Nonuniform Sampling and Maximum Entropy Reconstruction in Multidimensional NMR

Jeffrey C. Hoch[1*], Mark W. Maciejewski[1], Mehdi Mobli[2], Adam D. Schuyler[1], and Alan S. Stern[3]

[1]University of Connecticut Health Center, Farmington, CT USA

[2]Centre for Advanced Imaging, University of Queensland, St. Lucia, QLD AUS

[3]Rowland Institute at Harvard, Cambridge, MA USA

KEYWORDS nonuniform sampling, sparse sampling, compressed sensing, multidimensional NMR, spectrum analysis, non-Fourier methods, maximum entropy.

Maximum Entropy Reconstruction and Nonuniform Sampling in Multidimensional NMR

JEFFREY C. HOCH[1], MARK W. MACIEJEWSKI[1], MEHDI MOBLI[2], ADAM D. SCHUYLER[1], AND ALAN S. STERN[3]

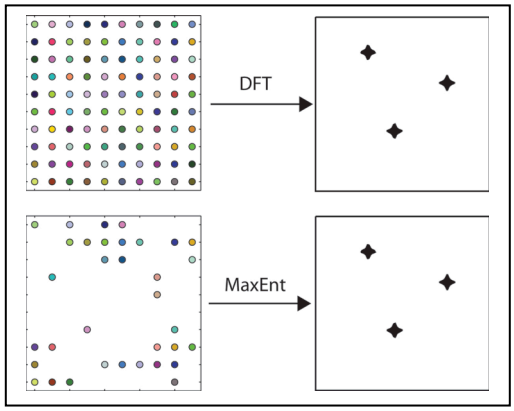
1University of Connecticut Health Center, Farmington, CT USA

2Center for Advanced Imaging, University of Queensland, St. Lucia, QLD AUS

3Rowland Institute at Harvard, Cambridge, MA USA

(DUE DATE NOVEMBER, 2013).

CONSPECTUS



NMR spectroscopy is one of the most powerful and versatile analytic tools available to chemists. The discrete Fourier transform (DFT) played a seminal role in the development of modern NMR, including the multidimensional methods that are essential for complex biomolecules, but it suffers from well-known limitations. Chief among these is the difficulty of obtaining high-resolution spectral estimates from short data records. For multidimensional NMR experiments, this imposes a sampling burden, because the time required to perform an experiment is proportional to the number of data samples. At high magnetic field, where spectral dispersion is greatest, the problem becomes particularly acute. Consequently multidimensional NMR experiments that rely on the DFT either must sacrifice resolution in order to be completed in reasonable time, or they must use inordinate amounts of time to achieve the potential resolution afforded by high-field magnets. Maximum entropy (MaxEnt) reconstruction is a non-Fourier method of spectrum analysis capable of providing high-resolution spectral estimates from short data records. It can also be used with nonuniformly sampled data sets. Since resolution is substantially determined by the largest evolution time sampled, nonuniform sampling enables high resolution while avoiding the need to uniformly sample at large numbers of evolution times. The Nyquist sampling theorem does not apply to nonuniformly sampled data, and artifacts that attend the use of nonuniform sampling can be viewed as frequency-aliased signals. Strategies for suppressing nonuniform sampling artifacts include careful design of the sampling scheme and special methods for computing the spectrum.

Time savings of a factor of three for each of the N-1 indirect dimensions of an N-dimensional NMR experiment are now routinely reported, making practical high-resolution 3- and 4-dimensional experiments that were previously prohibitively time consuming. Conversely, tailored sampling in the indirect dimensions has been utilized to improve sensitivity. Improvements in nonuniform sampling strategies appear poised to enable further reductions in sampling requirements for high resolution NMR spectra, and the combination of these strategies with robust non-Fourier methods of spectrum analysis (such as MaxEnt) represent a profound change in the way multidimensional experiments are conducted. The potential benefits will enable more advanced applications of multidimensional NMR spectroscopy to biological macromolecules, metabolomics, natural products, dynamic systems, and other areas where resolution, sensitivity, or experiment time are limiting. Just as the development of multidimensional NMR methods presaged multidimensional methods in other areas of spectroscopy, we anticipate that nonuniform sampling approaches will find application in other forms of spectroscopy.

Introduction

NMR spectroscopy can probe all states of matter and quantify the composition of mixtures, structures of molecules, dynamics of rate processes, and thermodynamics of association. This versatility comes at a price; useful sensitivity and high resolution requires expensive magnets and lengthy experiments. The introduction of Fourier Transform (FT) NMR enabled dramatic improvements in sensitivity and resolution[1]. In FT-NMR, the response of spins to a strong RF pulse is recorded, and the discrete FT (DFT) is used to compute the spectrum. In 2D NMR, for example, a delay between two RF pulses, representing an "indirect" time dimension, is parametrically sampled by repeating the experiment using different values for the time delay. Successive Fourier transformation along the rows and the columns of the resulting data matrix yields a two-dimensional spectrum. FT-NMR readily generalizes to arbitrary numbers of dimensions[2], enabling the resolution of individual nuclear resonances in complex systems.

The time required for a multidimensional NMR experiment is directly proportional to the number of samples in the indirect dimensions. Together, the requirements of uniform sampling (required by the DFT) with sufficiently small increments of the delay time to span the width of the spectrum (the Nyquist condition[3]) and long evolution times (for high resolution) mean that high-resolution spectra require lengthy experiments. Conversely, shorter experiments result in lower resolution spectra. Conventional uniform sampling in a high-resolution 3D experiment can require over a week of measuring time. While 3D experiments have become routine, resolution along the indirect dimensions is usually substantially less than the acquisition dimension. 4D experiments are far from routine, because of the time required to collect data sufficient for even moderate resolution.

The subject of this Account is the use of nonuniform sampling (NUS) methods in multidimensional NMR. NUS permits high-resolution spectra to be obtained from short data records, drastically reducing experiment times. NUS can also be tailored to increase sensitivity. We focus on maximum entropy (MaxEnt) reconstruction, one of a number of non-Fourier methods of spectrum analysis suitable for NUS data, because it is particularly versatile and robust. Fast NMR methods are a burgeoning area of development[4], and NUS represents just one approach, but one of the most general.

The MaxEnt method. MaxEnt reconstruction finds the spectrum that maximizes the entropy while maintaining consistency with the measured data. The use of entropy as a measure of missing information originated with Shannon and is the foundation for information theory[5]. Consistency of the computed spectrum **f** with the measured data **d** is defined by the condition

$$C(\mathbf{f},\mathbf{d}) = C_0 \qquad (1)$$

where C(**f**, **d**) is the unweighted χ-squared statistic,

$$C(\mathbf{f},\mathbf{d}) = \sum_{i=0}^{M-1} \left| m_i - d_i \right|^2 = \sum_{i=0}^{M-1} \left| \mathrm{iDFT}(\mathbf{f})_i - d_i \right|^2 \qquad (2)$$

and C_0 is an estimate of the noise level; iDFT is the inverse DFT, and **m** is a "mock data" vector given by iDFT(**f**). The constrained optimization problem is converted to an unconstrained optimization through introduction of a new objective function

$$Q(\mathbf{f},\mathbf{d}) = S(\mathbf{f}) - \lambda C(\mathbf{f},\mathbf{d}), \qquad (3)$$

where $S(\mathbf{f})$ is the entropy. The unconstrained problem is to find the **f** that minimizes $Q(\mathbf{f}, \mathbf{d})$, where the value of the Lagrange multiplier λ is adjusted to obtain $C = C_0$. $C(\mathbf{f},\mathbf{d})$ and $S(\mathbf{f})$, and thus $Q(\mathbf{f},\mathbf{d})$, readily generalize to multiple dimensions. The seminal development of the "Cambridge" algorithm[6], which is both robust and highly efficient, launched the modern application of the maximum entropy principle in NMR. Extensions to the Cambridge algorithm have provided additional performance gains and adapted it to the requirements of phase-sensitive NMR data[3].

A schematic diagram for MaxEnt reconstruction is shown in Figure 1. The algorithm begins with a trial spectrum equal to zero everywhere. At each iteration, **m** is computed from the current value of **f**. The algorithm constructs a small set of direction vectors, and computes a quadratic approximation to the entropy in the subspace spanned by these vectors. Since $C(\mathbf{f},\mathbf{d})$ is itself quadratic, it is possible to analytically maximize $S(\mathbf{f})$ subject to the constraint in this subspace. This results in **f** for the next iteration. Because it makes no assumptions about the nature of the signals, MaxEnt reconstruction can be applied to data with arbitrary lineshapes. The computation of $C(\mathbf{f},\mathbf{d})$ can be limited to arbitrary subsets of **m**; this is the basis for the application of MaxEnt with nonuniform sampling methods.

In principle multidimensional MaxEnt spectra can be reconstructed by computing the overall entropy of the full spectrum, or by computing a series of sub-spectra. For example, a 2D MaxEnt spectrum can be computed via a series of 1D MaxEnt reconstructions in the indirect dimension following Fourier processing of the acquisition dimension. If the constraint, C_0, is kept constant between sub-spectra, variation in the weighting (λ) of the constraint and the entropy can result. This will introduce small changes in the reconstruction between subspectra and may have a significant effect on peak shapes. By using a constant value for λ, one can

eliminate the variation of the nonlinearity between sub-spectra and obtain the same result as reconstructing the entire spectrum at once. A good estimate of λ can be made by finding representative sub-spectra where the constraint $C(\mathbf{f}, \mathbf{d})=C_0$ is satisfied and using the value of λ found for these sub-spectra to perform the complete reconstruction. The same strategy can be applied to higher dimensions, i.e. a 3D spectrum can be constructed as a series of 2D plane reconstructions. The approach of using a fixed value of λ, rather than a fixed value of C_0, is called the constant-λ algorithm. [7]

While the formal derivation of the MaxEnt algorithm specifies criteria for determining the value of C_0 and another parameter that appears in the complex entropy functional, applying those criteria in practice is challenging. Fortunately the results of MaxEnt reconstruction are not terribly dependent on the precise values of the parameters, over a wide range. A heuristic algorithm has been shown to automatically find useful values for the adjustable parameters.[8]

While numerical solution is required in the general case, there is a special case where MaxEnt reconstruction has an analytical solution that gives insights into how MaxEnt reconstruction works. When *N* (the number of points in the reconstructed spectrum) is equal to *M* (the number of experimental data points), Parseval's theorem[3] permits the constraint statistic to be computed in the frequency domain. The MaxEnt solution[9] corresponds to a nonlinear transformation, applied point-by-point to the DFT of the time domain data. Figure 2 illustrates the transformation $\delta_\lambda^{-1}(x)$ for various values of λ (panel A). The transformation depends on the value of λ, and has the effect of scaling every point in the spectrum down, but points closer to the baseline are scaled down more than points far above the baseline (panel B). This explains why noise near the baseline is suppressed more effectively than noise superimposed on top of broad

features. This result implies an important distinction between signal-to-noise-ratio (SNR) and sensitivity. Applying the same transformation to both the signal and the noise cannot improve sensitivity, since peaks that are comparable in height to the noise level will be reduced by the same amount as the noise. The SNR may increase, but small peaks will be just as difficult to distinguish as before. In this special case, gains in SNR in the MaxEnt reconstruction are purely cosmetic. In the more general case, there may be real sensitivity gains[10,11]. However, a prudent investigator will always question whether gains in SNR really correspond to gains in sensitivity.[9] Note that as λ increases, the relative weight given to the constraint term in the objective function increases, and the transformation becomes more nearly linear. This property has been used to perform MaxEnt reconstructions that are nearly linear.[12]

The nonlinearity of MaxEnt has important implications when quantification of peak intensities or volumes is required, such as nuclear Overhauser effect measurements. One approach is to tightly constrain the reconstruction to match the data, which forces the reconstruction to be nearly linear (although at the expense of noise suppression)[12,13]. Another is to inject synthetic signals into the time domain data prior to reconstruction. A calibration curve can then be constructed by comparing measured intensities or volumes to the known amplitudes of the injected signals[14].

MaxEnt is just one of a host of methods that have been developed as alternatives to the DFT for reconstructing spectra from NUS data. Some methods place restrictions on the way the data are sampled, for example along radial vectors in time. Others support arbitrary sampling schemes. Strengths and weaknesses of the various methods have been compared recently[4].

Sampling fundamentals. The Nyquist sampling theorem states that to unambiguously determine frequencies, the sampling interval Δt must be at least as short as the reciprocal of the spectral

width *SW* spanned by frequency components in the signal. Frequencies higher than 1/Δt are aliased, or mirrored about the spectral limits. The interval between frequency elements (the digital resolution) of the DFT is $1/N\Delta t$, where N is the number of samples collected; $N\Delta t$ is the maximum evolution time. The number of samples required to maintain a given maximum evolution time increases with magnetic field strength, because increasing field increases *SW*.

NUS schemes that sample a subset of the evolution times normally sampled using uniform sampling are called on-grid. In schemes such as radial, spiral or concentric ring, the samples do not fall on this Cartesian grid[15]. Exponentially-biased random (on-grid) sampling was the first general NUS approach applied to multidimensional NMR[16]. By analogy with matched filter apodization[17], biasing the sampling scheme toward shorter evolution times, using an exponential weighting to match the decay rate of the signal envelope, improves sensitivity. We refer to this as envelope-matched sampling (EMS). Generalizations to sine-modulated signals, where the signal is small at the beginning, and constant-time experiments, where the signal envelope does not decay, utilize the same rationale[18,19]. Distributions other than random have been employed; Poisson gap sampling[20] avoids long gaps between samples while ensuring the samples are approximately randomly distributed.

It bears emphasizing that Nyquist condition does not apply to NUS: NUS invariably introduces sampling artifacts that are a form of aliasing[21]. To a good approximation, the positions and amplitudes of the sampling artifacts relative to true signals can be derived *a priori* from the sampling scheme. The point-spread function (PSF) is the spectrum of a real-valued sampling function **K** consisting of the value 1 for samples included in the NUS scheme and the value zero for samples not included in the scheme. For on-grid sampling, the PSF can be computed using the DFT. **K** has the property that when it multiplies a uniformly sampled data vector, element-

wise, it results in a data vector in which the values not sampled in the NUS scheme have the value zero. The DFT of this zero-augmented NUS data (referred to as nonuniform DFT, nuDFT[22]) is equal to the convolution of the DFT spectrum of the uniformly sampled data with the PSF. Thus estimating the spectrum of an NUS data set is equivalent to deconvolving the PSF from the DFT spectrum of the zero-augmented data. While nuDFT provides useful insights into the nature of NUS artifacts, it is not a DFT of NUS data, nor is it a very good estimator of the spectrum of NUS data.

The PSF typically consists of a main central component at zero frequency surrounded by smaller non-zero frequency components. Because the PSF enters into the DFT of the zero-augmented data through convolution, each non-zero frequency component of the PSF gives rise to a sampling artifact for each signal component, with positions relative to the signal components that are the same as the relationship of the satellite peaks to the central component in the PSF. The amplitudes of the sampling artifacts are proportional to the amplitude of the signal component and the relative height of the satellite peaks in the PSF. Thus the largest sampling artifacts arise from the largest-amplitude components of the signal spectrum. The useful dynamic range (ratio between the magnitude of the largest and smallest detectable signal components) of the DFT spectrum of the zero-augmented data can be directly estimated from the PSF as the ratio between the amplitudes of the zero-frequency component and largest non-zero frequency component; this ratio is the peak-to-sidelobe ratio (PSR). The ability of a method of spectrum analysis to suppress sampling artifacts is ultimately limited by both the noise and the dynamic range of the signal.

In addition to the PSR, another useful metric for sampling schemes is the sensitivity relative to uniform sampling. The relative sensitivity depends on the sampling scheme and the

nature of the signals, principally the decay rate (or R_2^*) of the signal envelope; in contrast the PSR depends only on the sampling scheme. The relative sensitivity *r(***K***)* of a sampling scheme with sampling function **K** for a hypothetical signal can be estimated from the signal amplitude captured by a NUS scheme divided by that for uniform sampling having the same maximum evolution times t_{max}. For an exponentially decaying signal, the relative sensitivity of a scheme **K** spanning a two-dimensional grid with size n_1 by n_2 is approximately given by[23]

$$r(\mathbf{K}) = \frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} K_{ij}p_{ij}}{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} p_{ij}} \qquad (4)$$

where the elements of **p** are given by

$$p_{ij} = exp\left\{-\left(\frac{(i-1)R_2(1)}{SW(1)}\right) - \left(\frac{(j-1)R_2(2)}{SW(2)}\right)\right\} \quad , \qquad (5)$$

$R_2(1)$ and $R_2(2)$ are the signal envelope decay rates, and $SW(1)$ and $SW(2)$ are the spectral widths in the two dimensions. A more accurate estimate would include the amount of noise captured by the NUS scheme, compared to uniform sampling. Recently systematic efforts to improve sensitivity using NUS have been reported[12,24].

The magnitudes of artifacts in NUS spectra depend on the distribution of sampled evolution times and the sampling coverage $\gamma(\mathbf{K})=k/N$, with k equal to the number of nonzero entries in **K** and N the total number of elements in **K**, which is the fraction of the evolution times from a uniform grid that are sampled by **K**. For the example above, $\gamma(\mathbf{K}) = \frac{k}{(n_1 \times n_2)}$. In general, the PSR increases with increasing γ, with only the zero-frequency element of the PSF having a non-zero value for γ=1. Because large values of the nonzero frequency components result from correlations among the sampled evolution times, **K** composed of random evolution times will

have the smallest sampling artifacts, and highest PSR, for a given coverage. For decaying sinusoids, a random sampling scheme will not have the highest sensitivity. A compromise between sensitivity and small artifacts leads to biased random sampling distributions, such as EMS[17]. PSFs, together with PSRs and relative sensitivity are shown in Fig. 3 for some representative sampling schemes, for sampling coverages of 0.3, 0.1, and 0.05. The importance of randomness in sampling schemes for suppressing sampling artifacts has been explored in depth[21,25].

The resolution of any sampling scheme along a given dimension, whether uniform or nonuniform, is largely determined by t_{max}. Using the DFT, resolving spectral features separated by the natural linewidth requires sampling at evolution times of πT_2 or longer, but sampling beyond $1.26T_2$ results in diminishing returns on sensitivity [26]. With MaxEnt, sampling to $1.26T_2$ usually resolves spectral components separated by the natural linewidth, and thus represents a reasonable compromise between sensitivity and resolution for decaying signals. For experiments in which the evolution period is constant-time, the signal decay is determined mainly by field inhomogeneity (RF and B_0), and so practical limits on t_{max} are imposed by the inhomogeneity or length of the constant time period, rather than T_2.

The degree to which reducing sampling coverage via NUS can reduce experiment time, compared to uniform sampling, depends on a number of factors in addition to the randomness of the sampling scheme. Dynamic range of the signals and their amplitude relative to noise are key determinants. Because sampling artifacts enter through convolution, high dynamic range signals present challenges. Instead of being additive, the amplitudes of the largest sampling artifacts are determined by the amplitude of the strongest signal component. When the dynamic range is

large, these artifacts may exceed the amplitude of weak signal components. Thus more aggressive reductions in sampling coverage are feasible for high sensitivity experiments that have low dynamic range, and are more challenging for experiments with low sensitivity or high dynamic range. Dimensionality and sparsity (the fraction of values with amplitudes close to zero) of the spectrum have also been shown to play a role[27,28]. Increasing dimensionality helps in two ways, by decreasing the coherence of sampling and by increasing the sparsity of the spectrum. Sparsity helps because non-Fourier methods of spectrum analysis such as MaxEnt and l_1-norm reconstruction work best for recovering sparse spectra[28]. As we show below, sampling coverage can conservatively be around 1/3 for each NUS dimension (e.g. roughly an order of magnitude reduction relative to uniform sampling for two indirect dimensions), even for challenging signals with high dynamic range, while more aggressive reductions have been used successfully for low dynamic range signals.

Optimal sampling. As noted above in the discussion of sensitivity, optimizing sampling schemes can be challenging. Additional optimization can be realized by adjusting the sampling grid. Nonuniform sampling on an oversampled grid has been shown to shift artifacts to the edges of the spectrum, outside the desired spectral window, although the magnitude of sampling artifacts is not affected[29]. Because a sampling scheme that is optimal for one signal will not necessarily be optimal for a signal containing different frequency components, the design of efficient sampling schemes involves tradeoffs. Simply put, no single NUS scheme will be best suited for all experiments. Despite these challenges, prior knowledge about the signal can successfully inform the design of efficient sampling schemes. One approach is to use "greedy" or adaptive sampling, in which a sampling scheme is iteratively generated by asking what sample (corresponding to a specific combination of indirect evolution times), added to samples already measured, will most

improve some metric of performance. Suitable metrics can be derived from the PSF (to minimize sampling artifacts), the relative sensitivity, the ability to resolve expected resonances based on statistical knowledge of chemical shift distributions[30], or from characteristics of the spectrum reconstruction prior to the next sample[31,32]. A caveat is that while prior knowledge can greatly improve sampling efficiency when it is accurate, highly-tailored sampling schemes can be less robust than more general sampling schemes when there are deviations from the underlying assumptions[23] or high levels of experiment noise.

NUS in action. One compelling reason for adopting NUS in multidimensional NMR experiments is dramatic savings in data collection time, without loss of resolution. The potential savings increase with magnetic field strength and with dimensionality. The time required for a multidimensional experiment is directly proportional to the number of evolution times sampled in the indirect dimensions,

$$t_{exp} = (t_{acq} + t_{rc}) \times n_t \times k \times 2^{d-1} \qquad (6)$$

where t_{acq} is the time required to sample one FID, t_{rc} is the recycle time between transients, n_t is the number of FIDs co-averaged, k is the number of samples in the indirect dimensions (for uniform sampling $k = \prod_{j=1}^{d-1} n_j$ with n_j the number of samples in dimension j and the product is over the d-1 indirect dimensions of a d-dimensional experiment), and the factor of 2 per indirect dimension reflects quadrature detection. For a fairly typical uniformly sampled 3D experiment averaging two FIDs with 64 evolution times sampled in each of two indirect time dimensions, a t_{acq} of 0.6 s, t_{rc} of 1.2 s, t_{exp} is 16.4 hr. Contrast this time to an experiment in which the maximum evolution time in the indirect dimensions correspond to the Rovnyak limit[26] of $1.26T_2$ for optimizing sensitivity, or πT_2 for resolving components separated by the natural linewidth.

Typical ^{13}C and ^{15}N linewidths for a 20 kDa protein at 14.1 T (600 MHz for ^{1}H) are 17.5 and 5.8 Hz, respectively. The chemical shift dispersion for ^{13}C and ^{15}N is 10,500 and 2,100 Hz, respectively. With sampling intervals of 0.0952 ms and 0.476 ms required by the Nyquist condition, 241 samples in the ^{13}C dimension and 145 samples in the ^{15}N dimension would be required to sample uniformly to $1.26T_2$; 573 and 345 samples are needed to reach πT_2. t_{exp} for $1.26T_2$ is 140 hr, or 795 hr for πT_2. The total number of samples required for uniform sampling to either limit greatly exceeds the number typically acquired or the time devoted to data collection[33]. This means that higher dimensionality experiments that employ uniform sampling are usually sub-optimal both in sensitivity per unit time and in resolution.

In their seminal application of NUS and MaxEnt, Barna et al.[34] demonstrated rather conservative coverage ranging from 0.25 to 0.125. More substantial reductions in sampling coverage have subsequently been reported for 3D and 4D experiments, with coverages well below 0.01 common[35], and reaching 0.0016[36]. NUS is not the only means for reducing the time required for multidimensional NMR experiments; in the SOFAST approach, the time between FIDs is reduced[37]. As SOFAST (and related methods) and NUS are complementary, they can be combined, achieving greater speedup than either approach alone[29,38].

Rovnyak et al. exploited NUS to resolve separate resonances reflecting magnetically inequivalent ^{17}O nuclei in the unit cell of hydroxyapatite crystals[39]. NUS has also been used to obtain high-resolution spectra for disordered proteins, which exhibit narrow spectral dispersion and hence crowded spectra[38,40,41]. The higher resolution afforded by NUS has also enabled novel assignment strategies for protein spectra that are not practical with uniform sampling[35,42]. For example 4D HCC(CO)NH-TOCSY spectra obtained using NUS for the three indirect dimensions

can be obtained in 1.5 days, a rather dramatic speedup, rendering high-resolution 4D experiments practical. The resulting resolution in the indirect dimensions is sufficient to capture unique carbon-proton connectivities, enabling a novel and efficient scheme for assigning protein side-chain resonances. Similar approaches have also been reported for backbone resonance assignment employing 3D experiments [42,43].

Examples illustrating important characteristics of NUS are shown in Figs. 4 and 5. Fig. 4 depicts 2D cross-sections through the HNCO spectrum of ubiquitin; 1D cross-sections at the frequency indicated by the dashed line are shown above each contour plot. Panel A shows the spectrum obtained using conventional uniform sampling and DFT processing, requiring 34 hours to complete the experiment. Panel B shows the results using uniform sampling with a truncated data set requiring 25 minutes to collect; the reduction in resolution is severe. Panels C and D show the results from an experiment also using 25 minutes of measuring time, but using NUS instead of uniform sampling. In Panel C MaxEnt is used to compute the spectrum; in panel D, nuDFT was employed. Fig. 5 depicts contour plots for 2D projections of the HNCO spectrum for ubiquitin onto the ^{1}H-^{13}C plane. Panel A shows the projection obtained using the uniformly-sampled data set of Fig4a. Panels B and C show the projections obtained from the truncated uniform and NUS data sets (corresponding to Figs 4B and 4C, respectively). Panel D shows the projection obtained nuDFT instead of MaxEnt (4C); 4D dramatically reveals the poverty of nuDFT, because the coherent sampling artifacts are accumulated by the projection. The nearly 90-fold reduction in experiment time, with no loss of sensitivity or resolution, makes a convincing case for NUS and MaxEnt.

Concluding remarks. The debate over optimal sampling schemes and the best reconstruction method is far from settled. A comprehensive critical comparison remains elusive, in part because

metrics of spectral quality (sensitivity, resolution) that are valid for linear methods, such as DFT, are frequently not suitable for non-Fourier methods. In addition to a lack of consensus on appropriate metrics, critical comparison is made difficult by the absence of a "shared task" comprised of a standard set of data. Nevertheless, a number of basic tenets of NUS have achieved broad consensus. It is abundantly clear that NUS approaches are essential for fully realizing the potential resolution afforded by modern high field magnets in the indirect dimensions of multidimensional experiments. Also widely appreciated is the importance of randomness in the design of sampling schemes in order to minimize sampling artifacts. Although fine details of the design of efficient sampling schemes remain an active area of exploration, it is understood that sampling more frequently when the signal envelope has greater amplitude improves sensitivity. The flexibility of NUS approaches for reducing measuring time, increasing resolution, or enhancing sensitivity, and in some cases two or more of these simultaneously, make NUS an indispensable tool for enhancing the utility and power of multidimensional NMR. These improvements will enable new and challenging applications of multidimensional NMR to larger, more complex, less abundant, and fleetingly stable systems.

Figure 1. Schematic diagram for MaxEnt. MaxEnt reconstruction begins with empirical data and a preliminary trial spectrum **f** (typically a blank spectrum). **f** is inverted (DFT^{-1}) to create "mock" data (**m**) that is compared with the empirical data (**d**). An update to the trial spectrum is computed by searching along the gradients of the entropy and the constraint (the agreement between the empirical and mock data). The algorithm converges to the unique MaxEnt solution when the gradient of the objective function $O = S-\lambda C$ is zero and the gradients of S and C are antiparallel.

Figure 2. Nonlinear transformation for analytic MaxEnt. In the special case that MaxEnt is used to compute the *n*-element spectrum from an *n*-element FID, the MaxEnt spectrum is equivalent to applying a monotonic nonlinear transformation to the DFT of the FID. The nonlinear transformation (A) depends on the value of λ; in the limit of large λ (the constraint weighted more heavily than the entropy), the transformation becomes nearly linear. For small λ, the transformation scales down small amplitude signals more than large amplitude signals (B).

Figure 3. Examples of NUS schemes and PSFs. Examples of NUS sampling functions and PSFs in two nonuniformly sampled dimensions. Purely random sampling (third row) yields the smallest sampling artifacts for a given level of coverage (the central zero-frequency component is extremely narrow, making it hard to discern the red dot). Values for the relative sensitivity and PSR (both unitless) are displayed in the upper left and lower right, respectively, for each PSF.

Figure 4. 3D HNCO spectra for Ubiquitin at 14.1 T (600 MHz for ^{1}H). $^{13}C/^{1}H$ planar cross-sections at a ^{1}H frequency of 8.14 ppm. The one-dimensional cross-sections through the plots are of the ^{13}C row at the weakest peak (^{15}N frequency of 120.5 ppm), scaled so that the highest and lowest amplitudes are aligned. A) Using a full dataset, 6656 data samples, processed using LP extrapolation in each indirect dimension, shifted sine-bell apodization, and DFT; this data set required 36 hours of data collection B) using 100 uniformly-sampled data points (10 increments in each indirect dimension); this data requires 25 minutes of data collection. The spectrum was computed by LP extrapolation in each indirect dimension, apodization using a shifted sine-bell, and DFT. C) using NUS, with 100 random samples selected according to an exponentially weighted distribution, reconstructed using MaxEnt; this data also requires 25 minutes of data collection. D) Same as C), except using nuDFT instead of MaxEnt. The weak peak near the center of the ^{13}C trace in A is the "tail" of a peak at a nearby ^{1}H frequency; this disappears because of the narrower peaks in C.

Figure 5. 3D HNCO spectra for Ubiquitin at 14.1 T (600 MHz for ^{1}H). Projections of the full spectrum onto the $^{13}C/^{1}H$ plane. The one-dimensional cross-sections through the plots are expansions depicted by the rectangular boxes, scaled to align the maxima and minima. A) Using the full 6656 sample data set, processed using LP extrapolation in the indirect dimensions, shifted sine-bell apodization, and DFT (36 hours of data collection) B) using 400 uniformly-sampled data points (20 increments in each indirect dimension, 25 minutes of data collection). The spectrum was computed by LP extrapolation in each indirect dimension, apodization using a shifted sine-bell, and DFT. C) using NUS, with 400 random samples selected according to an exponentially weighted distribution, reconstructed using MaxEnt; this data also requires 25 minutes of data collection. D) Same as C), except using nuDFT instead of MaxEnt.

AUTHOR INFORMATION

Corresponding Author

* University of Connecticut Health Center, Dept. of Molecular, Microbial, and Structural Biology, 263 Farmington Ave., Farmington, CT 06030-3305 USA

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENT

We are grateful to many generous colleagues for inspiration and fruitful collaboration: Haribabu Arthanari, David Donoho, Dominque Frueh, Vitaliy Gorbatyuk, Sven Hyberts, Glenn King, Tatyana Polenova, David Rovnyak, Peter Schmieder, Jim Sun, and Gerhard Wagner. Support from the US National Institutes of Health (RR-020125; GM-047467; GM-102366) and the Australian Research Council (FTl10100925) is gratefully acknowledged.

Jeffrey Hoch is Professor of Molecular, Microbial, and Structural Biology and Director of the Gregory P. Mullen NMR Structural Biology Facility at the University of Connecticut Health Center. He obtained is Ph.D. from Harvard (1983) with Martin Karplus and Chris Dobson.

Mark Maciejewski is Assistant Professor of Molecular, Microbial, and Structural Biology and Manager of the Gregory P. Mullen NMR Structural Biology Facility at the University of

Connecticut Health Center. For the past 20 years he has utilized NMR for structural characterization of biomolecules and worked on NMR data processing.

Mehdi Mobli is an Australian Research Council Future Fellow at the Centre for Advanced Imaging at the University of Queensland, Australia. His research interests lie in developing and applying advanced NMR methods to solving difficult biological problems, including the structure and function of membrane proteins as well as characterizing weak intermolecular interactions.

Adam Schuyler was born in New York City, NY in 1978. He received his BA with Honors from Williams College (2000) and his Ph.D. from Johns Hopkins University (2006). He is presently a postdoctoral research fellow at the University of Connecticut Health Center. His research interests include computational aspects of structural biology.

Alan Stern is a Staff Computational Scientist at Harvard's Rowland Institute. He has worked on NMR data processing for over 25 years.

REFERENCES

(1) Ernst, R. R.; Anderson, W. A. Application of Fourier Transform Spectroscopy to Magnetic Resonance. *Rev. Sci. Instrum.* **1966**, *37*, 93-102.

(2) Jeener, J.: Oral Presentation, Ampere International Summer School, Basko Polje Yugoslavia. 1971.

(3) Hoch, J. C.; Stern, A. S.: *NMR Data Processing*; Wiley-Liss: New York, 1996.

(4) Mobli, M.; Hoch, J. C.; King, G. F.: Fast Acquisition Methods in Multidimensional NMR. In *Advances in Biomedical spectroscopy*; Dingley, A. J., Pascal, S. M., Eds.; IOS Press: Amsterdam, 2011.

(5) Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379-423.

(6) Skilling, J.; Bryan, R. Maximum entropy image reconstruction: general algorithm. *Mon. Not. R. Astron. Soc.* **1984**, *211*, 111-124.

(7) Schmieder, P.; Stern, A. S.; Wagner, G.; Hoch, J. C. Quantification of maximum-entropy spectrum reconstructions. *J Magn Reson* **1997**, *125*, 332-339.

(8) Mobli, M.; Maciejewski, M. W.; Gryk, M. R.; Hoch, J. C. An automated tool for maximum entropy reconstruction of biomolecular NMR spectra. *Nat Methods* **2007**, *4*, 467-468.

(9) Donoho, D. L.; Johnstone, I. M.; Stern, A. S.; Hoch, J. C. Does the maximum entropy method improve sensitivity? *Proc Natl Acad Sci U S A* **1990**, *87*, 5066-5068.

(10) Jones, J. A.; Hore, P. J. The Maximum-Entropy Method - Appearance and Reality. *J. Magn. Reson.* **1991**, *92*, 363-376.

(11) Jones, J. A.; Hore, P. J. The Maximum-Entropy Method and Fourier Transformation Compared. *J. Magn. Reson.* **1991**, *92*, 276-292.

(12) Paramasivam, S.; Suiter, C. L.; Hou, G.; Sun, S.; Palmer, M.; Hoch, J. C.; Rovnyak, D.; Polenova, T. Enhanced sensitivity by nonuniform sampling enables multidimensional MAS NMR spectroscopy of protein assemblies. *J Phys Chem B* **2012**, *116*, 7416-7427.

(13) Hyberts, S. G.; Heffron, G. J.; Tarragona, N. G.; Solanky, K.; Edmonds, K. A.; Luithardt, H.; Fejzo, J.; Chorev, M.; Aktas, H.; Colson, K.; Falchuk, K. H.; Halperin, J. A.; Wagner, G. Ultrahigh-resolution (1)H-(13)C HSQC spectra of metabolite mixtures using nonlinear sampling and forward maximum entropy reconstruction. *J Am Chem Soc* **2007**, *129*, 5108-5116.

(14) Schmieder, P.; Stern, A. S.; Wagner, G.; Hoch, J. C. Quantification of Maximum Entropy Reconstructions. *J. Magn. Reson.* **1997**, *125*, 332-339.

(15) Maciejewski, M. W.; Mobli, M.; Schuyler, A. D.; Stern, A. S.; Hoch, J. C.: Data Sampling in Multidimensional NMR: Fundamentals and Strategies. In *Top Curr Chem*; Billeter, M., Orekhov, V., Eds.; Springer: Berlin, 2012; pp 1-29.

(16) Barna, J. C. J.; Laue, E. D.; Mayger, M. R.; Skilling, J.; Worrall, S. J. P. Exponential sampling: an alternative method for sampling in two dimensional NMR experiments. *J. Magn. Reson.* **1987**, *73*, 69.

(17) Ernst, R. R. Sensitivity Enhancement in Magnetic Resonance. *Adv. Magn. Reson.* **1966**, *2*, 1-135.

(18) Schmieder, P.; Stern, A. S.; Wagner, G.; Hoch, J. C. Application of nonlinear sampling schemes to COSY-type spectra. *J. Biomol. NMR* **1993**, *3*, 569-576.

(19) Schmieder, P.; Stern, A. S.; Wagner, G.; Hoch, J. C. Improved resolution in triple-resonance spectra by nonlinear sampling in the constant-time domain. *J. Biomol. NMR* **1994**, *4*, 483-490.

(20) Hyberts, S. G.; Takeuchi, K.; Wagner, G. Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data. *J Am Chem Soc* **2010**, *132*, 2145-2147.

(21) Bretthorst, G. L. Nonuniform sampling: Bandwidth and Aliasing. *Concepts Magn. Reson.* **2008**, *32A*, 417-435.

(22) Kazimierczuk, K.; Kozminski, W.; Zhukov, I. Two-dimensional Fourier transform of arbitrarily sampled NMR data sets. *J Magn Reson* **2006**, *179*, 323-328.

(23) Schuyler, A. D.; Maciejewski, M. W.; Arthanari, H.; Hoch, J. C. Knowledge-based nonuniform sampling in multidimensional NMR. *J Biomol NMR* **2011**, *50*, 247-262.

(24) Hyberts, S. G.; Robson, S. A.; Wagner, G. Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra. *J Biomol NMR* **2013**, *55*, 167-178.

(25) Hoch, J. C.; Maciejewski, M. W.; Filipovic, B. Randomization improves sparse sampling in multidimensional NMR. *J Magn Reson* **2008**, *193*, 317-320.

(26) Rovnyak, D.; Hoch, J. C.; Stern, A. S.; Wagner, G. Resolution and sensitivity of high field nuclear magnetic resonance spectroscopy. *J Biomol NMR* **2004**, *30*, 1-10.

(27) Lustig, M.; Donoho, D.; Pauly, J. M. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn Reson Med* **2007**, *58*, 1182-1195.

(28) Donoho, D. L.; Johnstone, I. M.; Stern, A. S.; Hoch, J. C. Maximum Entropy and the Nearly Black Object (with discussion). *J Roy Stat Soc B* **1992**, *54*, 41-81.

(29) Maciejewski, M. W.; Qui, H. Z.; Rujan, I.; Mobli, M.; Hoch, J. C. Nonuniform sampling and spectral aliasing. *J Magn Reson* **2009**.

(30) Eghbalnia, H. R.; Bahrami, A.; Tonelli, M.; Hallenga, K.; Markley, J. L. High-resolution iterative frequency identification for NMR as a general strategy for multidimensional data collection. *J Am Chem Soc* **2005**, *127*, 12528-12536.

(31) Wong, L. E.; Masse, J. E.; Jaravine, V.; Orekhov, V.; Pervushin, K. Automatic assignment of protein backbone resonances by direct spectrum inspection in targeted acquisition of NMR data. *J Biomol NMR* **2008**, *42*, 77-86.

(32) Cornilescu, G.; Bahrami, A.; Tonelli, M.; Markley, J. L.; Eghbalnia, H. R. HIFI-C: a robust and fast method for determining NMR couplings from adaptive 3D to 2D projections. *J Biomol NMR* **2007**, *38*, 341-351.

(33) Hyberts, S. G.; Arthanari, H.; Wagner, G. Applications of non-uniform sampling and processing. *Top Curr Chem* **2012**, *316*, 125-148.

(34) Barna, J. C. J.; Laue, E. D.; Mayger, M. R.; Skilling, J.; Worrall, S. J. P. Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments. *J Magn Reson* **1987**, *73*, 69-77.

(35) Mobli, M.; Stern, A. S.; Bermel, W.; King, G. F.; Hoch, J. C. A non-uniformly sampled 4D HCC(CO)NH-TOCSY experiment processed using maximum entropy for rapid protein sidechain assignment. *J Magn Reson* **2010**, *204*, 160-164.

(36) Stanek, J.; Augustyniak, R.; Koźmiński, W. Suppression of sampling artefacts in high-resolution four-dimensional NMR spectra using signal separation algorithm. *J Magn Reson* **2012**, *214*, 91-102.

(37) Schanda, P.; Kupce, E.; Brutscher, B. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR* **2005**, *33*, 199-211.

(38) Marion, D. Combining methods for speeding up multi-dimensional acquisition. Sparse sampling and fast pulsing methods for unfolded proteins. *J Magn Reson* **2010**, *206*, 81-87.

(39) Rovnyak, D.; Filip, C.; Itin, B.; Stern, A. S.; Wagner, G.; Griffin, R. G.; Hoch, J. C. Multiple-quantum magic-angle spinning spectroscopy using nonlinear sampling. *J Magn Reson* **2003**, *161*, 43-55.

(40) Pannetier, N.; Houben, K.; Blanchard, L.; Marion, D. Optimized 3D-NMR sampling for resonance assignment of partially unfolded proteins. *J Magn Reson* **2007**, *186*, 142-149.

(41) Motáčková, V.; Nováček, J.; Zawadzka-Kazimierczuk, A.; Kazimierczuk, K.; Zídek, L.; Sanderová, H.; Krásný, L.; Koźmiński, W.; Sklenář, V. Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution-enhanced 5D experiments. *J Biomol NMR* **2010**, *48*, 169-177.

(42) Sun, Z. Y.; Frueh, D. P.; Selenko, P.; Hoch, J. C.; Wagner, G. Fast assignment of 15N-HSQC peaks using high-resolution 3D HNcocaNH experiments with non-uniform sampling. *J Biomol NMR* **2005**, *33*, 43-50.

(43) Frueh, D. P.; Sun, Z. Y.; Vosburg, D. A.; Walsh, C. T.; Hoch, J. C.; Wagner, G. Non-uniformly sampled double-TROSY hNcaNH experiments for NMR sequential assignments of large proteins. *J Am Chem Soc* **2006**, *128*, 5757-5763.