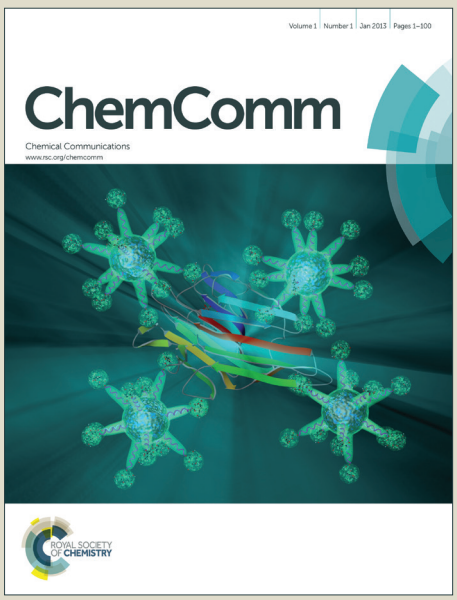
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

Journal Name

RSCPublishing

COMMUNICATION

Causality principle in reconstruction of sparse NMR spectra

Cite this: DOI: 10.1039/x0xx00000x

M. Mayzel[a], K. Kazimierczuk[b], and V. Yu. Orekhov[a],*

Received 00th January 2012,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

Sparse sampling offers dramatic increase in power and efficiency of magnetic resonance techniques in, chemistry, molecular structural biology, and other fields. Here we show that use of the causality property of an NMR signal is a general approach for major reduction of measuring time and quality improvement of the sparsely detected spectra.

The invention of multidimensional magnetic resonance (MR) experiments 40 years ago led to success of the modern MRI and NMR spectroscopy in medicine, chemistry, molecular structural biology, and other fields. The approach, however, has an important weakness: the detailed site-specific information and ultimate resolution obtained in two and higher dimensional experiments are contingent on the lengthy data collection required for systematic uniform sampling of the large multidimensional space spanned by the indirectly detected spectral dimensions[1]. A fundamental solution for this problem stems from an observation that upon appropriate transform, e.g. from the NMR time to frequency domain, MR signal becomes nearly-black or sparse, i.e. essentially zero in the vast majority of points and thus largely redundant. Darkness of the MR images and NMR spectra is a key for the remarkable success and rapid development of the non-uniform sampling (NUS) methods [1, 2, 3, 4-7]. The darker an object is, the less experimental measurements are needed for its recovery [8]. The transform that brings data into a dark presentation is called sparsifying transform. In NMR, the Fourier transform connects complex free induction decay (FID) signal in time domain and frequency spectrum. A properly phased spectrum consists of the real absorption part used for the analysis and the redundant imaginary dispersion part. Since an absorption signal is much narrower than the dispersion, the latter contribute the most to the total spectrum brightness. The main result of this paper is the notion that for most of the currently used algorithms, e.g. compressed sensing [4, 5], SIFT [9], maximum entropy [2, 7], MINT [6], etc. it is the dispersion part that sets the lower limit for the amount of measured data required for the high quality spectrum reconstruction from the NUS signal. We show that the causality property of the NMR signal can be used to construct a sparsifying transform, which eliminates the spectral dispersion part from the time domain signal and, thus, allows spectrum reconstruction with better fidelity and from fewer measurements. In NMR, the causality reflects the fact that the FID signal is only observed after excitation of the spin system, e.g. by a radiofrequency pulse, and is zero before the excitation.

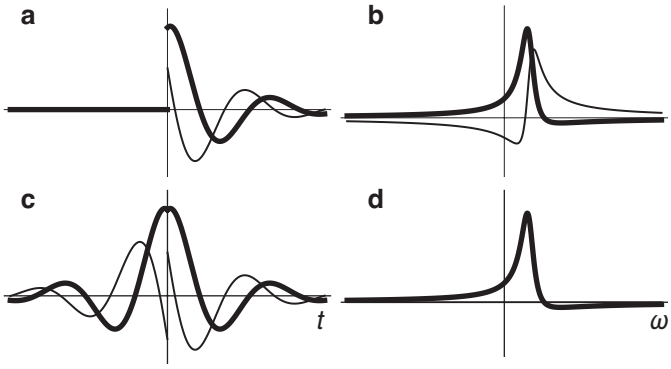


Figure 1. Illustration of the Kramers–Kronig relations. (a) FID and (c) virtual-echo representations of the NMR time domain signals with the corresponding spectra (b and d, respectively). Real and imaginary parts are shown in bold and thin lines, respectively. Note that the spectrum in panel (d) has zero imaginary part. Small zero order phase 0.15π is used to illustrate the effect of non-zero phase on the signal in the time and frequency domains.

It is well known that the Fourier transform of a causal time signal *S(t)* leads to spectrum, whose real and imaginary parts can be produced from each other using the Kramers–Kronig relations also known as the Hilbert transform [10]. The Kramers–Kronig relations are illustrated in Fig. 1. Signal $\boldsymbol{S_{FID}(t)}$ (Fig. 1a) and the corresponding spectrum in Fig. 1b are related via the Fourier transform. Spectrum in Fig. 1d is produced from the one in Fig. 1b by zeroing its imaginary part. The inverse Fourier transform of real spectrum in panel d gives a complex time domain signal (Fig. 1c), whose real and imaginary parts are essentially even and odd parts of the real and imaginary components of the FID (Fig 1a), respectively. Thus, signal in Fig. 1c is can also be produced by time reversal and complex conjugate of the FID.

$$S_{VE}(t) = \begin{cases} S_{FID}(t) & t \geq 0 \\ S^*_{FID}(-t) & t < 0 \end{cases} \quad (1)$$

In the following, we call $\boldsymbol{S_{VE}(t)}$ signal in Eq. 1 virtual-echo (VE). The original signal $\boldsymbol{S_{FID}(t)}$ can be obtained from $\boldsymbol{S_{VE}(t)}$ by zeroing the signal for negative time. Direct transition from panel d to panel b in Fig. 1 is done by the Hilbert transform. In practice, the Hilbert transform algorithm takes the detour d → c → a → b (Fig. 1) in order to use the computationally efficient fast Fourier transform.

The spectrum (Fig. 1d) obtained from the VE representation (Fig. 1c) consists of the traditionally looking real part and zero imaginary part. Depending on the signal phase, the real part can contain absorption, dispersion, or mixture of the both modes. Given *a priori*, the phase, Eq. 1 allows us to obtain time domain signal corresponding to the pure absorption spectrum and, thus, to construct a sparsifying transform that produces significantly darker spectrum than the traditional Fourier transform of the original FID.

Obtaining NMR spectrum from a time-domain signal is a typical example of the mathematical inverse problem. When all data points in the signal are present, the solution of the problem is trivial and is given by the Discrete Fourier Transform (DFT). In case of NUS, most of the data in the time-domain signal is missing and the unconstrained inverse problem has infinite number of solutions (i.e. spectra). A unique and "correct" spectrum is obtained by introducing additional assumptions such as minimal power, maximum entropy, maximal sparseness, etc. The VE presentation is equally applicable to traditional fully sampled and NUS signals. When the former is processed using DFT, FID and VE presentations lead to the equivalent spectra as illustrated in Fig. 1. However, when reconstructing spectra form NUS signal and in some other cases [11], use of the Kramers–Kronig relations, namely path a → c → d in Fig. 1 represents a significant advantage over the traditional processing, which is a → b → d.

Fig. 2 demonstrates benefits of the VE signal for two modern spectra recovering algorithms used for NUS signal: Spectroscopy by Integration of Frequency and Time Domain (SIFT)[9] and Compressed Sensing by Iterative Reweighted Least Squares (CS-IRLS)[4, 12]. Similar results for alternative CS algorithm, Iterative Soft Thresholding (CS-IST)[4, 13, 14] are presented in Supplementary Fig. S3. Both CS algorithms and SIFT can be applied without modifications to either traditional FID or VE signal. With SIFT making use of the prior knowledge about positions of dark regions in a spectrum and CS searching for the darkest among all possible spectra consistent with the measured data, both methods are expected to benefit from the darker representation of the spectrum provided by VE.

For a given number of NUS measurements, quality of the SIFT reconstruction improves, when the larger fraction of the spectrum area is free from signals and contains only the baseline noise. In our calculations, the signal-free area is defined by a mask, which excludes rectangles of a defined size around all peaks in the spectrum. This corresponds, for example, to a set-up in relaxation and kinetics studies [15], where the peak positions are known and only their intensities or integrals need to be defined. Figs. 2a and 2b show reconstructions of a 2D ^{1}H-^{15}N HSQC spectrum of human alpha-synuclein obtained using only 15% of the data from the full experiment.

By avoiding broad dispersion peaks, the VE signal ensures that larger fraction of the spectrum is "dark" and thus SIFT produces much better spectrum (Fig. 2b, Supplementary Fig. S4) and more accurate peak intensities in comparison to the reconstruction from the original FID (Fig. 2e and Supplementary Fig. S5). Fig. 2e (inset) illustrates that prior information about the signal phase does not have to be exact. For SIFT example, the peak intensities in the VE reconstruction obtained for uncorrected phase up to 15° are still better reproduced then those measured in the spectrum calculated for the traditional FID representation. Similar behaviour is also observed for the CS algorithms. For most of the multidimensional experiments, zero order phases for the indirect spectral dimensions are known and thus can be corrected in the time domain to values close to zero prior to the spectrum reconstruction.

Similarly to SIFT, CS also assumes, that the major part of a spectrum is dark. However, no assumption is made about the exact location of the dark regions, which creates an apparently unsolvable combinatorial problem. Yet, it has been recently reformulated as a relatively simple task of spectral l_p-norm ($0<p\leq1$) minimization[16]:

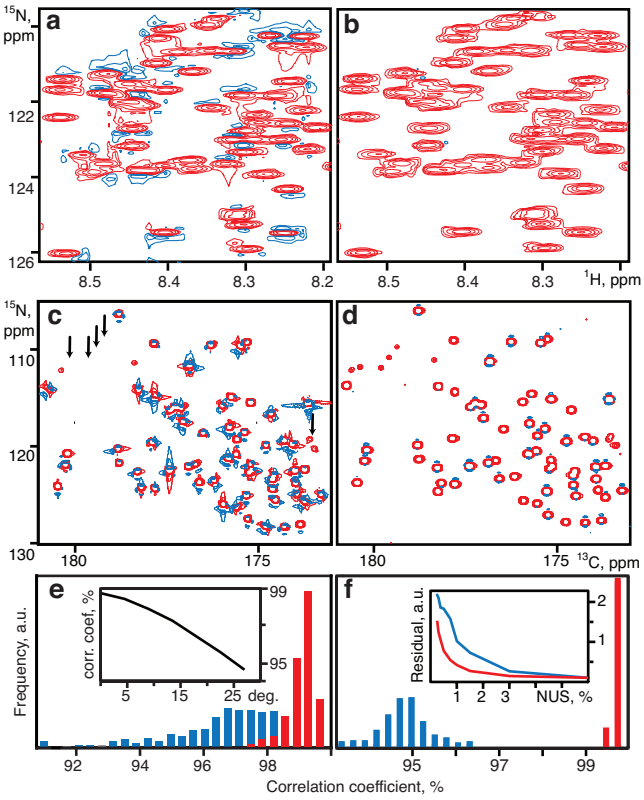


Figure 2. Comparison of SIFT (a,b,e) and CS (c,d,f) spectral reconstructions obtained using time-domain signal in traditional FID (a,c) and VE (b,d) presentations. (a,b) 2D 1H-15N HSQC of alpha-synuclein (15% NUS). (c,d) 13C-15N projection from a 3D HNCO spectrum of ubiquitin (0.7% NUS). In the pairs of spectra, the contours are shown at the same level. Arrows in panel (c) indicate several true weak signals present in the VE reconstruction (d) but missing in panel (c). Histograms (e,f) show distribution of the correlation coefficients between signal intensities measured in the reference spectrum and the spectra reconstructed with VE (red) and FID (blue) (e) SIFT: 500 resampling trials with 15% NUS. Inset in panel (e) the median (over 25 resampling trials) of correlation coefficients for the VE processing versus uncorrected zero order phase. (f) CS: 200 resampling trials with 0.7% NUS. Inset in panel (f) Residual of the CS-IRLS reconstructions versus the sampling level obtained using FID (blue) and VE signal representations (red line). The residuals are defined as an RMSD of the difference between the reference spectrum and the corresponding CS-IRLS reconstruction measured over the signal regions (±50 Hz in all spectral dimensions around every peak in a complete manually verified peak list). As the reference we use 6% NUS HNCO averaged over the reconstructions obtained with and without VE.

$$min_{\boldsymbol{F}} \, |\boldsymbol{A}\cdot\boldsymbol{F}-\boldsymbol{S}|_{l2}^{2}+|\boldsymbol{F}|_{lp}^{p} \qquad (2)$$

where $\boldsymbol{F}$ and $\boldsymbol{S}$ are the frequency spectrum and time domain signal, respectively; $\boldsymbol{A}$ is matrix derived from the inverse Fourier transform matrix; and l_p-norm is defined as:

$$|\boldsymbol{F}|_{\boldsymbol{lp}}=(|F_1|^p+|F_2|^p+\cdots+|F_N|^p)^{1/p} \qquad (3)$$

In the present paper the $p=1$ is used for IST algorithm [13] and lp-norm with p iteratively approaching 0 for IRLS algorithm [4, 17]. The applicability of the CS method in NMR spectroscopy has been commented recently by many authors [4, 5, 18, 19], with important conclusions on the limited applicability to non-random sampling [20] and superior performance of non-convex lp-norms ($p<1$) [19, 21].

Here we apply CS IRLS algorithm [4] to reconstruct a 3D HNCO spectrum sampled at the level of 0.7 %, without VE (Fig. 2c) and with VE in both indirect dimensions (Fig. 2d). It can be seen, that VE improves the reconstruction significantly by providing better line shapes, more accurate peak intensities (Fig. 2f), and revealing low intensity signals. Supplementary Fig. S3 shows a notable improvement for the 2D ^{1}H-^{15}N HSQC spectrum of intrinsically disordered protein alpha-synuclein processed with CS-IST.

The effect can be explained by the basic CS theorem, binding the number of properly reconstructed spectral points, which is essentially a measure of spectrum darkness, with the sampling level [16]. With the VE, fewer points contribute to each peak in the spectrum and thus relatively low sampling level is sufficient to fulfil the condition for the successful CS reconstruction. It should be emphasized, that the striking advantage of the VE demonstrated in Fig. 2 and Supplementary Fig. S3-5, is mostly due to the very low sampling level. Without the VE, high quality reconstructions by CS and SIFT are also possible, but require at least twice as many sampling points for the presented spectra (inset in Fig. 2f and Supplementary Fig. S4).

As it was pointed by Donoho, at al. [8] there is an unambiguous relation between the darkness of NMR spectrum and quality of the spectral reconstruction by the Maximum Entropy or minimum l_1-norm minimisation. It is therefore likely that most of related methods including FM-reconstruction [22], MINT [6], hmsIST [14], QME [7] etc. will also benefit from the VE signal.

Conclusions

We show that the causality property of the NMR signal can be exploited to dramatically enhance performance of the CS, SIFT and probably many other algorithms commonly used for the reconstruction of NUS spectra. Our findings open a way to significant reduction in measurement time and improvement of the quality of NUS spectra and thus to increase of power and appeal of multidimensional NMR spectroscopy in multitude of its existing and future applications. The method is particularly useful for short living systems, time resolved measurements, and high-dimensional experiments on intrinsically disordered proteins.

The work was supported by the Swedish Research Council (research grant 2011-5994); Swedish National Infrastructure for Computing (grant SNIC 001/12-271); Polish National Centre of Science (grant DEC-2012/07/E/ST4/ 01386); Polish Ministry of Science and Higher Education (grant IP2011 023171); Foundation for Polish Science, TEAM programme. We thank Dina Katabi and Haitham Hassanieh (Dept Electr Eng & Comput. Sci., Massachusetts Institute of Technology) for an inspiring discussion and Anna Zawadzka-Kazimierczuk (Biological and Chemical Research Centre, University of Warsaw established from EU Regional Development Fund) for HSQC spectrum of alpha-synuclein.

Notes and references

[a] Swedish NMR Centre, University of Gothenburg, Box 465, S-405 30 Göteborg, Sweden;

[b] Centre of New Technologies, University of Warsaw, Żwirki i Wigury 93, 02-089 Warsaw, Poland

* Correspondence author: E-mail: vladislav.orekhov@nmr.gu.se

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/c000000x/

1. M. Billeter and V. Y. Orekhov, in *Novel Sampling Approaches in Higher Dimensional NMR*, eds. M. Billeter and V. Y. Orekhov, Springer, Heidelberg Dordrecht London New York, 2012, vol. 316, pp. ix-xiv.
2. J. C. J. Barna, E. D. Laue, M. R. Mayger, J. Skilling and S. J. P. Worrall, *Biochem Soc T*, 1986, **14**, 1262-1263 ; S. G. Hyberts, K. Takeuchi and G. Wagner, *J Am Chem Soc*, 2010, **132**, 2145-2147
3. V. Y. Orekhov, I. Ibraghimov and M. Billeter, *J Biomol NMR*, 2003, **27**, 165-173 ; B. E. Coggins, R. A. Venters and P. Zhou, *J Am Chem Soc*, 2004, **126**, 1000-1001 ; R. Bruschweiler and F. L. Zhang, *J Chem Phys*, 2004, **120**, 5253-5260 ; V. Tugarinov, L. E. Kay, I. Ibraghimov and V. Y. Orekhov, *J Am Chem Soc*, 2005, **127**, 2767-2775 ; D. Marion, *J Biomol NMR*, 2006, **36**, 45-54 ; V. Jaravine, I. Ibraghimov and V. Y. Orekhov, *Nat Methods*, 2006, **3**, 605-607 ; M. Lustig, D. Donoho and J. M. Pauly, *Magn Reson Med*, 2007, **58**, 1182-1195 ; S. Hiller, R. G. Garces, T. J. Malia, V. Y. Orekhov, M. Colombini and G. Wagner, *Science*, 2008, **321**, 1206-1210 ; D. Sakakibara, A. Sasaki, T. Ikeya, J. Hamatsu, T. Hanashima, M. Mishima, M. Yoshimasu, N. Hayashi, T. Mikawa, M. Wälchli, B. O. Smith, M. Shirakawa, P. Güntert and Y. Ito, *Nature*, 2009, **457**, 102-105
4. K. Kazimierczuk and V. Y. Orekhov, *Angew Chem-Int Edit*, 2011, **50**, 5556-5559
5. D. J. Holland, M. J. Bostock, L. F. Gladden and D. Nietlispach, *Angew Chem Int Ed*, 2011, **50**, 6548-6551
6. S. Paramasivam, C. L. Suiter, G. Hou, S. Sun, M. Palmer, J. C. Hoch, D. Rovnyak and T. Polenova, *J Phys Chem B*, 2012, **116**, 7416-7427
7. J. Hamatsu, D. O'Donovan, T. Tanaka, T. Shirai, Y. Hourai, T. Mikawa, T. Ikeya, M. Mishima, W. Boucher, B. O. Smith, E. D. Laue, M. Shirakawa and Y. Ito, *J Am Chem Soc*, 2013, **135**, 1688-1691
8. D. L. Donoho, I. M. Johnstone, J. C. Hoch and A. S. Stern, *J Royal Statist Soc B*, 1992, **54**, 41-81
9. Y. Matsuki, M. T. Eddy and J. Herzfeld, *J Am Chem Soc*, 2009, **131**, 4648-4656
10. S. H. Hall and H. L. Heck, Wiley : IEEE,, Hoboken, N.J., 2009; E. Bartholdi and R. R. Ernst, *J Magn Reson*, 1973, **11**, 9-19
11. A. Gibbs and G. A. Morris, *J Magn Reson*, 1991, **91**, 77-83
12. E. J. Candes, M. B. Wakin and S. P. Boyd, *J Fourier Anal Appl*, 2008, **14**, 877
13. A. Papoulis, *IEEE T Circuits Syst*, 1975, **22**, 735-742
14. S. G. Hyberts, A. G. Milbradt, A. B. Wagner, H. Arthanari and G. Wagner, *J Biomol NMR*, 2012, **52**, 315-327
15. D. M. Korzhnev, I. V. Ibraghimov, M. Billeter and V. Y. Orekhov, *J Biomol NMR*, 2001, **21**, 263-268 ; Y. Matsuki, T. Konuma, T. Fujiwara and K. Sugase, *J Phys Chem B*, 2011, **115**, 13740-13745 ; P. Selenko, D. P. Frueh, S. J. Elsaesser, W. Haas, S. P. Gygi and G. Wagner, *Nat Struct Mol Biol*, 2008, **15**, 321-329
16. E. J. Candes and M. B. Wakin, *IEEE Sig Proc Mag*, 2008, **25**, 21-30
17. A. E. Yagle, http://webeecsumichedu/~aey/sparsehtml, 2008
18. I. Drori, *Eurasip J Advances in Signal Processing*, 2007 20248.

19. X. B. Qu, D. Guo, X. Cao, S. H. Cai and Z. Chen, *Sensors*, 2011, **11**, 8888-8909
20. A. S. Stern, D. L. Donoho and J. C. Hoch, *J Magn Reson*, 2007, **188**, 295-300
21. K. Kazimierczuk and V. Y. Orekhov, *J Magn Reson*, 2012, **223**, 1-10
22. S. G. Hyberts, D. P. Frueh, H. Arthanari and G. Wagner, *J Biomol NMR*, 2009, **45**, 283-294

