruce, "Reduction of axial kinetic energy induced perturbations on observed cyclotron frequency," *Journal of the American Society for Mass Spectrometry*, vol. 19, no. 4, pp. 467–478, 2008.
33. M. Knobeler and W. K. P., "Suppression, amplification and application of the third harmonic of the cyclotron frequency in ion cyclotron resonance spectrometry," *International Journal of Mass Spectrometry and Ion Processes*, vol. 125, no. 2–3, pp. 127–134, 1993.
34. "Introduction to data acquisition." <http://www.ni.com/white-paper/3536/en/>. [Accessed 18 September 2014].
35. "PXI Express specification tutorial." <http://www.ni.com/white-paper/2876/en/>. [Accessed 18 September 2014].
36. S. Guan and P. R. Jones, "Personal computer based Fourier transform ion cyclotron resonance mass spectrometer," *Review of Scientific Instruments*, vol. 59, no. 12, pp. 2573–2576, 1988.
37. S. C. Beu and D. A. Laude, "Modular data system for selective wave-form excitation and trapping experiments in Fourier transform mass spectrometry," *Analytical Chemistry*, vol. 63, no. 19, pp. 2200–2203, 1991.
38. M. W. Senko, J. D. Canterbury, S. Guan, and A. G. Marshall, "A high-performance modular data system for Fourier transform ion cyclotron resonance mass spectrometry," *Rapid Communications in Mass Spectrometry*, vol. 10, no. 14, pp. 1839–1844, 1996.

Bibliography

39. M. A. Freitas, E. King, and S. D.-H. Shi, "Tool command language automation of the modular ion cyclotron data acquisition system (MIDAS) for data-dependent tandem fourier transform ion cyclotron resonance mass spectrometry," *Rapid Communications in Mass Spectrometry*, vol. 17, no. 4, pp. 363–370, 2003.
40. T. H. Mize, I. Taban, M. Duursma, M. Seynen, M. Konijnenburg, A. Vijftigschild, C. V. Doornik, G. V. Rooij, and R. M. A. Heeren, "A modular data and control system to improve sensitivity, selectivity, speed of analysis, ease of use, and transient duration in an external source FTICR-MS," *International Journal of Mass Spectrometry*, vol. 235, no. 3, pp. 243–253, 2004.
41. A. N. Vilkov, C. M. Gamage, A. S. Misharin, V. M. Doroshenko, D. A. Tolmachev, I. A. Tarasova, O. N. Kharybin, K. P. Novoselov, and M. V. Gorshkov, "Atmospheric pressure ionization permanent magnet Fourier transform ion cyclotron resonance mass spectrometry," *Journal of the American Society for Mass Spectrometry*, vol. 18, no. 8, pp. 1552–1558, 2007.
42. G. T. Blakney, C. L. Hendrickson, and A. G. Marshall, "Predator data station: A fast data acquisition system for advanced FT-ICR MS experiments," *International Journal of Mass Spectrometry*, vol. 306, no. 2–3, pp. 246–252, 2011. Special Issue: In Honor of Tino Gaumann.
43. "Data streaming architectures in PXI systems." <http://www.ni.com/white-paper/3221/en/>. [Accessed 18 September 2014].
44. G. T. Blakney, N. Kaiser, B. Ruddy, R. P. Rodgers, and A. G. Marshal, "Advanced experimental event sequences and hardware for highest mass accuracy of complex mixtures by FT-ICR MS," in *Proceeding of the 61st ASMS Conference on Mass Spectrometry and Allied Topics*, Minneapolis, MN, June 9–13 2013.
45. "NI PXIe-5122: 100 MS/s, 14-Bit Oscilloscope/Digitizer." <http://sine.ni.com/nips/cds/view/p/lang/en/nid/203623>. [Accessed 18 September 2014].
46. "NI PXIe-8135: 2.3 GHz Quad-Core PXI Express Controller." <http://sine.ni.com/nips/cds/view/p/lang/en/nid/210545>. [Accessed 18 September 2014].
47. "NI PXIe-1062Q: 8-Slot 3U PXI Express Chassis with AC — Up to 3 GB/s." <http://sine.ni.com/nips/cds/view/p/lang/en/nid/202664>. [Accessed 18 September 2014].
48. "High-speed data streaming: Programming and benchmarks." <http://www.ni.com/white-paper/5897/en/>. [Accessed 18 September 2014].
49. "An introduction to peer-to-peer streaming." <http://www.ni.com/white-paper/10801/en/>. [Accessed 18 September 2014].
50. M. V. Gorshkov, C. D. Masselon, G. A. Anderson, H. R. Udseth, R. Harkewicz, and R. D. Smith, "A dynamic ion cooling technique for FTICR mass spectrometry," *Journal of the American Society for Mass Spectrometry*, vol. 12, no. 11, pp. 1169–1173, 2001.

51. "Boston University Data Analysis." <http://www.bumc.bu.edu/ftms/buda/>. [Accessed 18 September 2014].
52. "The official home of the Python Programming Language." <https://www.python.org/>. [Accessed 18 September 2014].
53. A. Goloborodko, L. Levitsky, M. Ivanov, and M. Gorshkov, "Pyteomics—a Python framework for exploratory data analysis and rapid software prototyping in proteomics," *Journal of The American Society for Mass Spectrometry*, vol. 24, no. 2, pp. 301–304, 2013.
54. S. O'Callaghan, D. De Souza, A. Isaac, Q. Wang, L. Hodkinson, M. Olshansky, T. Erwin, B. Appelbe, D. Tull, U. Roessner, A. Bacic, M. McConville, and V. Likić, "PyMS: a Python toolkit for processing of gas chromatography-mass spectrometry (GC-MS) data. Application and comparative study of selected tools," *BMC Bioinformatics*, vol. 13, no. 1, 2012.
55. H. L. Röst, U. Schmitt, R. Aebersold, and L. Malmström, "pyOpenMS: A Python-based interface to the OpenMS mass-spectrometry algorithm library," *Proteomics*, vol. 14, no. 1, pp. 74–77, 2014.
56. O. Rübél, A. Greiner, S. Cholia, K. Louie, E. W. Bethel, T. R. Northen, and B. P. Bowen, "OpenMSI: A high-performance web-based platform for mass spectrometry imaging," *Analytical Chemistry*, vol. 85, no. 21, pp. 10354–10361, 2013.
57. P. Kiefer, U. Schmitt, and J. A. Vorholt, "emZed: an open source framework in Python for rapid and interactive development of LC/MS data analysis workflows," *Bioinformatics*, vol. 29, no. 7, pp. 963–964, 2013.
58. "Scientific Computing Tools for Python." <http://www.scipy.org/about.html>. [Accessed 18 September 2014].
59. M. Frigo and S. Johnson, "The design and implementation of FFTW3," *Proceedings of the IEEE*, vol. 93, pp. 216–231, Feb 2005.
60. F. James and M. Roos, "MINUIT — a system for function minimization and analysis of the parameter errors and correlations," *Computer Physics Communications*, vol. 10, no. 6, pp. 343–367, 1975.
61. S. D.-H. Shi, J. J. Drader, M. A. Freitas, C. L. Hendrickson, and A. G. Marshall, "Comparison and interconversion of the two most common frequency-to-mass calibration functions for Fourier transform ion cyclotron resonance mass spectrometry," *International Journal of Mass Spectrometry*, vol. 195–196, pp. 591 – 598, 2000.
62. E. B. Ledford, D. L. Rempel, and M. L. Gross, "Space charge effects in Fourier transform mass spectrometry. II. Mass calibration," *Analytical Chemistry*, vol. 56, no. 14, pp. 2744–2748, 1984.

Bibliography

63. T. J. Francl, M. G. Sherman, R. L. Hunter, M. J. Locke, W. D. Bowers, and R. T. M. Jr., "Experimental determination of the effects of space charge on ion cyclotron resonance frequencies," *International Journal of Mass Spectrometry and Ion Processes*, vol. 54, no. 1–2, pp. 189–199, 1983.
64. M. Gorshkov, D. Good, Y. Lyutvinskiy, H. Yang, and R. Zubarev, "Calibration function for the orbitrap FTMS accounting for the space charge effect," *Journal of the American Society for Mass Spectrometry*, vol. 21, no. 11, pp. 1846–1851, 2010.
65. M. B. Comisarow and J. D. Melka, "Error estimates for finite zero-filling in Fourier transform spectrometry," *Analytical Chemistry*, vol. 51, no. 13, pp. 2198–2203, 1979.
66. A. Serreqi and M. B. Comisarow, "Frequency interpolation of discrete, apodized, magnitude lineshapes," *Applied Spectroscopy*, vol. 41, no. 2, pp. 288–295, 1987.
67. A. Tolmachev, C. Masselon, G. Anderson, H. Udseth, and R. Smith, "Frequency shifts due to the interference of resolved peaks in magnitude-mode Fourier-transform ion cyclotron resonance mass spectra," *Journal of the American Society for Mass Spectrometry*, vol. 13, no. 4, pp. 387–401, 2002.
68. K. O. Zhurov, A. N. Kozhinov, and Y. O. Tsybin, "Evaluation of high-field orbitrap Fourier transform mass spectrometer for petroleomics," *Energy & Fuels*, vol. 27, no. 6, pp. 2974–2983, 2013.
69. L. S. Brown and G. Gabrielse, "Geonium theory: Physics of a single electron or ion in a penning trap," *Rev. Mod. Phys.*, vol. 58, pp. 233–311, 1986.
70. M. E. Belov, R. Zhang, E. F. Strittmatter, D. C. Prior, K. Tang, and R. D. Smith, "Automated gain control and internal calibration with external ion accumulation capillary liquid chromatography-electrospray ionization-Fourier transform ion cyclotron resonance," *Analytical Chemistry*, vol. 75, no. 16, pp. 4195–4205, 2003.
71. J. E. P. Syka, J. A. Marto, D. L. Bai, S. Horning, M. W. Senko, J. C. Schwartz, B. Ueberheide, B. Garcia, S. Busby, T. Muratore, J. Shabanowitz, and D. F. Hunt, "Novel linear quadrupole ion trap FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications," *Journal of Proteome Research*, vol. 3, no. 3, pp. 621–626, 2004.
72. A. Makarov, E. Denisov, A. Kholomeev, W. Balschun, O. Lange, K. Strupat, and S. Horning, "Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer," *Analytical Chemistry*, vol. 78, no. 7, pp. 2113–2120, 2006.
73. A. Makarov, E. Denisov, O. Lange, and S. Horning, "Dynamic range of mass accuracy in LTQ orbitrap hybrid mass spectrometer," *Journal of the American Society for Mass Spectrometry*, vol. 17, no. 7, pp. 977–982, 2006.

74. L.-K. Zhang, D. Rempel, B. N. Pramanik, and M. L. Gross, "Accurate mass measurements by Fourier transform mass spectrometry," *Mass Spectrometry Reviews*, vol. 24, no. 2, pp. 286–309, 2005.
75. M. L. Easterling, T. H. Mize, and I. J. Amster, "Routine parts-per-million mass accuracy for high mass ions: space-charge effects in MALDI FT-ICR," *Analytical Chemistry*, vol. 71, no. 3, pp. 624–632, 1999.
76. R. Burton, K. Matuszak, C. Watson, and J. Eyler, "Exact mass measurements using a 7 Tesla Fourier transform ion cyclotron resonance mass spectrometer in a good laboratory practices-regulated environment," *Journal of the American Society for Mass Spectrometry*, vol. 10, no. 12, pp. 1291–1297, 1999.
77. C. Masselon, A. Tolmachev, G. Anderson, R. Harkewicz, and R. Smith, "Mass measurement errors caused by "local" frequency perturbations in FTICR mass spectrometry," *Journal of the American Society for Mass Spectrometry*, vol. 13, no. 1, pp. 99–106, 2002.
78. P. Taylor and I. Amster, "Space charge effects on mass accuracy for multiply charged ions in ESI-FTICR," *International Journal of Mass Spectrometry*, vol. 222, no. 1–3, pp. 351–361, 2003.
79. D. C. Muddiman and A. L. Oberg, "Statistical evaluation of internal and external mass calibration laws utilized in Fourier transform ion cyclotron resonance mass spectrometry," *Analytical Chemistry*, vol. 77, no. 8, pp. 2406–2414, 2005.
80. R. Wong and I. Amster, "Sub part-per-million mass accuracy by using stepwise-external calibration in Fourier transform ion cyclotron resonance mass spectrometry," *Journal of the American Society for Mass Spectrometry*, vol. 17, no. 12, pp. 1681–1691, 2006.
81. R. L. Wong and I. J. Amster, "Experimental evidence for space-charge effects between ions of the same mass-to-charge in Fourier-transform ion cyclotron resonance mass spectrometry," *International Journal of Mass Spectrometry*, vol. 265, no. 2–3, pp. 99–105, 2007. Jean H. Futrell Honour Issue.
82. D. Smith, A. Kharchenko, M. Konijnenburg, I. Klinkert, L. Paša-Tolić, and R. Heeren, "Advanced mass calibration and visualization for FT-ICR mass spectrometry imaging," *Journal of The American Society for Mass Spectrometry*, vol. 23, no. 11, pp. 1865–1872, 2012.
83. A. Kharchenko, G. Vladimirov, R. Heeren, and E. Nikolaev, "Performance of orbitrap mass analyzer at various space charge and non-ideal field conditions: Simulation approach," *Journal of The American Society for Mass Spectrometry*, vol. 23, no. 5, pp. 977–987, 2012.
84. L. Chen, C. E. Cottrell, and A. G. Marshall, "Effect of signal-to-noise ratio and number of data points upon precision in measurement of peak amplitude, position and width in Fourier transform spectrometry," *Chemometrics and Intelligent Laboratory Systems*, vol. 1, no. 1, pp. 51–58, 1986.

Bibliography

85. M. B. Comisarow and A. G. Marshall, "Selective-phase ion cyclotron resonance spectroscopy," *Canadian Journal of Chemistry*, vol. 52, no. 10, pp. 1997–1999, 1974.
86. E. C. Craig, I. Santos, A. G. Marshall, and N. M. M. Nibbering, "Dispersion versus absorption (DISPA) method for automatic phasing of Fourier transform ion cyclotron resonance mass spectra," *Rapid Communications in Mass Spectrometry*, vol. 1, no. 2, pp. 33–37, 1987.
87. B. A. Vining, R. E. Bossio, and A. G. Marshall, "Phase correction for collision model analysis and enhanced resolving power of Fourier transform ion cyclotron resonance mass spectra," *Analytical Chemistry*, vol. 71, no. 2, pp. 460–467, 1999.
88. S. C. Beu, G. T. Blakney, J. P. Quinn, C. L. Hendrickson, and A. G. Marshall, "Broadband phase correction of FT-ICR mass spectra via simultaneous excitation and detection," *Analytical Chemistry*, vol. 76, no. 19, pp. 5756–5761, 2004.
89. F. Xian, C. L. Hendrickson, G. T. Blakney, S. C. Beu, and A. G. Marshall, "Automated broadband phase correction of Fourier transform ion cyclotron resonance mass spectra," *Analytical Chemistry*, vol. 82, no. 21, pp. 8807–8812, 2010.
90. Y. Qi, M. P. Barrow, H. Li, J. E. Meier, S. L. Van Orden, C. J. Thompson, and P. B. O'Connor, "Absorption-mode: The next generation of Fourier transform mass spectra," *Analytical Chemistry*, vol. 84, no. 6, pp. 2923–2929, 2012.
91. F. Xian, Y. E. Corilo, C. L. Hendrickson, and A. G. Marshall, "Baseline correction of absorption-mode Fourier transform ion cyclotron resonance mass spectra," *International Journal of Mass Spectrometry*, vol. 325–327, pp. 67–72, 2012. Eugene N. Nikolaev 65th Birthday Honor Issue.
92. D. P. A. Kilgour, R. Wills, Y. Qi, and P. B. O'Connor, "Autophaser: An algorithm for automated generation of absorption mode spectra for FT-ICR MS," *Analytical Chemistry*, vol. 85, no. 8, pp. 3903–3911, 2013.
93. O. Lange, "Methods and apparatus for producing a mass spectrum," US Patent 2011 0240841 A1.
94. O. Lange, E. Damoc, A. Wiegand, and A. Makarov, "Enhanced Fourier transform for Orbitrap mass spectrometry," in *Proceeding of the 59th ASMS Conference on Mass Spectrometry and Allied Topics*, Denver, CO, June 5–9 2011.
95. R. T. Hilger, P. J. Wyss, R. E. Santini, and S. A. McLuckey, "Absorption mode Fourier transform electrostatic linear ion trap mass spectrometry," *Analytical Chemistry*, vol. 85, no. 17, pp. 8075–8079, 2013.
96. O. Lange, E. Damoc, A. Wiegand, and A. Makarov, "Enhanced Fourier transform for Orbitrap mass spectrometry," in *Proceeding of the 59th ASMS Conference on Mass Spectrometry and Allied Topics*, Denver, CO, June 5–9 2011.

97. A. Rahbee, "High-resolution mass spectrometry using the maximum entropy method," *International Journal of Mass Spectrometry and Ion Processes*, vol. 72, no. 1–2, pp. 3–13, 1986.
98. J. E. Meier and A. G. Marshall, "Bayesian versus Fourier spectral analysis of ion cyclotron resonance time-domain signals," *Analytical Chemistry*, vol. 62, no. 2, pp. 201–208, 1990.
99. J. E. Meier and A. G. Marshall, "Pure absorption-mode spectra from Bayesian maximum entropy analysis of ion cyclotron resonance time-domain signals," *Analytical Chemistry*, vol. 63, no. 6, pp. 551–560, 1991.
100. J. Loo, M. D. Krahling, and F. T.C., "Accurate ion abundance measurements in ion cyclotron resonance mass spectrometry by linear prediction," *Rapid Communications in Mass Spectrometry*, vol. 4, no. 8, pp. 297–299, 1990.
101. S. Guan and A. G. Marshall, "Linear prediction Cholesky decomposition vs Fourier transform spectral analysis for ion cyclotron resonance mass spectrometry," *Analytical Chemistry*, vol. 69, no. 6, pp. 1156–1162, 1997.
102. T. C. Farrar, J. W. Elling, and M. D. Krahling, "Application of linear prediction to Fourier transform ion cyclotron resonance signals for accurate relative ion abundance measurements," *Analytical Chemistry*, vol. 64, no. 22, pp. 2770–2774, 1992.
103. R. Grothe, "Estimation of ion cyclotron resonance parameters in Fourier transform mass spectrometry," US Patent 2013 8431886 B2.
104. R. Grothe, "Mass spectrometry systems," US Patent 2013 0013274 A1.
105. E. N. Nikolaev, R. M. A. Heeren, A. M. Popov, A. V. Pozdnev, and K. S. Chingin, "Realistic modeling of ion cloud motion in a fourier transform ion cyclotron resonance cell by use of a particle-in-cell approach," *Rapid Communications in Mass Spectrometry*, vol. 21, no. 22, pp. 3527–3546, 2007.
106. A. G. Marshall, M. B. Comisarow, and G. Parisod, "Relaxation and spectral line shape in Fourier transform ion resonance spectroscopy," *The Journal of Chemical Physics*, vol. 71, no. 11, pp. 4434–4444, 1979.
107. E. N. Nikolaev, I. A. Boldin, R. Jertz, and G. Baykut, "Initial experimental characterization of a new ultra-high resolution FTICR cell with dynamic harmonization," *Journal of The American Society for Mass Spectrometry*, vol. 22, no. 7, pp. 1125–1133, 2011.
108. Z. Starcuk, K. Bartusek, and Z. Starcuk, "First-data-point problem and the baseline distortion in Fourier-transform NMR spectroscopy with simultaneous sampling," *Journal of Magnetic Resonance, Series A*, vol. 108, no. 2, pp. 177–188, 1994.
109. G. C. McAlister, E. L. Huttlin, W. Haas, L. Ting, M. P. Jedrychowski, J. C. Rogers, K. Kuhn, I. Pike, R. A. Grothe, J. D. Blethrow, and S. P. Gygi, "Increasing the multiplexing capacity

Bibliography

- of TMTs using reporter ion isotopologues with isobaric masses," *Analytical Chemistry*, vol. 84, no. 17, pp. 7469–7478, 2012.
110. R. Roy, B. Sumpter, G. Pfeffer, S. Gray, and D. Noid, "Novel methods for spectral analysis," *Physics Reports*, vol. 205, no. 3, pp. 109–152, 1991.
111. V. Mandelshtam, "FDM: the filter diagonalization method for data processing in NMR experiments," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 38, no. 2, pp. 159–196, 2001.
112. D. Neuhauser, "Bound state eigenfunctions from wave packets: Time to energy resolution," *The Journal of Chemical Physics*, vol. 93, no. 4, 1990.
113. M. R. Wall and D. Neuhauser *The Journal of Chemical Physics*, vol. 102, no. 20, 1995.
114. V. A. Mandelshtam and H. S. Taylor, "A low-storage filter diagonalization method for quantum eigenenergy calculation or for spectral analysis of time signals," *The Journal of Chemical Physics*, vol. 106, no. 12, 1997.
115. V. A. Mandelshtam and H. S. Taylor, "Harmonic inversion of time signals and its applications," *The Journal of Chemical Physics*, vol. 107, no. 17, 1997.
116. V. A. Mandelshtam and H. S. Taylor, "Erratum: "harmonic inversion of time signals and its applications" [J. Chem. Phys. 107, 6756 (1997)]," *The Journal of Chemical Physics*, vol. 109, no. 10, 1998.
117. H. Hu, Q. N. Van, V. A. Mandelshtam, and A. Shaka, "Reference deconvolution, phase correction, and line listing of NMR spectra by the 1D filter diagonalization method," *Journal of Magnetic Resonance*, vol. 134, no. 1, pp. 76–87, 1998.
118. J. W. Pang, T. Dieckmann, J. Feigon, and D. Neuhauser, "Extraction of spectral information from a short-time signal using filter-diagonalization: Recent developments and applications to semiclassical reaction dynamics and nuclear magnetic resonance signals," *The Journal of Chemical Physics*, vol. 108, no. 20, 1998.
119. K. Aizikov and P. B. O'Connor, "Use of the filter diagonalization method in the study of space charge related frequency modulation in Fourier transform ion cyclotron resonance mass spectrometry," *Journal of the American Society for Mass Spectrometry*, vol. 17, no. 6, pp. 836–843, 2006.
120. K. Aizikov, R. Mathur, and P. B. O'Connor, "The spontaneous loss of coherence catastrophe in Fourier transform ion cyclotron resonance mass spectrometry," *Journal of the American Society for Mass Spectrometry*, vol. 20, no. 2, pp. 247–256, 2009.
121. A. G. Marshall and S. Guan, "Advantages of high magnetic field for Fourier transform ion cyclotron resonance mass spectrometry," *Rapid Communications in Mass Spectrometry*, vol. 10, no. 14, pp. 1819–1823, 1996.

122. N. Karabacak, M. Easterling, N. Agar, and J. Agar, "Transformative effects of higher magnetic field in fourier transform ion cyclotron resonance mass spectrometry," *Journal of the American Society for Mass Spectrometry*, vol. 21, no. 7, pp. 1218–1222, 2010.
123. N. Kaiser, C. Weisbrod, J. Quinn, G. T. Blakney, S. Beu, T. Chen, C. L. Hendrickson, and M. A. G., "Development of an FT-ICR mass spectrometer in preparation for 21 Telsa," in *Proceeding of the 62nd ASMS Conference on Mass Spectrometry and Allied Topics*, Baltimore, MD, June 15–19 2014.
124. J. E. Bruce, G. A. Anderson, C.-Y. Lin, M. Gorshkov, A. L. Rockwood, and R. D. Smith, "A novel high-performance Fourier transform ion cyclotron resonance cell for improved biopolymer characterization," *Journal of Mass Spectrometry*, vol. 35, no. 1, pp. 85–94, 2000.
125. S. Kim, M. C. Choi, M. Hur, H. S. Kim, J. S. Yoo, C. L. Hendrickson, and A. G. Marshall, "The 'hybrid cell': a new compensated infinity cell for larger radius ion excitation in Fourier transform ion cyclotron resonance mass spectrometry," *Rapid Communications in Mass Spectrometry*, vol. 22, no. 9, pp. 1423–1429, 2008.
126. A. Brustkern, D. Rempel, and M. Gross, "An electrically compensated trap designed to eighth order for FT-ICR mass spectrometry," *Journal of the American Society for Mass Spectrometry*, vol. 19, no. 9, pp. 1281–1285, 2008.
127. C. R. Weisbrod, N. K. Kaiser, G. E. Skulason, and J. E. Bruce, "Trapping ring electrode cell: A FTICR mass spectrometer cell for improved signal-to-noise and resolving power," *Analytical Chemistry*, vol. 80, no. 17, pp. 6545–6553, 2008.
128. A. Tolmachev, E. Robinson, S. Wu, H. Kang, N. Lourette, L. Paša-Tolić, and R. Smith, "Trapped-ion cell with improved dc potential harmonicity for FT-ICR MS," *Journal of the American Society for Mass Spectrometry*, vol. 19, no. 4, pp. 586–597, 2008.
129. A. M. Brustkern, D. L. Rempel, and M. L. Gross, "Ion behavior in an electrically compensated ion cyclotron resonance trap," *International Journal of Mass Spectrometry*, vol. 300, no. 2–3, pp. 143–148, 2011. John Fenn Honor Issue.
130. A. V. Tolmachev, E. W. Robinson, R. D. Smith, F. E. L. III, J. H. Futrell, and L. Paša-Tolić, "A conceptual approach for FT-ICR cell harmonization utilizing external shim electrodes," *International Journal of Mass Spectrometry*, vol. 325–327, pp. 45–50, 2012. Eugene N. Nikolaev 65th Birthday Honor Issue.
131. Y. Pan, D. P. Ridge, J. Wronka, A. L. Rockwood, and A. G. Marshall, "Resolution improvement by using harmonic detection in an ion cyclotron resonance mass spectrometer," *Rapid Communications in Mass Spectrometry*, vol. 1, no. 7-8, pp. 120–121, 1987.
132. Y. Pan, D. Ridge, and A. L. Rockwood, "Harmonic signal enhancement in ion cyclotron resonance mass spectrometry using multiple electrode detection," *International Journal of Mass Spectrometry and Ion Processes*, vol. 84, no. 3, pp. 293–304, 1988.

Bibliography

133. E. N. Nikolaev, M. V. Gorshkov, A. V. Mordehai, and V. L. Talrose, "Ion cyclotron resonance signal-detection at multiples of the cyclotron frequency," *Rapid Communications in Mass Spectrometry*, vol. 4, no. 5, pp. 144–146, 1990.
134. A. S. Hebert, A. L. Richards, D. J. Bailey, A. Ulbrich, E. E. Coughlin, M. S. Westphall, and J. J. Coon, "The one hour yeast proteome," *Molecular & Cellular Proteomics*, vol. 13, no. 1, pp. 339–347, 2014.
135. Y. O. Tsybin, A. N. Kozhinov, and K. O. Nagornov, "A device and method for ion cyclotron resonance mass spectrometry," PCT/IB2014/060709 (pending).
136. K. O. Zhurov, A. N. Kozhinov, and Y. O. Tsybin, "Hexagonal class representation for fingerprinting and facile comparison of petroleomic samples," *Analytical Chemistry*, vol. 85, no. 11, pp. 5311–5315, 2013.

Curriculum vitae

Anton N. Kozhinov

Biomolecular Mass Spectrometry Laboratory
École Polytechnique Fédérale de Lausanne
EPFL SB ISIC LSMB, 1015 Lausanne, Switzerland

telephone: +41 (0)21 693 93 80
mobile: +41 (0)76 268 86 38
e-mail: anton.kozhinov@phystech.edu

Education and Research

- 2010-2014 PhD Student in Chemistry and Chemical Engineering at Biomolecular Mass Spectrometry Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.
Thesis director: Prof. Yury O. Tsybin
Specialization: Fourier transform mass spectrometry
- 2004-2010 MSc in Applied Mathematics and Physics at Moscow Institute of Physics and Technology (PhysTech), Moscow, Russia.
Graduated with GPA: 4.94/5.00

Fellowships and Awards

- 2013 SCNAT/SCS Chemistry Travel Award, Swiss Academy of Sciences.
2013 IJMS Best Student Paper Award, International Journal of Mass Spectrometry, presented at the 61st ASMS Conference, Minneapolis, MN, USA.
2008 Fellowship from Nix Corporation, Moscow, Russia.
2007 First degree diploma for the presentation at the Scientific Conference of Photochemistry Center of Russian Academy of Sciences, Moscow, Russia.
2006, 2007 Fellowship from the Foundation for Developments in Innovations and Education in the field of Natural Sciences, Moscow, Russia. (twice)
2004 Award of the President of Mari El Republic of Russian Federation.
2004 Golden Medal Award for course studies excellence.
2002-2004 Prize-winner of Regional & University Olympiads in Physics & Mathematics.

Publications

Patent applications: **2** (*pending*);

Peer-reviewed articles, published: **11**; submitted for publication: **2**; in preparation: **8**

Conference presentations

Talks: **4**, including at *Am. Soc. Mass Spectrom. Conf.*: **1**; *Int. Mass Spectrom. Conf.*: **1**

Posters: **7**, including at *Am. Soc. Mass Spectrom. Conf.*: **3**

Teaching assistance (practical work)

Fundamentals of Mass Spectrometry (EPFL/UNIL, MSc level); Structure Analysis (EPFL/UNIL, BSc level); Current Topics in Mass Spectrometry (EPFL, PhD level); Projects in Analytical Sciences (EPFL, MSc level)

Skills and Experience

Fourier transform mass spectrometry: FT-ICR MS, Orbitrap FTMS

Data processing: Python, C, bioinformatics

Design of measurement systems: PXI Express, NI-SCOPE, LabVIEW

Professional Associations

American Society for Mass Spectrometry
Swiss Chemical Society