Journal of Magnetic Resonance 197 (2009) 219–228

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr

Narrow peaks and high dimensionalities: Exploiting the advantages of random sampling

Krzysztof Kazimierczuk, Anna Zawadzka, Wiktor Koźmiński *

Department of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warszawa, Poland

ARTICLE INFO

Article history:
Received 7 November 2008
Revised 17 December 2008
Available online 13 January 2009

Keywords:
Multidimensional NMR spectroscopy
Sparse Multidimensional Fourier Transform
Random data sampling
Fast multidimensional data acquisition
Proteins

ABSTRACT

Level of artifacts in spectra obtained by Multidimensional Fourier Transform has been studied, considering randomly sampled signals of high dimensionality and long evolution times. It has been shown theoretically and experimentally, that this level is dependent on the number of time domain samples, but not on its relation to the number of points required in appropriate conventional experiment. Independence of the evolution time domain size (in the terms of both: dimensionality and evolution time reached), suggests that random sampling should be used rather to design new techniques with large time domain than to accelerate standard experiments. 5D HC(CC-TOCSY)CONH has been presented as the example of such approach. The feature of Multidimensional Fourier Transform, namely the possibility of calculating spectral values at arbitrary chosen frequency points, allowed easy examination of resulting spectrum. We present the example of such approach, referred to as Sparse Multidimensional Fourier Transform.

1. Introduction

Multidimensional NMR techniques are routine tools for determination of biomolecular structure. Separation of peaks in additional spectral dimensions, proposed in 1970 by Jeener [1], is exceptionally effective in studies of proteins, where the number of resonances exceeds hundreds. However, the conventional approach to this problem is limited by the need for fulfilling of the Nyquist Theorem. The Theorem determines the sampling rate to be twice higher than the highest frequency expected in the signal. In fact, this is an implicit limit for the maximum evolution time sampled in indirect dimensions with given number of samples (i.e. given experimental time). Because of this, even in 3D experiments FID signals are rarely measured long enough to reach the relaxation limits in indirectly sampled dimensions. For spectra of higher dimensionality it is practically impossible. Nowadays, this problem is getting relatively important comparing to the problems of sensitivity which are being solved by hardware improvements like magnets of higher fields or cryogenically cooled probes. Moreover, it should be noted that at higher fields the required Nyquist rate grows together with spectral width.

Truncated time domain signal results in broadened spectral peaks, according to Fourier Uncertainty Principle [2]. The problem of sampling-determined line-widths arises with the number of dimensions because Nyquist Theorem has to be fulfilled in each of time dimensions separately.

Many techniques have been invented to overcome the sampling barrier for spectral resolution. Usually, their goal is to obtain the spectrum from sparse time domain data i.e. from the number of points lower than in conventional sampling schedule. Sparse data sets may be of different kind, requiring various methods of processing in order to obtain spectra. Some of these methods are: Back-Projection Reconstruction [3–5], Reduced Dimensionality [6–8] and Multi-Way Decomposition [9] employing radial sampling, Polar Fourier Transform with radial [10–12] or concentric ring sampling [13], Maximum Entropy Methods [14,15], Hyperdimensional Spectroscopy [16] and Multidimensional Decomposition [17] for sampling using points placed randomly (often on Cartesian grid). Measurements of multidimensional spectra in a single scan where also presented, employing spatial encoding [18] or correlating data of lower dimensionality [19]. Enhanced spectral resolution can also be obtained from conventional data set. The example of such approach is Covariance Spectroscopy [20] (which is, however, not limited to conventional sampling) or Hadamard spectroscopy [21]. Truncated or block sampling may be processed by Filter Diagonalization Method [22].

An important issue regarding Nyquist Theorem should be considered when discussing various advanced methods for processing of non-linearly sampled signal. Although Fourier Transform (FT) of signal sampled at constant rate is prone to peak folding, it is not true that FT necessarily requires such data as an input. This is required by Fast Fourier Transform (FFT) algorithm which is one (but not only) possibility for calculating Fourier integral. Admittedly, FFT is very efficient numerically, but for multidimensional signal it requires sampling of time domain with points placed on

* Corresponding author. Fax: +48 22 822 59 96.
E-mail address: kozmin@chem.uw.edu.pl (W. Koźmiński).

1090-7807/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2009.01.003

knots of Cartesian grid. Each dimension is sampled and transformed separately, which is very fast in terms of computing time but not in terms of spectrometer time. That is because sampling points in time space that have the same t_i coordinate provide identical information about frequency dimension ω_i. Thus, for k-dimensional conventional experiment with N points in each dimension, the total experiment time t_{exp} increases exponentially with number of dimensions:

$$t_{exp} = N^{k-1}2^{k-1}rt \tag{1}$$

where rt is the repetition time of the direct measurement of FID and 2^{k-1} represents number of data sets required by quadrature detection rules.

In general, however, the numerical approximation of multidimensional integral can be performed using any kind of samples, even not suitable for FFT. We have proposed the off-grid random sampling, which is in our opinion the best option for non-linear sampling, i.e. produces Point Spread Function (PSF) of the best signal-to-artifact ratio [23,24]. The time domain signal is sampled evenly and each point differs from the others in values of all its coordinates (or, in perfect case, there is infinitesimally small probability that it will not differ). This means, that information about the frequencies of multidimensional correlation signals becomes more and more emphasized with each sampling point i.e. aliasing is suppressed in all directions.

The influence of sampling grid on the resulting spectra is worth deeper discussion, as random on-grid sampling is chosen very often. This is because on-grid sampling, allows separation of time (and, consequently, frequency) variables during processing. The separation is possible even if the sampling is sparse, as shown recently by Coggins and Zhou [25]. However, the fact that sparse sampling points are on knots of Cartesian grid results in problem known as primary and secondary aliasing [26]. Briefly, the primary aliasing manifests itself by the existence of folded peaks, which are at the same positions as folded peaks resulting from conventional sampling performed on the same grid but using full set of points. Moreover, they are of the same intensities as the "true" peak. In other words, while using on-grid random sampling the primary aliasing effect is neither avoided nor even attenuated. Secondary aliasing is the phenomenon of more complicated nature. It results in many false spectral peaks featuring lower intensity than original, "true" peak. Often, they are called artifacts or "sampling noise". The shape and distribution of this artifacts may be predicted by separating sampling schedule into combination of "component grids", which have smaller number of points but larger distance between them, and by predicting the effect of primary aliasing for each of them. The sum of this "component primary aliases" is called secondary aliasing. The magnitude of secondary aliasing effect depends on the number of "component grids" and the number of points occupying them. The bigger their number with smaller number of points on each, the more false peaks of lower intensity appear in the spectrum. In other words, the artifacts are more evenly spread over spectral domain (which is favorable). Of course, using more dense sampling grid during generation of on-grid random sampling pattern shifts away the false peaks originating from primary aliasing and makes the secondary aliasing artifact pattern more even.

The processing of on-grid data may be fast, even if the grid is quite dense, as it allows obtaining spectrum by sequential FFT. The example of such approach was shown recently in application to 4D HCCH-TOCSY experiment presented by Coggins and Zhou [25]. However, in application for really large evolution time domains (which in our opinion are the best target for random sampling), the sequential FFT may quickly become problematic because of hard disk space requirements. For example, in the case of 5D experiment with 256 samples in each dimension, the FFT has to handle with data matrix which is in order of terabytes. This is because standard FFT has to transform full spectral width in each dimension, with equal frequency intervals, of number equal to at least the number of time domain intervals (rounded up to 2^N).

However, simple modification of Fourier Transformation algorithm allows reducing of size of transformed spectrum to the regions of interest which is discussed in further part of this paper. We have recently proposed the Multidimensional Fourier Transformation (MFT) [11,23] as a general approach for processing of arbitrary sampled data sets. The essence of MFT may be presented as follows: instead of separating time variables by conventional, on-grid sampling followed by sequence of one-dimensional FFT 's:

$$f_1(t_1,t_2,\ldots,t_m) \overset{FT(t_1)}{\rightarrow} f_2(\omega_1,t_2,\ldots,t_m) \overset{FT(t_2)}{\rightarrow} \ldots \overset{FT(t_m)}{\rightarrow} S(\omega_1,\omega_2,\ldots,\omega_m) \tag{2}$$

we approximate really multidimensional Fourier integral:

$$f_1(t_1,t_2,\ldots,t_m)^{\mathrm{MFT}(t_1,t_2,\ldots,t_m)} \rightarrow S(\omega_1,\omega_2,\ldots,\omega_m) \tag{3}$$

by summation over whole time domain for each point of frequency domain:

$$S(\omega_1,\omega_2,\ldots,\omega_m) = \sum_{n=1}^{N} f_1(t_1^n,t_2^n,\ldots,t_m^n)\exp(-i_1\omega_1 t_1^n - i_2\omega_2 t_2^n \ldots - i_m\omega_m t_m^n) \tag{4}$$

where $\omega_1,\omega_2,\ldots,\omega_m$ are spectral frequencies, t_1^n, t_2^n, t_m^n are coordinates of time point number n, m is the number of spectral dimensions, i_1, i_2,...i_m are imaginary units for commutative Clifford algebra of m-th order and N is the number of time domain points.

2. Theory

Spectra of randomly sampled signals have many important features concerning artifact level, which are however not well understood and thus not fully exploited. Number of time domain points per unit of evolution time space is often used as a parameter evaluating the level of "sparseness" of sampling. Most often, the ratio Θ between the particular sampling density level and Nyquist density (i.e. points density in conventional experiment) is held into account:

$$\Theta = \frac{N}{t_{1\max}t_{2\max},\ldots,t_{n\max}sw_1sw_2,\ldots,sw_n} \tag{5}$$

where N is the total number of time points in sparse experiment, $t_{k\max}$ is maximum evolution time in dimension k and sw_k is respective spectral width. The bottom of above fraction is the number of points required to fulfill Nyquist Theorem for each sw_k while reaching $t_{k\max}$. Thus, Θ shows the sparseness of sampling in the terms of time saved comparing to conventional experiment. In most of the sampling schemes this parameter is related to the level and shape of artifacts. However, this is not the case of off-grid random sampling processed by MFT. In our recent works [23,24] we have shown that artifacts from random sampling appear even when sampling density is above Nyquist condition. Moreover, they do not origin from integration imperfections, but rather from breaking of conventional Sampling Theorem, which stands that function with finite spectrum may be fully specified by its values only if they are at least at Nyquist density, like in interlaced sampling schemes [27]. Even randomly sampled data at Nyquist density can result in artifact-free spectra by application of unorthogonal transform [26]. In the further part of this paper we are going to estimate the level of random sampling artifacts, and prove that it depends only on the number of time points, in the same way as thermal noise. First, we shall analyze the Fourier Transform of FID signal $f(t)$.

For one-dimensional $f(t)$:

$$S(\omega) = \int dt f(t) e^{-i\omega t} \tag{6}$$

For multidimensional $f(t)$:

$$S(\omega_1, \omega_2, \ldots, \omega_m) = \int dt_1 \int dt_2 \ldots \int dt_m f(t_1, t_2, \ldots, t_m) e^{-i_1\omega_1 t_1} e^{-i_2\omega_2 t_1} \ldots e^{-i_m\omega_m t} \tag{7}$$

For better analysis of Fourier Transform integral it is convenient to make use of Convolution Theorem:

$$FT[u(t)v(t)] = FT[u(t)] * FT[v(t)] \tag{8}$$

which means that Fourier Transform of product of two functions is convolution (*) of their Fourier Transforms.

In the above case of perfect continuous Transform of infinite, exponentially decaying signal one obtains spectrum consisting only of peaks of Lorentzian shapes. This is because simplest FID containing one correlation signal is the product of two functions of time $u(t)v(t)$:

decaying exponential function,

$$u(t) = e^{-R_1 t_1} e^{-R_2 t_2}, \ldots, e^{-R_m t_m} \tag{9}$$

whose Fourier Transformation is Lorentzian function:

$$\mathrm{Re}\{FT[u(t)]\} = \frac{R_1}{R_1^2 + \left(\frac{\omega_1}{2\pi}\right)^2} \frac{R_2}{R_2^2 + \left(\frac{\omega_2}{2\pi}\right)^2}, \cdots, \frac{R_m}{R_m^2 + \left(\frac{\omega_m}{2\pi}\right)^2} \tag{10}$$

and oscillatory function,

$$v(t) = e^{-i_1\Omega_1 t_1} e^{-i_2\Omega_2 t_2}, \ldots, e^{-i_m\Omega_m t_m} \tag{11}$$

whose Fourier Transformation gives multidimensional delta function:

$$FT[v(t)] = \delta(\omega_1 - \Omega_1, \omega_2 - \Omega_2, \ldots, \omega_m - \Omega_m) \tag{12}$$

Convolution of delta function and Lorentzian function gives spectrum consisting of Lorentzian-shaped peaks at positions of resonances. The effect of exponential decay on the spectrum can be easily extracted and evaluated independently of other modifications of function of time. Therefore, we will neglect the relaxation in further discussion. FID frequencies can also be neglected (set to 0) as they influence only position, not shape of spectral peak. Thus, constant non-decaying signal will be considered in studies of sampling effects presented below. Its Fourier Transform involving all of these effects is referred to as Point Spread Function.

Now, two unavoidable experimental limitations will be considered by employing Convolution Theorem: the influence of finite acquisition time and discrete sampling.

First, finite evolution time can be described by multiplying infinite FID by rectangular function $\prod(t_1, t_2, \ldots, t_m)$:

$$\prod(t_1, t_1, \ldots, t_m) = \begin{cases} 1 & \forall t_i < t_{i\max} \\ 0 & \exists t_i > t_{i\max} \end{cases} \tag{13}$$

where $t_{i\max}$ are maximum evolution times set in appropriate spectral dimensions. Its Fourier Transformation gives so called *sinc* function in each spectral dimension:

$$FT\left\{\prod(t)\right\} = \frac{\sin(\omega_1/t_{1\max})}{\omega_1/t_{1\max}} \frac{\sin(\omega_2/t_{2\max})}{\omega_2/t_{2\max}} \cdots \frac{\sin(\omega_m/t_{m\max})}{\omega_m/t_{m\max}} \tag{14}$$

This result is easy to interpret: truncation of signal leads to *sinc*-shaped peaks. Again, like in the case of Lorentzian line shape caused by exponential decay, this effect can be separated from other effects. Simulations of PSF presented in the further part of this work do not reveal *sinc*-shaped peaks despite finite maximum evolution times $t_{i\max}$ because the distance between spectral points is set to:

$$\Delta\nu_i = \frac{\Delta\omega_i}{2\pi} = \frac{n}{2t_{i\max}}, \quad \text{where } n = 1, 2, 3, \ldots \tag{15}$$

which, for signal of zero frequency, means that spectral values are calculated exactly on the peak, and for the zeroes of *sinc* function.

Then, sampling should be considered. Depending on the sampling scheme, sampling function which multiplies signal of interest, may be of different shape. In the conventional experiment sampling points are equally spaced and sampling function is so called *shah* function (plotted in Fig. 1a), namely:

$$III(t_1, t_2, \ldots, t_m) = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \cdots \sum_{n_m=1}^{N_m} \delta(t_1 - n_1\tau_1, t_2 - n_2\tau_2, \ldots, t_m - n_m\tau_m) \tag{16}$$

where $\tau_1, \tau_2, \ldots, \tau_m$ are distances between samples according to Nyquist Theorem.

Thus, multidimensional *shah* is a set of deltas placed at knots of Cartesian grid. Its Fourier transform is of similar shape (see Fig. 1b), namely:

$$FT\{III(t_1, t_2, \ldots, t_m)\} = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \cdots \sum_{n_m=1}^{N_m} \delta\left(\omega_1 - 2\pi\frac{n_1}{\tau_1}, \omega_2 - 2\pi\frac{n_2}{\tau_2}, \cdots, \omega_m - 2\pi\frac{n_m}{\tau_m}\right) \tag{17}$$

Note, that distances between peaks in spectral domain are inversely proportional to appropriate distances in time domain. Convolution of the frequency domain *shah* function (Eq. (17)) with perfect spectrum produces infinite number of its equivalent copies, which leads to signal aliasing, if distance between time points is greater than inverse spectral width in appropriate dimension. This is a direct consequence of already mentioned Nyquist Theorem.

Suitable sampling function for non-linear (i.e. with unequal sampling intervals) sampling, denoted $\widetilde{III}$, has to be defined in more general way:

$$\widetilde{III}(t_1, t_2, \ldots, t_m) = \sum_{n=1}^{N} \delta(t_1 - t_1^n, t_2 - t_2^n, \ldots, t_m - t_m^n) \tag{18}$$

where $t_1^n, t_2^n, \ldots, t_m^n$ are time coordinates of sample number n. We are going to discuss random sampling i.e. these coordinates will be random numbers of some distribution (uniform, Gaussian etc.). For graphical presentation – see Fig. 1c. Fourier Transform of $\widetilde{III}(t_1, t_2, \ldots, t_m)$ shown in Fig. 1d, is of different shape than that of $III(t_1, t_2, \ldots, t_m)$. It contains only one, dominating peak and infinite spectrum of artifacts (secondary aliases).

It is impossible to find general, analytical shape of PSF in the case of random sampling. There is, however, simple relation

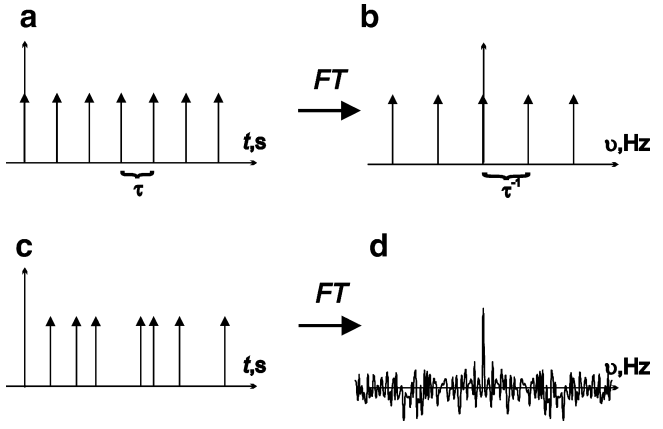

Fig. 1. Time (a and c) and frequency (b and d) domain representation of conventional (a and b) and random (c and d) sampling schemes.

between magnitude of sampling artifact at particular point of frequency domain and time domain points distribution. To find it, some simple, statistical laws have to be used [28]. In the further analysis only a cosine Fourier Transform will be considered, as it is easier to visualize than complex FT, providing the same conclusions for signal of zero frequency.

At particular point of spectral domain $\omega_1, \omega_2, \ldots, \omega_m$ the value of PSF (i.e. Fourier Integral of signal $f(t_1, t_2, \ldots, t_m) = 1$)is:

$$S(\omega_1, \omega_2, \ldots, \omega_m) = \sum_{n=1}^{N} \cos(\omega_1 t_1^n)\cos(\omega_2 t_2^n)\ldots\cos(\omega_m t_m^n) \tag{19}$$

where $t_1^n, t_2^n, \ldots, t_m^n$ are some random numbers generated using known Probability Density Functions (PDF) $f_1(t_1)$, $f_2(t_2)$,...$f_m(t_m)$. Thus, spectral values are sums of multidimensional cosine function of m random variables. Now, two simple observations can be made.

First, we shall consider cosines in the above sum: $y_i = g_i(t_i) = \cos(\omega_i t_i)$ as a new set of random variables. According to basic statistical laws, PDF $h_i(y_i)$ of monotonic function $g(t_i)$ of random variable t_i is equal to:

$$h_i(y_i) = \frac{f_i(g^{-1}(y_i))}{\left|\frac{dy_i}{dt_i}\right|} \tag{20}$$

Cosine is not monotonic, and thus has to be considered separately in monotonic parts (see Fig. 2). Because probability of finding y_i inside the interval $y, y + dy$ is equal to probability that t_i is inside one of the regions:

$$P(y_i \in (y, y + dy)) = \sum_{t:g(t)=y} P(t_i \in (t, t + dt)) \tag{21}$$

Thus, PDF for non-monotonic function is equal to:

$$h_i(y_i) = \sum_{t_i:g_i(t_i)=y} \frac{f_i(t_i)}{\left|\frac{dy_i}{dt_i}\right|} \tag{22}$$

PDF for variable t_i can be chosen arbitrary, depending on the shape of signal (decaying or not) and affects line shape in the same way as weighting function [24]. Thus, for simplicity, uniform PDF will be considered for generation of sampling points. We assume that t is randomized uniformly within the band (0, $t_{\max}$) with $t_{\max} \to \infty$. To avoid being imprecise, we also assume, that maximum evolution time is equal to $k_{\max}\frac{2\pi}{|\omega_i|}$ i.e. to whole number of periods. For the infinite signal this is true for every ω_i (only $k_{\max}$ differs between frequencies, which does not influence the results). The shape of PDF for generation of time points is thus:

$$f_i(t_i) = \frac{|\omega_i|}{2k_{\max}\pi} \quad \text{for } t_i \in \left(0, k_{\max}\frac{2\pi}{|\omega_i|}\right)$$
$$\text{otherwise}, \quad f_i(t_i) = 0 \tag{23}$$

To find PDF for cosine function of frequency ω_i and random time variable t_i we shall denote following statements:

$$y_i = \cos(\omega_i t_i) \tag{24a}$$

$$\left|\frac{dy_i}{dt_i}\right| = |\omega_i \sin(\omega_i t_i)| = |\omega_i|\sqrt{1 - y_i^2} \tag{24b}$$

Now, using the Eq. (22) one can obtain:

$$h_i(y_i) = \sum_{k=0}^{2k_{\max}} \frac{|\omega_i|}{2\pi k_{\max}|\omega_i \sin(\omega_i t_i)|} \tag{25}$$

Thus, using $|\sin(\omega_i t_i)| = \sqrt{1 - y_i^2}$:

$$h_i(y_i) = \sum_{k=0}^{2k_{\max}} \frac{1}{2\pi k_{\max}\sqrt{1 - y_i^2}} = 2k_{\max}\frac{1}{2\pi k_{\max}\sqrt{1 - y_i^2}} = \frac{1}{\pi\sqrt{1 - y_i^2}} \tag{26}$$

Obtained PDF is very simple. It is frequency independent, which means that values of all randomly sampled cosine terms in Eq. (19) are generated with the same PDF (see Fig. 3). Because evolution times $t_1, t_2, \ldots, t_n$ are randomized separately (i.e. they are independent variables), the PDF of product of cosines in Eq. (19) is simply the product of their PDFs:

$$h_{\text{total}}(y_1, y_2, \ldots, y_m) = \prod_{i=0}^{m} \frac{1}{\prod\sqrt{1 - y_i^2}} \tag{27}$$

Considering the above result, the estimation of PDF for $S(\omega_1, \omega_2, \ldots, \omega_m)$ from Eq. (19) is now easy. Spectral point $S(\omega_1, \omega_2, \ldots, \omega_m)$ is a sum of independent random numbers of the

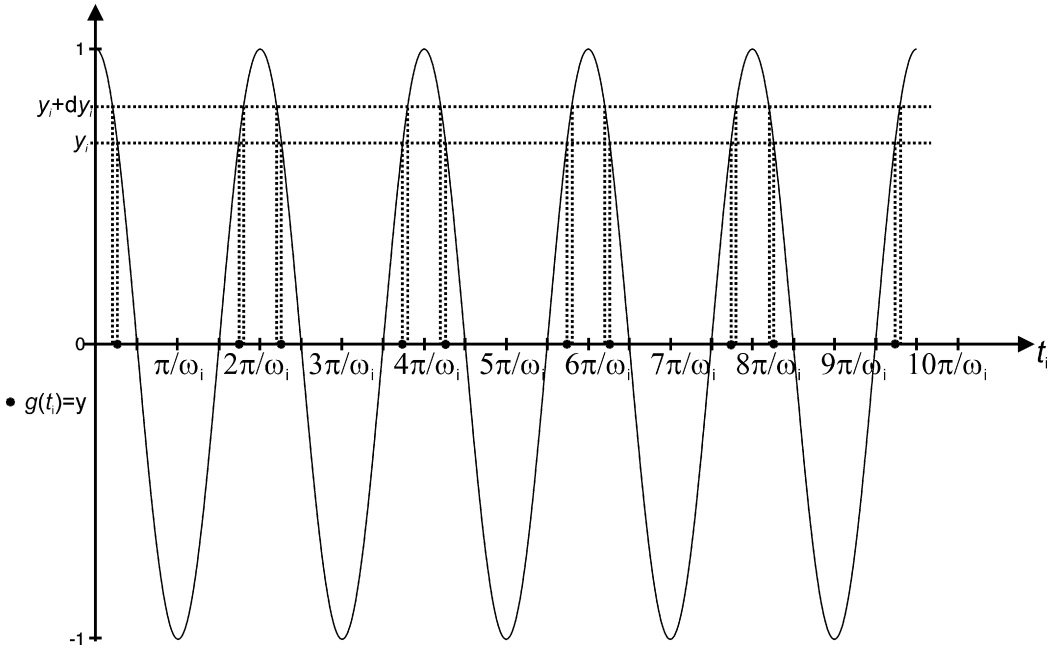

Fig. 2. Function $y_i = \cos(\omega_i t_i)$. Probability of generating y_i from the band (y_i, $y_i + dy_i$) by randomizing argument t_i is equal to the probability that t_i is inside one of the marked regions.

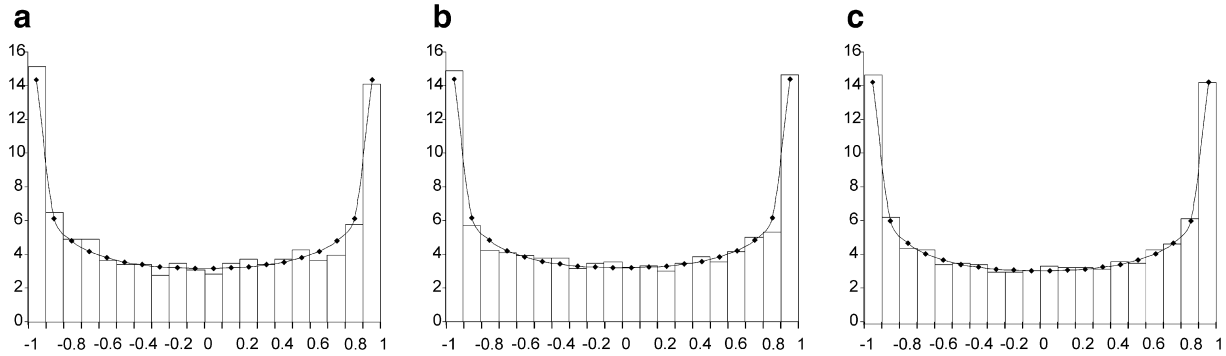

Fig. 3. Graphical representation of Probability Density Function $h_i(y_i) = \frac{1}{\pi\sqrt{1-y_i^2}}$ (line) and normalized histograms of values of $y_i = \cos(\omega_i t_i)$ obtained by randomizing t_i using uniform distribution and different frequencies: (a) 100 Hz (b) 1000 Hz (c) 10,000 Hz. Four thousand nine hundred six samples were used to generate histograms. Shape of histograms converges to that of Eq. (27) independently of frequency.

same PDFs $h_{\text{total}}(y_1, y_2, \ldots, y_m)$ Basic statistical law, known as Central Limit Theorem stands that in such case the $S(\omega_1, \omega_2, \ldots, \omega_m)$ is variable of normal distribution. Because mean value of all cosine terms is 0, then also distribution of $S(\omega_1, \omega_2, \ldots, \omega_m)$ reveals the same mean value of 0. Moreover, according to Central Limit Theorem, the variance of spectral value σ_S depends only on the number of summed samples N and their individual variance σ_n:

$$\sigma_S = \sigma_n/\sqrt{N} \tag{28}$$

In other words, intensity of spectral artifact originating from random sampling behaves in the same way as thermal noise. This explains the experimental fact reported before [23] that signal-to-artifact ratio is proportional to the $\sqrt{N}$ and may be also derived from similar proofs for one-dimensional signals published in fields different than NMR. For example, in the work of Tarczyński and Allay [29] the equations for expected value of spectrum estimator in weighted samples method (in our case there is no weighting i.e. $w(t) = 1$) and its standard deviation were derived. They are of the following shape (Eqs. (8) and (21) in [29]):

Expected value of spectrum estimator:

$$E\{X_{WS}(\omega)\} = X_W(\omega) \tag{29}$$

where $E\{X_{WS}(\omega)\}$ is the expected value of spectral estimator and $X_W(\omega)$ is the value of continuous FT of signal $x(t)$. This proves that estimator is unbiased.

Standard deviation of spectrum estimator:

$$\sigma_{WS}(\omega) = \sqrt{\frac{TE_{WS} - |X_W(\omega)|^2}{N}} \tag{30}$$

where T is the acquisition time of the signal, E_{WS} is the energy of the signal in the period T, and N is the number of samples.

For constant signal $f(t) = 1$ discussed above, we obtain:

$$E\{X_{WS}(\omega)\} = \int_0^T dt 1 \cdot \exp(-i\omega t) = \begin{cases} T & \text{for } \omega = 0 \\ 0 & \text{for } \omega = 0 \end{cases} \tag{31a}$$

$$E_{WS} = \int_0^T dt (x(t))^2 = T \tag{31b}$$

In the Eq. (31a) *sinc* was neglected, similarly as in proof above. Now, substitution of (31a) and (31b) to (30) gives:

$$\sigma_{WS} = \sqrt{\frac{T^2}{N}} \quad \text{for } \omega \neq 0 \text{ (artifact amplitude)} \tag{32}$$

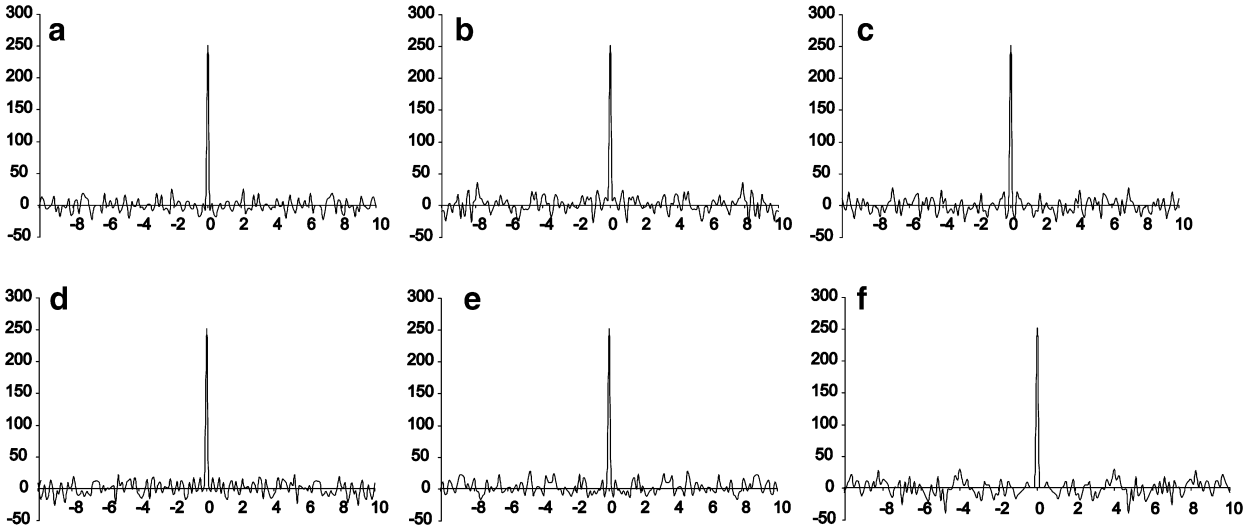

Fig. 4. Point Spread Functions for non-relaxing signal sampled using 250 randomly distributed time points and different maximum evolution times t_{max}: (a) t_{max} = 4 s; (b) t_{max} = 8 s; (c) t_{max} = 16 s; (d) t_{max} = 32 s; (e) t_{max} = 1 month; (f) t_{max} = 1 year. The magnitude of artifacts is constant, independently of number of time points per unit of time, and thus independently of the part of Nyquist Density used which is: (a) Θ = 3.12; (b) Θ = 1.56; (c) Θ = 0.78; (d) Θ = 0.39; (e) Θ = 4.8 10^{-6}; (f) Θ = 3.96 10^{-7}. The distance between spectral points is set to 0.125 Hz in order to zero the effect of *sinc*. Because of this digital resolution the effect of line narrowing can not be observed, but obviously exists.

a

number of time point	t_1, ms	t_2, ms	t_3, ms
1	142	377	236
2	22	46	113
3	282	39	270
4	400	209	66
5	392	139	280
...	.....	.....	.....
250	346	272	314

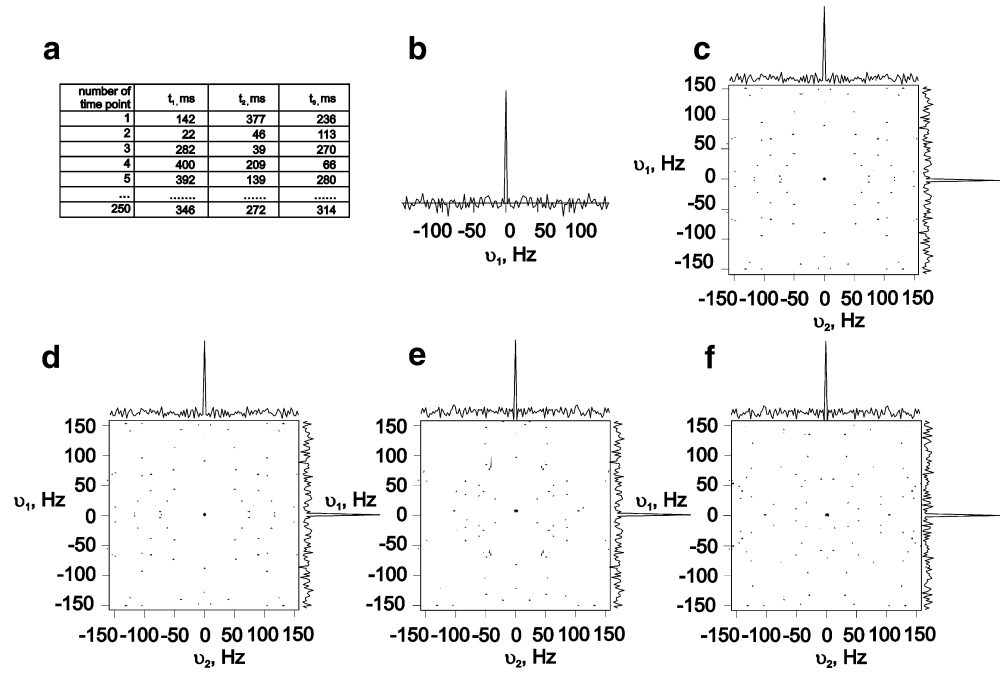

Fig. 5. Dimensionality and artifacts – simulations. (a) Sample of time domain points coordinates (number of points is 250); (b) 1D peak shape (PSF) in simulation employing first column (t_1); (c) 2D peak shape employing t_1 and t_2; (d–f) orthogonal cross-sections showing 3D peak shape employing t_1, t_2 and t_3. In all simulations the level of artifacts is similar, independently of number of dimensions and thus independently of the relative sampling density Θ, equal to 1.94, 0.015, and 0.00012, in the 1D, 2D and 3D simulation, respectively. Evolution time of 0.4 s, spectral width of 322.5 Hz and 129 spectral points were used in all dimensions. Because the distance between spectral points is 2.5 Hz the effect of convolution with *sinc* function is zeroed. Threshold is set at the 8% of peak intensity.

a

number of time point	t_1, ms	t_2, ms	t_3, ms
1	1,78820	24,53712	18,14699
2	2,89460	13,34076	15,06451
3	0,90547	17,08408	14,89350
4	3,67498	24,21573	8,908401
5	4,55924	7,50531	1,99194
...	.....	.....	.....
2500	5,01899	6,37636	3,74418

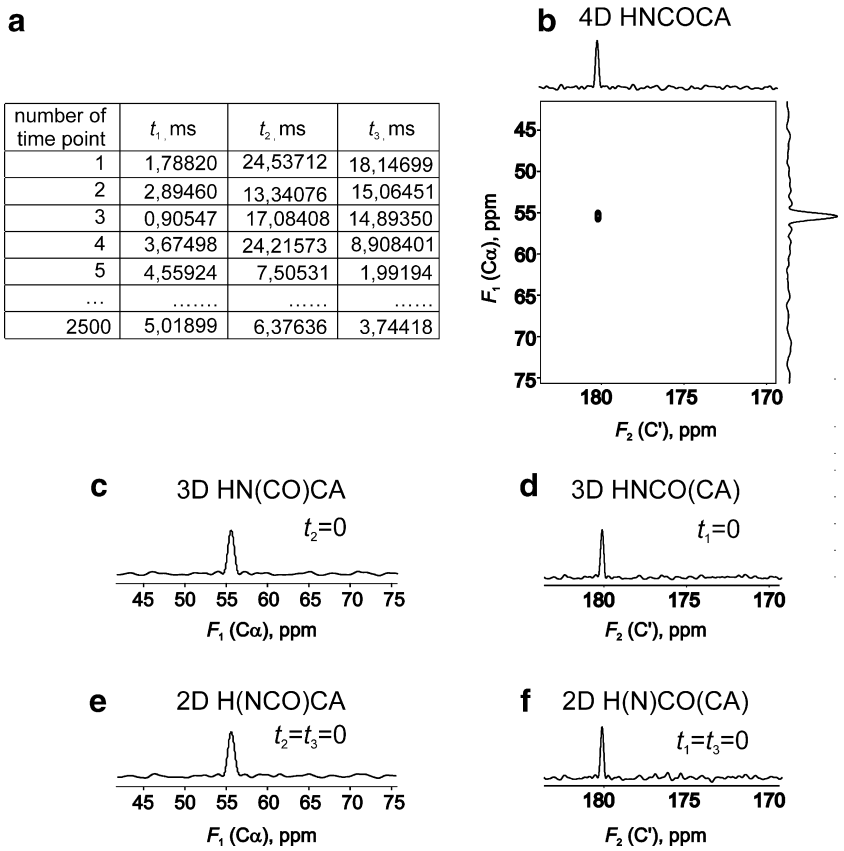

Fig. 6. Dimensionality and artifacts – 4D HNCOCA experiment. Cross-sections of A28 signal. (a) Sample of time domain points coordinates (number of points is 2500); (b) Cα-C′ 2D cross-section and 1D slices from 4D HNCOCA (Θ = 0.013); (c) Cα 1D cross-section from 3D HN(CO)CA (t_2 column in table in panel (a) was zeroed, Θ = 0.54); (d) C′ 1D cross-section from 3D HNCO(CA) (t_1 column in table in panel a) was zeroed, Θ = 0.97); (e) Cα 1D cross-section from 2D H(NCO)CA (t_2 and t_3 columns in table in panel a were zeroed, Θ = 33.33); (f) C′ 1D cross-section from 2D H(N)CO(CA) (t_1 and t_3 columns in table in panel a were zeroed, Θ = 59.52).

Signal-to-artifact ratio is therefore:

$$S/N = \frac{E\{X_{WS}(0)\}}{\sigma_{WS}(\omega)} = \sqrt{N} \tag{33}$$

It is noteworthy, that although both peak value $E\{X_{WS}(0)\}$ and artifact magnitude $\sigma_{WS}(\omega)$ depend on T, their ratio is independent of it.

Similar proof may be derived from Eq. (5) in the work of Dippé and Wold [30]. These authors, however, have chosen different approach – further they discussed infinite signal with constant sampling rate (which, obviously, makes all conclusions independent of number of samples, which is infinite). Small modification of their proof i.e. setting of finite sampling period X in their Eq. (6) leads to the dependence of signal-to-noise ratio on number of samples rather than sampling rate (which is more practical for NMR). Moreover, these authors have shown similar relations for different types of random sampling, as for example jittered and Poisson sampling, also used in NMR [31].

In fact, the conclusion about signal-to-artifact ratio is quite intuitive. The artifact magnitude in Point Spread Function of randomly sampled signal, at the frequency point ω is equal to the sum of values of $cos(\omega t_k)$ at sampling points t_k. The expected value of this sum does not depend on the sampled period and ω, as derived in Eq. (27). It does not depend on the dimensionality, too. Only increasing the number of points will change the value of this sum.

Above facts lead to quite surprising conclusions. The most important is that signal-to-artifact ratio in ideal experiment (without relaxation) employing random sampling does not depend on anything but the number of time domain samples. It is independent of the particular spectral width, maximum evolution times or number of dimensions (see Figs. 4 and 5). This means that with random sampling the resolution of obtained spectra is limited only by thermal noise level and transverse relaxation rate. Moreover, resolving spectrum by adding any number of dimensions does not cause increase of the artifact level.

In other words, dimensionality, spectral widths and maximum evolution times can be chosen arbitrary, with no effect on experiment time, which determines only signal-to-artifact and signal-to-noise ratios. In contradiction, in conventional sampling-limited experiment all these factors influence the experiment time.

The practical consequence of above relations is that it is beneficial to set the number of scans per point as low as possible (according to basic phase cycle) and increase number of random time domain points for given experimental duration. This approach causes both signal-to-artifact and signal-to-noise ratio to rise proportionally to $\sqrt{N}$, while increasing number of scans at the expense of reducing number of points improves only the latter.

Thus, the advantages of random sampling can be fully exploited using extended number of dimensions or ultra-long evolution times. This is in contradiction to the common way of employing random sampling to accelerate obtaining information from the same evolution space volume as in conventional experiments. The key feature of random sampling and MFT processing is ability of acquiring ultra-high-resolution spectra of high dimensionality in the same experiment duration, which is usually more interesting as it could provide more information on physical phenomena. Recently we have shown the example of ultra-high-resolution 3D spectra which enabled determination of small spin–spin couplings [32].

3. Sparse Multidimensional Fourier Transform (SMFT)

The improved resolution, being a great advantage, may cause data handling problems. Large number of spectral points has to be used in order to avoid missing narrow peaks between frequency domain points, i.e. the distance between spectral points should be smaller than expected line-width. This digital resolution determines the amount of data to be stored. For example three-dimensional spectrum containing 1024 real frequency points in each dimension has size of 4 GB (while using floating point variables). The problem is getting more serious with increasing the dimensionality of spectra: e.g. four-dimensional spectrum of the same resolution in each dimension would require four terabytes of disk space.

It is possible to solve this problem by exploiting the fact that peaks in high dimensional NMR spectra are rather dispersed, occupying only a small part of entire frequency domain. Therefore, Fourier transformation can be focused on the regions of interest, which may be determined using simpler experiments. At first the two- or three-dimensional experiment of very high-resolution should be performed (again, random sampling can be employed here). Then peak coordinates need to be extracted. On their basis, calculation of full 4 or higher dimensional spectrum may be reduced to the 1D, 2D or 3D regions of interest, featuring as high digital resolution as necessary. For example, positions of resonances of amide protons, nitrogens and carbonyl carbons, obtained from 3D HNCO spectrum, can determine regions of interest for processing of data from 5D HC(CC-TOCSY)CONH experiment (see flowchart given in

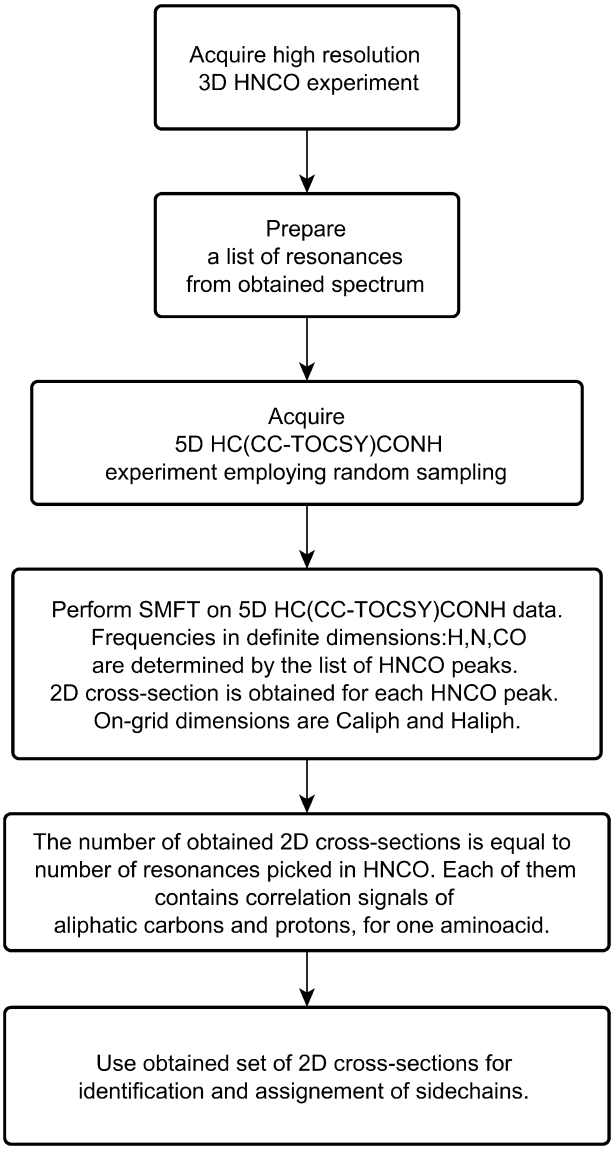

Fig. 7. Example of side-chain assignment procedure employing SMFT. Transformation of5DHC(CC-TOCSY) data on the base of 3D HNCO peak list results in set of 2D cross-sections.

Fig. 7). In this case, processing results with a set of 2D (C_{aliph}, H_{aliph}) cross-sections, one for each resonance from 3D HNCO spectrum. In fact, well known resonance assignment strategies employ similar way of thinking. For example, peak selection in 3D HNCO is usually performed using peak coordinates from ^{15}N HSQC.

Above approach, later referred to as Sparse Multidimensional Fourier Transform (SMFT) may be defined using MFT formula (Eq (4)), but with set of frequencies $(\omega_1, \omega_2, \ldots, \omega_m)$ chosen in an arbitrary way, e.g. some of them may be set according to knowledge obtained from other spectra. In conventional, non-sparse approach, $(\omega_1, \omega_2, \ldots, \omega_m)$ is one of spectral domain points laid out in a grid pattern. Its coordinates ω_i^{conv} are defined by:

$$\omega_i^{conv} = -\frac{sw_i}{2} + k_i \Delta v_i \tag{34}$$

where sw_i is spectral width in dimension i consisting of N_i points, $\Delta v_i = \frac{sw_i}{N_i - 1}$ is the distance between spectral points in this dimension, and $k_i \in \{0, 1, \ldots, N_i\}$. In SMFT frequency coordinates ω_i^{SMFT} may be arbitrary. In some dimensions they may be conventional ones, the same as in the Eq. (29). Others, however, may consist of arbitrary chosen values Ω_i, like peak coordinates known from simpler spectra. If the dimension i consists of conventional on-grid frequency values, it will be denoted as *on-grid* dimension. Otherwise it will be called *definite* dimension.

$$\omega_i^{SMFT} = \begin{cases} -\frac{sw_i}{2} + k\Delta v_i & \text{if } i \text{ is } \textit{on-grid} \text{ dimension} \\ \Omega_i & \text{if } i \text{ is } \textit{definite} \text{ dimension} \end{cases} \tag{35}$$

It is noteworthy, that such arbitrary trajectory in frequency domain can not be obtained by application of sequential FFT or any other processing method requiring on-grid frequency points.

The example of such approach is illustrated in Fig. 8 presenting 5D HC(CC-TOCSY)CONH as the set of 2D cross-sections, obtained in the manner described above. The separate 2D cross-sections are obtained for each amino acid residue, thus enabling identification and assignment of spin systems of all aliphatic side-chain proton and carbon nuclei. Such method of processing saves disk space, and does not exclude possibility of different choice of definite dimensions, or even full 5D processing using the same data. Recently, different example of 5D HC(CC-TOCSY)CONH experiment using APSY procedure has been presented [33]. In this case the 5D information is obtained from analysis of the set of projections obtained from automated radial sampling. It is noteworthy, that data obtained by radial sampling can also be processed by MFT, resulting in a full dimensional spectrum (identical to these obtained by Projection Reconstruction), or by SMFT resulting in the set of spectra. Example of similar sparse transformation of radially sampled signal was presented before [34,35]. However, it is much more convenient to use random than radial sampling, due to reduced artifact amplitude and ability to sample far below Nyquist condition in a case of the former.

4. Experimental

All spectra were recorded for 1.5 mM ^{13}C, ^{15}N-double labeled human ubiquitin in 9:1 H_2O/D_2O at pH = 4.5 at 298 K on a Varian

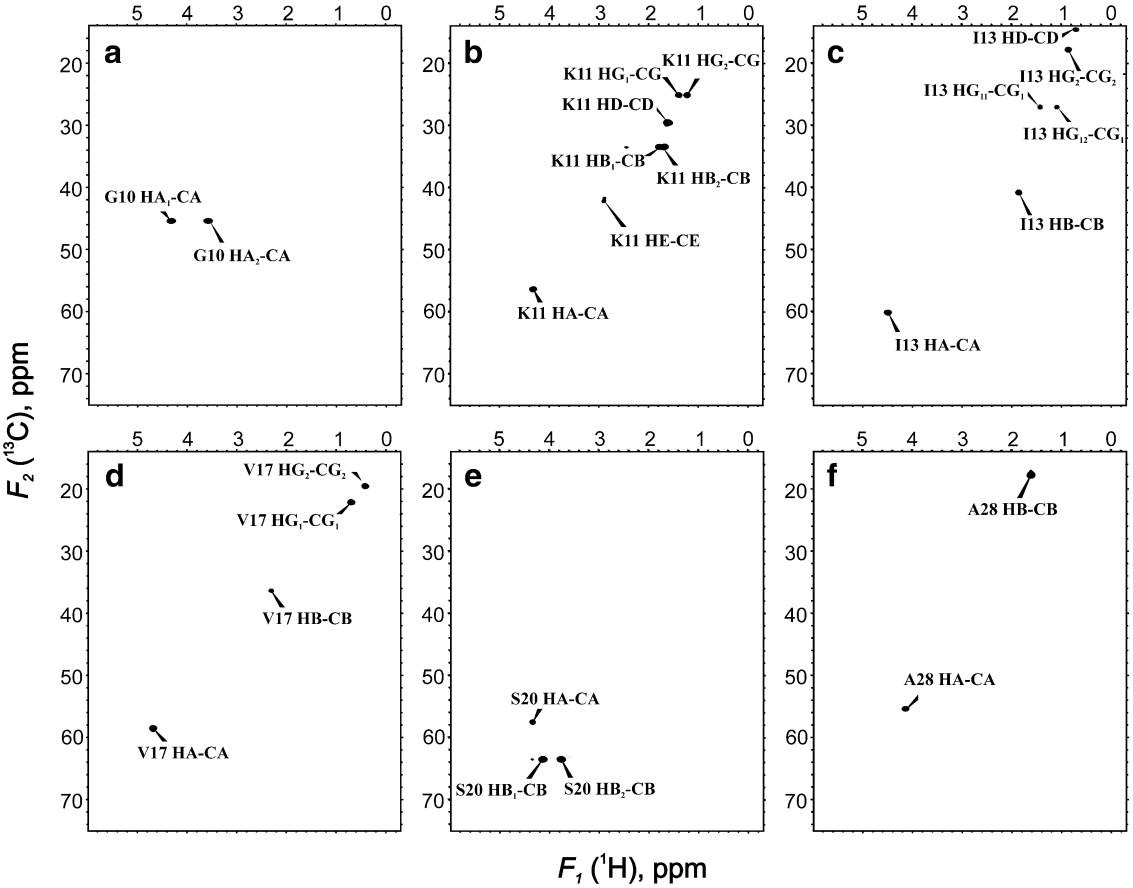

Fig. 8. Example of F_1–F_2 cross-sections from 5D HC(CC-TOCSY)CONH spectrum. Each cross-section was obtained by SMFT with HN_i, N_i, C'_{i-1}, coordinates determined by vectors consisting of HNCO peak positions. The 2D cross-sections reveal aliphatic $^1H_{i-1}$–$^{13}C_{i-1}$ correlations for each side-chain. (a)K11-NH (resonances of G10 side-chain); (b)T12-NH (resonances of K11 side-chain); (c) T14-NH (resonances of I13 side-chain); (d) Q18-NH (resonances of V17 side-chain); (e) D21-NH (resonances of S20 side-chain); (f) K29-NH (resonances of A28 side-chain). Experiment lasted 61 h, 2000 time domain points were used ($\Theta = 5.4 \times 10^{-5}$).

NMR System 700 spectrometer equipped with a Performa XYZ PFG unit and using the 5 mm ^{1}H, ^{13}C, ^{15}N – triple resonance probehead with high power ^{1}H, ^{13}C, and ^{15}N $\pi/2$ pulses of 5.9, 13.5, and 31.0 μs, respectively. The pulse sequences were adapted from the Varian Userlib BioPack package. The resulting spectra were saved and displayed in the format of SPARKY program [36].

For the 4D HNCOCA experiment (Fig. 6) spectral width of $6000 \times 2500 \times 2500 \times 12000$ Hz was set in $F_1(C\alpha)$, $F_2(C')$, $F_3(N)$, and $F_4(H_N)$, respectively. The maximum evolution times t_1, t_2 and t_3 of 7.0, 30.0, and 24.5 ms, were used, respectively. The ^{15}N evolution in t_3 was performed in constant time mode. Four scans were coherently added in all eight data sets for 2500 randomly distributed $t_1/t_2/t_3$ data points. Therefore, the sampling density reached 1.3% of required for conventional experiment ($\Theta = 1.3\ 10^{-2}$). The 3D HN(CO)CA, 3D HNCO(CA), 2D H(NCO)CA, and 2D H(N)CO(CA) experiments were acquired in exactly the same conditions with one or two time coordinates set to zeroes.

The 5D HC(CC-TOCSY)CONH experiment spectral width of $4500 \times 12000 \times 2770 \times 2300 \times 12000$ Hz was set in $F_1(H_{alif})$, $F_2(C_{alif})$, $F_3(C)$, $F_4(N)$,and $F_5(H_N)$, respectively. The maximum evolution times t_1, t_2, t_3, and t_4 of 15.0, 12.0, 20.0, and 30.0 ms, were employed, respectively. The ^{15}N evolution in t_3 was performed in semi constant time mode. Four scans were coherently added in all sixteen data sets for 2000 random $t_1/t_2/t_3/t_4$ data points with time decaying ($\sigma = 0.5$) Poisson disk distribution. The sampling density reached 0.0054% of required for conventional experiment ($\Theta = 5.4 \times 10^{-5}$). The 14.5 ms DIPSI-3 spinlock with $\gamma B_2/2\pi = 7500$ Hz, was used for isotropic mixing of ^{13}C magnetization.

In all cases acquisition time of 85 ms and relaxation delay of 1.2 s was employed. For processing of directly detected dimension cosine square weighting function was applied prior to Fourier transformation with zero-filling to 2048 complex points.

5. Conclusions

One of the most popular ways of evaluating sparse sampling is to compare the sampling density with the so called Nyquist density (i.e. number of points per time domain volume unit in conventional experiment). Such comparison gives no information about signal-to-artifact ratio, as the Nyquist density depends on the spectral width, maximum evolution times and number of dimensions, which have nothing in common with the artifact level. Due to this fact, it seems that use of random sampling becomes beneficial for high dimensionality and long maximum evolution times.

However, such an approach, resulting in high-resolution spectra of high dimensionality causes problems, as the spectrum requires hard disk space which is so far inconvenient (or even impossible) to be handled by modern computers. Moreover, besides problems with data handling, there is still a problem of human perception and imagination. Spectra of dimensionality higher than 3, although very useful, are quite hard to be analyzed. Nevertheless, some of their peak coordinates are well known from other, simpler measurements. Thus, the Fourier integral should be calculated in regions of interest only. For example, some dimensions may be calculated in a conventional way (*on-grid* spectral dimensions), while the rest only at the positions of the peaks, which are known. It is noteworthy, that spectral resolution may be very high, not causing significant problems with computing time or data size, as only small spectral regions are calculated. Thus, SMFT provides high precision of frequency determination.

Acknowledgments

This work was supported by Grant No.: N301 07131/2159, founded by Ministry of Science and Higher Education in years 2006–2009. The NMR measurements were accomplished at the Structural Research Laboratory, Chemistry Department, University of Warsaw, Poland.

References

[1] J. Jeener, Lecture presented at Ampere International Summer School II, Basko Polje, Yugoslavia, 1971.
[2] C. Szantay, NMR and the uncertainty principle: how to and how not to interpret homogeneous line broadening and pulse nonselectivity. III. Uncertainty?, Concept Magn Reson. A 32A (2008) 302–325.
[3] P.C. Lauterbur, Image formation by induced local interactions: examples employing nuclear magnetic resonance, Nature 242 (1973) 190–191.
[4] Ē. Kupče, R. Freeman, Projection-reconstruction of three-dimensional NMR Spectra, J. Am. Chem. Soc. 125 (2003) 13958–13959.
[5] B.E. Coggins, R.A. Venters, P. Zhou, Filtered backprojection for the reconstruction of a high-resolution (4,2)D CH_3–NH NOESY spectrum on a 29 kDa Protein, J. Am. Chem. Soc. 127 (2005) 11562–11563.
[6] K. Ding, A.M. Gronenborn, Novel 2D triple-resonance NMR experiments for sequential resonance assignments of proteins, J. Magn. Reson. 156 (2002) 262–268.
[7] S. Kim, T. Szyperski, GFT NMR, a new approach to rapidly obtain precise high dimensional NMR spectral information, J. Am. Chem. Soc. 125 (2003) 1385–1393.
[8] W. Koźmiński, I. Zhukov, Multiple quadrature detection in reduced dimensionality experiments, J. Biomol. NMR 26 (2003) 157–166.
[9] D. Malmodin, M. Billeter, Robust and versatile interpretation of spectra with coupled evolution periods using multi-way decomposition, Magn. Reson. Chem. 44 (2006) 185–195.
[10] B.E. Coggins, P. Zhou, Polar Fourier transforms of radially sampled NMR data, J. Magn. Reson. 182 (2006) 84–95.
[11] K. Kazimierczuk, W. Koźmiński, I. Zhukov, Two-dimensional Fourier transform of arbitrarily sampled NMR data sets, J. Magn. Reson. 179 (2006) 323–328.
[12] D. Marion, Processing of ND NMR spectra sampled in polar coordinates: a simple Fourier transform instead of reconstruction, J. Biomol. NMR 36 (2006) 45–54.
[13] B.E. Coggins, P. Zhou, Sampling of the NMR time domain along concentric rings, J. Magn. Reson. 184 (2007) 207–221.
[14] D. Rovnyak, D.P. Frueh, M. Sastry, Z.-Y.J. Sun, A.S. Stern, J.C. Hoch, G. Wagner, Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction, J. Magn. Reson. 170 (2004) 15–21.
[15] A.S. Stern, J.C. Hoch, Maximum entropy reconstruction in NMR, in: D.M. Grant, R. Harris (Eds.), Encyclopedia of Nuclear Magnetic Resonance, vol. 8, John Wiley & Sons, Chichester, UK, 1996.
[16] Ē. Kupče, R. Freeman, Hyperdimensional NMR spectroscopy, Prog. Nucl. Magn. Reson. 52 (2008) 22–30.
[17] V. Tugarinov, L.E. Kay, I. Ibraghimov, V.Y. Orekhov, High-resolution four-dimensional H-1–C-13 NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional decomposition, J. Am. Chem. Soc. 127 (2005) 2767–2775.
[18] L. Frydman, T. Scherf, A. Lupulescu, The acquisition of multidimensional NMR spectra within a single scan, Proc. Natl. Acad. Sci. USA 99 (2002) 15858–15862.
[19] Ē. Kupče, R. Freeman, Fast multi-dimensional NMR by minimal sampling, J. Magn. Reson. 191 (2008) 164–168.
[20] F.L. Zhang, R. Bruschweiler, Indirect covariance NMR spectroscopy, J. Am. Chem. Soc. 126 (2004) 13180–13181.
[21] Ē. Kupče, T. Nishida, R. Freeman, Hadamard NMR Spectroscopy, Prog. NMR Spectrosc. 42 (2003) 95–122.
[22] X. Meng, B.D. Nguyen, C. Ridge, A.J. Shaka, Enhanced spectral resolution by high-dimensional NMR using the filter diagonalization method and "hidden" dimensions, J. Magn. Reson. 196 (2009) 12–22.
[23] K. Kazimierczuk, A. Zawadzka, W. Koźmiński, I. Zhukov, Random sampling of evolution time space and Fourier transform processing, J. Biomol. NMR 36 (2006) 157–168.
[24] K. Kazimierczuk, A. Zawadzka, W. Koźmiński, I. Zhukov, Lineshapes and artifacts in Multidimensional Fourier Transform of arbitrary sampled NMR data sets, J. Magn. Reson. 188 (2007) 344–356.
[25] B.E. Coggins, P. Zhou, High resolution 4-D spectroscopy with sparse concentric shell sampling and FFT-CLEAN, J. Biomol. NMR 42 (2008) 225–239.
[26] Ivars Bilinskis, Digital Alias-Free Signal Processing, Wiley, 2007.
[27] R.N. Bracewell, The Fourier Transform and its Applications, McGraw-Hill Higher Education, 2000. pp. 224.
[28] I.N. Bronshtein, K.A. Semendyayev, G. Musiol, H. Muehlig, Handbook of Mathematics, fifth ed., Springer, 2007.
[29] A. Tarczyński, N. Allay, Spectral analysis of randomly sampled signals: suppression of aliasing and sampler jitter, IEEE Trans. Signal Process. 52 (2004) 3324–3334.
[30] M.A.Z. Dippé, E.H. Wold, Antialiasing through stochastic sampling, ACM SIGGRAPH Comput. Graphics 19 (1985) 69–78.
[31] K. Kazimierczuk, A. Zawadzka, W. Koźmiński, Optimization of random time domain sampling in multidimensional NMR, J. Magn. Reson. 192 (2008) 123–130.

[32] K. Kazimierczuk, A. Zawadzka, W. Kozminski, I. Zhukov, Determination of Spin–Spin couplings from ultrahigh resolution 3D NMR spectra obtained by optimized random sampling and multidimensional Fourier transformation, J. Am. Chem. Soc. 130 (2008) 5404–5405.
[33] S. Hiller, R. Joss, G. Wider, Automated NMR assignment of protein side-chain resonances using Automated Projections Spectroscopy (APSY), J. Am. Chem. Soc. 130 (2008) 12073–12079.
[34] B.E. Coggins, R.A. Venters, P. Zhou, Generalized reconstruction of n-D NMR spectra from multiple projections: application to the 5D HACACONH spectrum of protein G B1 domain, J. Am. Chem. Soc. 126 (2004) 1000–1001.
[35] B.E. Coggins, P. Zhou, PR-CALC: a program for the reconstruction of NMR spectra from projections, J. Biomol. NMR 34 (2006) 179–195.
[36] T.D. Goddard, D.G. Kneller, SPARKY 3, University of California, San Francisco.