



Resolution and sensitivity of high field nuclear magnetic resonance spectroscopy

D. Rovnyak^a, J.C. Hoch^b, A.S. Stern^c & G. Wagner^c

^aDepartment of Chemistry, Bucknell University, Lewisburg, PA 17837, U.S.A.; ^bDepartment of Molecular, Microbial and Structural Biology, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, CT 06030-3305. ^cRowland Institute at Harvard, 100 Edwin H. Land Boulevard, Cambridge MA 02142, U.S.A.; ^dDepartment of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115, U.S.A.; MIT/Harvard Center for Magnetic Resonance, 150 Albany Street, Cambridge, MA 02139, U.S.A.

Received 22 december 2003; Accepted 22 April 2004

Key words: maximum entropy, NMR, non-linear sampling, protein, resolution, sensitivity

Abstract

The arrival of very high field magnets and cryogenic circuitries, and the development of relaxation-optimized pulse sequences have added powerful tools for increasing sensitivity and resolution in NMR studies of biomacromolecules. The potential of these advances is not fully realized in practice, however, since current experimental protocols do not permit sufficient data sampling for optimal resolution in the indirect dimensions. Here we analyze quantitatively how increasing resolution in indirect dimensions affects the S/N ratio and compare this with currently used sampling routines. Optimal resolution would require sampling up to $\sim 3R_2^{-1}$, and the S/N reaches a maximum at $\sim 1.2R_2^{-1}$. Currently used data acquisition protocols rarely sample beyond $0.4R_2^{-1}$, and extending evolution times would result in prohibitively long experiments. We show that a general solution to this problem is to use non-uniform sampling, where only a small subset of data points in the indirect sampling space are measured, and possibly different numbers of transients are collected for different evolution times. Coupled with modern methods of spectrum analysis, this strategy delivers substantially improved resolution and/or reduced measuring times compared to uniform sampling, without compromising sensitivity. Higher resolution in the indirect dimensions will facilitate the use of automated assignment programs.

The capabilities of nuclear magnetic resonance (NMR) spectroscopy for investigating biomolecules have improved dramatically since the pulsed Fourier-transform method (FT-NMR) was first developed (Ernst and Anderson, 1966). Isotopic enrichment schemes, cryogenic detection circuits, and superconducting magnets operating at fields up to 21.1 Tesla have enabled the routine study of large biomolecules at sub-millimolar concentration in aqueous solutions. Despite these developments, biomolecules larger than 30 kDa yield spectra that are sufficiently crowded to make structural studies difficult. Deuteration strategies and transverse relaxation optimized spectroscopy (TROSY) mitigate the problem of

spectral overlap by reducing the natural linewidths (Venters et al., 2002; Pervushin et al., 1997). Yet the potential of such methods to provide higher resolution using high field instruments is not routinely realized due to the practical inability to record sufficient data. In this paper we discuss such practical barriers and a general strategy for surmounting them.

The resolution of an NMR experiment is determined by a number of factors. The digital resolution, Δf , is given by

$$\Delta f = \frac{1}{N\Delta t}, \quad (1)$$

where Δt is the time between samples (the dwell time) of the free induction decay (FID) and N is the num-

ber of points in the spectrum. The spectral width (the range of frequencies that can be detected without aliasing) is equal to $1/\Delta t$. (Note that N is generally larger than N_s , the number of time-domain data samples.) Since the spectral dispersion of nuclear resonances increases linearly with magnetic field strength, so must the spectral width. However, coupling constants, and in a first approximation the line widths, do not vary with field strength, so the resolution required to distinguish coupled peaks is independent of field strength. As an example, at a field strength of 11.7 T ($\nu(^1\text{H}) = 500$ MHz) a 36-ppm ^{15}N spectral width corresponds to 1800 Hz, or $\Delta t = 556 \mu\text{s}$. To achieve a digital resolution of 2 Hz requires $N \geq 900$. At 21.1 T, 36 ppm corresponds to 3240 Hz, or $\Delta t = 309 \mu\text{s}$, requiring $N \geq 1620$ for 2-Hz digital resolution. For a three-dimensional experiment, the total number of points in each $f_1 - f_2$ plane required to maintain the same digital resolution increases by a factor of $(9/5)^2 = 3.24$ at 900 MHz as compared to 500 MHz. Given the amount of instrument time typically available, it is impractical to increase the number of FIDs by the same factor. Hence the ratio of N_s to N is forced to be lower at higher field strengths, to the detriment of the achievable resolution.

The ability to resolve peaks depends not only on the digital resolution, but also on the natural linewidth of the peaks,

$$L = \frac{R_2}{\pi}, \quad (2)$$

where R_2 is the relaxation rate of the detected coherence. For the special case $N = N_s$, the digital resolution (Equation 1) is related to the natural linewidth by the formula:

$$\Delta f = L \left(\frac{\pi}{R_2} \right) \left(\frac{1}{N_s \Delta t} \right) = L \left(\frac{\pi R_2^{-1}}{t_{\max}} \right), \quad (3)$$

where t_{\max} is approximately the largest delay time sampled, equal to $N_s \Delta t$ for uniform sampling. Ideally, the digital resolution should be comparable to the natural linewidth, which would require t_{\max} to be as long as πR_2^{-1} . However, there is no point in making Δf significantly smaller, which means that t_{\max} should not be much larger than πR_2^{-1} . It is common practice to set t_{\max} considerably smaller than πR_2^{-1} , and increasing t_{\max} by collecting additional samples would lead to a genuine improvement in resolution.

This analysis ignores the use of nonlinear techniques, such as extrapolation by linear prediction (LP), which in the case of high S/N are capable of increases

in resolution without the need for additional sampling. The formal results on the influence of t_{\max} on sensitivity and resolution that we present here are indicative of the *increase* in t_{\max} needed to maintain resolution as the magnetic field is increased, but the specific number of samples needed will depend on the extent of LP extrapolation, if employed. The use of nonuniform sampling, in contrast to LP, permits an increase in t_{\max} without increasing N_s . A detailed analysis of the influence of N_s on resolution using LP was reported previously (Stern et al., 2003).

Table 1 lists the values of N yielding $t_{\max} = \pi R_2^{-1}/2$ (sufficient to distinguish signals separated by twice the natural linewidths, and yielding nearly optimal S/N, see below). The linewidth estimates were obtained by empirically analyzing correlation times for several proteins as a function of molecular weight (Wagner, 1997) and calculating relaxation rates using standard methods (Peng and Wagner, 1992; Yamazaki et al., 1994; Pervushin et al., 1997; Venters et al., 2002; Cavanagh et al., 1996). Results are shown for (a) antiphase $^{15}\text{N}_x^1\text{H}_z$ coherence (b) $^{15}\text{N}_x^1\text{H}_z$ coherence via TROSY and (c) $^{13}\text{C}_x^2\text{H}_z$ coherence (details in the figure caption). These computations provide useful guidelines for average linewidths that may be achieved, but effects from remote nuclei, internal dynamics, and secondary and tertiary structure will cause further perturbations in observed correlation times (τ_c) and decay rates (R_2).

More efficient relaxation at higher fields increases R_2 values slightly, but this effect does not scale linearly. On the other hand, deuteration and TROSY experiments yield decreased R_2 values ($R_2^{-1} \sim 10^2$ ms) for ^{15}N and ^{13}C evolution in proteins with molecular weights of 20 kDa and higher. Even so, the guidelines in Table 1 indicate that data collection needs to be increased by a factor of 2–20 from the 32–64 samples that are typically obtained in indirect evolution periods of triple resonance experiments. It is clear from Table 1 that typical values of t_{\max} are very short compared to R_2^{-1} values even at a field strength of 14 Tesla ($\nu(^1\text{H}) = 600$ MHz). In principle, spectral folding (which increases Δt and thus t_{\max} for fixed N_s) can bring the requirements down to the 32–64 range of N_s , but such folding is not generally acceptable. Zero-filling (extending the measured data by appending zeros) can be used to increase N , but does not improve the ability to resolve closely-spaced resonances. Although LP can improve the resolution of a spectrum, it has limits and disadvantages; it assumes

Table 1. Number of uniform intervals (N) required to sample to an ^{15}N evolution period of $t_{\pi/2} = \pi R_2^{-1}/2$ (ms) for several protein sizes and static fields, using computed line widths for (a) antiphase $^{15}\text{N}_x^1\text{H}_z$ coherence assuming an internuclear distance $d_{\text{NH}} = 1.02 \text{ \AA}$ and a symmetric chemical shift tensor with $\Delta\sigma(^{15}\text{N}) = -160.0$ ppm (b) $^{15}\text{N}_x^1\text{H}_z$ coherence via TROSY with $d_{\text{NH}} = 1.02 \text{ \AA}$, $\Delta\sigma(^{15}\text{N}) = -160.0$ ppm, and taking the angle between the principal axis of the chemical shift tensor and the N-H bond to be 15 degrees and (c) $^{13}\text{C}_x^2\text{H}_z$ coherence using $d_{\text{CD}} = 1.05 \text{ \AA}$, $\Delta\sigma(^{13}\text{C}) = 25.0$ ppm but considering no additional effects from other nuclei outside the C-H spin pair. For (a) and (b) two non-covalently associated protons, 2.8 \AA each from the covalently attached proton, are included in the computation. Additionally, spectrometer inhomogeneity of 1.5 Hz is assumed for all cases. Spectral windows are taken to be 36 ppm for ^{15}N and 32 ppm for ^{13}C (e.g., $^{13}\text{C}_\alpha$ spectral region), based on the distribution of shifts reported in the BioMagResBank (Seavey et al., 1991)

MW/kDa (τ_c)	Static field									
	14.1 T		16.4 T		17.6 T		18.8 T		21.1 T	
	N	$t_{\pi/2}/\text{ms}$	N	$t_{\pi/2}/\text{ms}$	N	$t_{\pi/2}/\text{ms}$	N	$t_{\pi/2}/\text{ms}$	N	$t_{\pi/2}/\text{ms}$
(a) $^{15}\text{N}_x^1\text{H}_z$:										
10 (5.46 ns)	247	113	279	109	292	107	304	104	327	99.3
20 (9.98 ns)	170	77.5	188	73.6	196	71.5	202	69.4	214	65.1
30 (14.3 ns)	130	59.2	143	55.8	148	54.0	152	52.2	160	48.7
(b) $^{15}\text{N}_x^1\text{H}_z$ (TROSY):										
30 (14.3 ns)	312	143	394	154	437	159	476	163	554	169
50 (22.7 ns)	235	108	302	118	336	123	368	126	431	131
100 (43.0 ns)	147	67.2	192	75.0	215	78.5	237	81.4	280	85.0
(c) $^{13}\text{C}_x^2\text{H}_z$:										
30 (14.3 ns)	1041	215	1170	208	1233	205	1294	201	1395	193
50 (22.7 ns)	866	179	963	171	1008	167	1052	163	1120	155
100 (43.0 ns)	616	127	673	120	699	116	724	112	759	105

N = number of uniform intervals that span $t_{\pi/2} = \pi R_2^{-1}/2$ (ms); τ_c (ns) = rotational correlation time.

Lorentzian lineshapes and, depending on the signal-to-noise ratio (S/N) of the data, can introduce false peaks and small frequency errors (Stern et al., 2002). Expanding the dimensionality of an NMR experiment can improve the dispersion of signals, and thus the resolution. However it is not always feasible to use more than two indirect evolution periods since further coherence transfer and evolution steps may unacceptably attenuate the observable signal for large biomolecules.

Increasing t_{max} will undoubtedly improve the resolution in indirect dimensions. However the amount of sensitivity lost by sampling to longer times is less obvious. To analyze these losses, we show how the choice of t_{max} affects the S/N ratio in uniformly sampled spectra and conclude that increasing resolution only moderately diminishes the S/N . An ideal method for improving resolution should not sacrifice S/N , should be applicable to all pulse sequences, should use no more data acquisition time than current practice affords, should robustly handle data with low S/N , should apply to arbitrary lineshapes, and should not introduce errors or bias. Below we will describe

the technique of non-uniform sampling, which closely meets these criteria.

Theory

The total duration of an experiment is proportional to the number of FIDs, N_s , and the number of transients, n_t , recorded for each FID. The most important constraint on these values is the available instrument time. Increasing n_t will improve the S/N ratio, whereas increasing N will primarily improve resolution. As will be shown below, increasing N_s also increases the S/N ratio (not as much as increasing n_t) up to $t_{\text{max}} = 1.26R_2^{-1}$ but reduces the S/N for longer values. A balance must be struck based on the desired S/N ratio, the desired resolution, the inherent S/N of the sample, and the natural linewidths. We will first consider the situation where experiment time is not limited, exploring how sensitivity and resolution vary with N_s for fixed n_t .

The height, S , of a line in an FT-NMR spectrum is proportional to the integral of its corresponding time

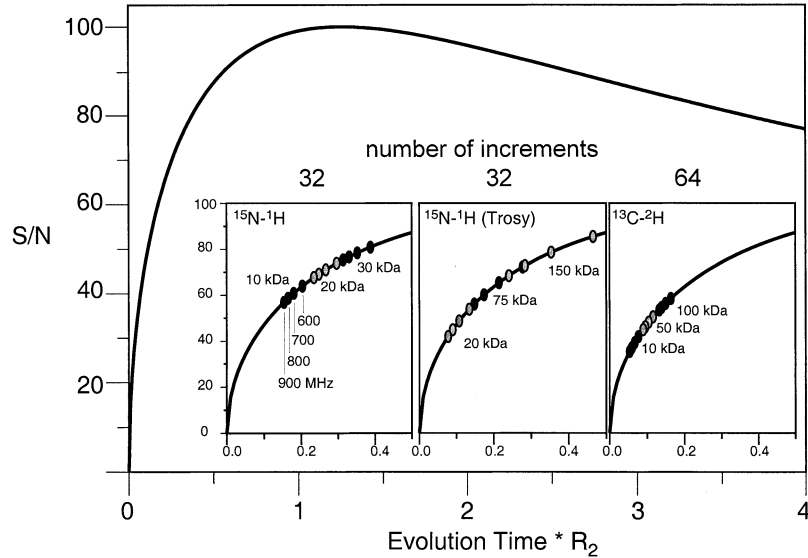


Figure 1. The signal to noise ratio as a function of t_{\max} , given by Equation 6, is plotted for an exponentially decaying signal containing Gaussian distributed noise. Inset: evolution periods that would result from using 32 samples and a spectral width of 36 ppm (e.g., protein ^{15}N amide region) are shown for ^{15}N - $^1\text{H}_z$ evolution (left panel), and TROSY- ^{15}N - $^1\text{H}_z$ evolution (middle panel). In the right panel, ^{13}C - $^2\text{H}_z$ evolution periods are shown assuming 64 samples and a spectral width of 32 ppm (e.g., protein C_α region). Positions of evolution times relative to R_2^{-1} were estimated for different molecular weights and field strengths as indicated.

domain signal over the acquisition time, t_{\max} . The dependence on t_{\max} and R_2 can be approximated by

$$S \propto \int_0^{t_{\max}} e^{-tR_2} dt. \quad (4)$$

The amount of noise for the same acquisition period is

$$\text{Noise} \propto \sqrt{t_{\max}}. \quad (5)$$

The S/N is then

$$S_N(t_{\max}) \propto \frac{(1 - e^{-t_{\max}R_2})}{R_2\sqrt{t_{\max}}}, \quad (6)$$

which yields a maximum when

$$t_{\max} = R_2^{-1} \ln(2R_2 t_{\max} + 1) \quad (7)$$

or $t_{\max} \approx 1.26R_2^{-1}$, since a closed solution for Equation 7 is not possible. A similar result is obtained for discrete sampling (Stern et al., 2002). Equation 6 is essentially a special case of the S/N per unit measurement time, used by Ernst and coworkers to demonstrate the sensitivity of Fourier spectroscopy (Ernst et al., 1987). Since sensitivity is defined to be S/N per unit time, we obtain the sensitivity by dividing Equation 6 by $\sqrt{t_{\max}}$. The preceding analysis can be validated also by noticing that if a non-decaying signal is assumed in Equation 4 (i.e $R_2 = 0$) then one finds the well known result that $S_N(t_{\max}) \propto \sqrt{t_{\max}}$.

Equation 6 is plotted in Figure 1 and we have found it to be an accurate model of experimentally acquired data, even in the presence of non-ideal lineshapes and other experimental artifacts (Figure 2). Insets in Figure 1 illustrate the times t_{\max} that would result when using 32–64 samples for the classes of experiments considered here. It is evident in all cases of Table 1 and Figure 1 that $\Delta f \gg L$. It is worth noting that our analysis is related to but different from that of Levitt et al. (1984) who considered the S/N per unit time, as a function of spectral width (or dwell time). By contrast, we are considering the total S/N , with the spectral width held fixed, as a function of t_{\max} .

We may now evaluate the S/N gain or loss from extending the evolution time by a factor n using uniform sampling,

$$\varepsilon_{\text{time}} = \frac{(1 - e^{-nt_{\max}})}{(1 - e^{-t_{\max}})} \frac{1}{\sqrt{n}}, \quad (8)$$

where $R_2 = 1$. Alternatively, the number of transients could be increased by a factor of n , leading to an enhancement $\varepsilon_{\text{transients}} = \sqrt{n}$. We then express the *penalty* in S/N for acquiring a decaying signal with uniform sampling to longer times versus acquiring additional transients as

$$\varepsilon_1(n, t) = \frac{\varepsilon_{\text{time}}}{\varepsilon_{\text{transients}}} = \frac{(1 - e^{-nt_{\max}})}{(1 - e^{-t_{\max}})} \frac{1}{n}, \quad (9)$$

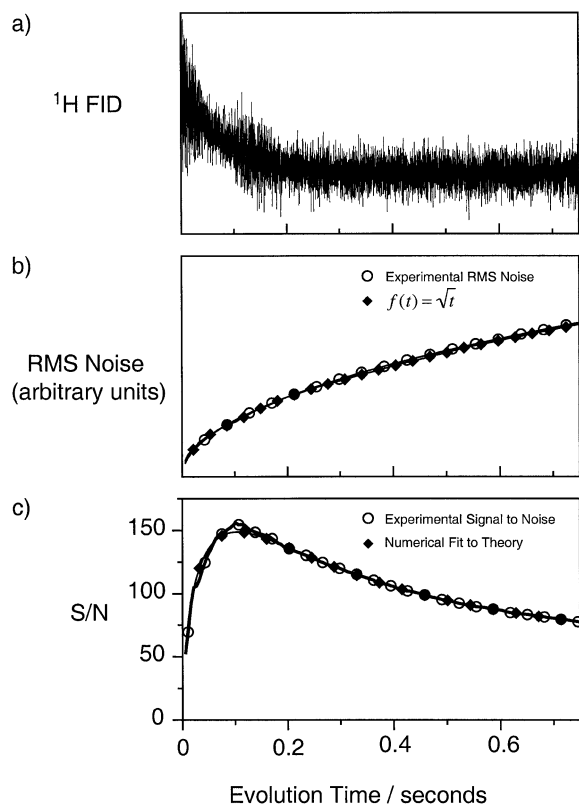


Figure 2. Analysis of an experimental decaying signal containing noise: (a) The ^1H time domain signal from the residual water in a sample of 99% D_2O doped with 0.1 mg/ml GdCl_3 , following a $0.1 \mu\text{s}$ excitation pulse using a UnityPlus (Varian Inc.) spectrometer operating at $\nu(^1\text{H}) = 400 \text{ MHz}$. The first 0.75 seconds of the FID are displayed, however the full FID used 16384 samples for an acquisition time of 1.365 seconds, and also includes a small ethanol signal that is useful to evaluate the shimming quality. The residual water line was shimmed to a non-spinning line width (full width at half maximum) of 3.88 Hz, as determined by nonlinear least squares fitting of the frequency domain line shape to a Lorentzian function (b) the RMS noise level as a function of acquisition time is plotted for the experimental data (open circles); the function $f(t) = \sqrt{t}$ is superimposed without fitting (filled diamonds); (c) the signal to noise ratio is plotted for the experimental data (open circles), and for a nonlinear least squares fit of the data to Equation 6 (filled diamonds), which resulted in a linewidth of 3.84 Hz.

which is plotted for several values of n in Figure 3. Although a factor of $n = 2$ avoids serious decreases in S/N , we will show that this rarely suffices for obtaining the maximum resolution available. For example, if $t_{\text{max}} = R_2^{-1}/2$, a signal must be extended by a factor of $n = 6$ to gain the full resolution; this would give just 40% of the S/N that would have resulted if 6 times as many transients had been collected. We note that a typical procedure for setting up a n -D NMR experiment is to first determine the number of transients that provide

acceptable S/N , and set the number of increments in the indirect evolution periods to fill the allotted measuring time. This procedure is analogous to choosing short evolution periods and minimizing n in order to reduce signal losses, which trades away resolution in favor of sensitivity.

The preceding analysis assumes that samples are collected at uniform intervals with a fixed number of transients for each sample. We refer to this as uniform sampling. By contrast, non-uniform sampling can be defined as collecting different numbers of transients for different sample times. As a practical matter, the number of transients must be an integer multiple of the length of the minimum phase cycle. Frequently the minimum phase cycle yields adequate sensitivity; in these experiments the number of transients collected for each sample time is either zero or the length of the minimum phase cycle. The obvious advantage of nonuniform sampling is that t_{max} can be larger than $N_s \Delta t$ (since some of the intermediate times are not sampled), yielding improved resolution for a given N_s .

A ‘sampling schedule’ is a specification of the number of transients collected for each sample time. In this paper we only consider sampling schedules in which the number of transients is zero or the length of the minimum phase cycle. Thus a sampling schedule will simply be a non-consecutive list of evolution times. To optimize S/N , an efficient strategy is for the sampling schedule to include more samples when the signal intensity is high and fewer when it is low (Barnett et al., 1987). By analogy with the matched filter, for exponentially decaying signals this leads to sampling schedules with an exponentially decaying density of samples. Figure 4a shows the S/N computed for several exponentially decaying signals using a range of exponential sampling schedules. The signals have line widths of 5 Hz, 10 Hz, and 15 Hz and the curves were computed assuming a 36 ppm ^{15}N spectral width and a static field of 21.1 T. Figure 4c illustrates the sampling schedules. Each schedule used 64 samples distributed over a progressively greater range of sample times, up to a factor of 4 times the span of $64 * \Delta t$. Increasing t_{max} decreases the S/N by up to 10% for the 5 Hz signal and 20% for the 15 Hz signal. In general, as the sampling schedule scaling factor increases, the rate at which S/N is reduced by progressively extending t_{max} with non-uniform sampling decreases. We show in Figure 4b the S/N of TROSY- $^{15}\text{N}_x$ - $^1\text{H}_z$ signals computed using the same R_2 values as in Table 1 for a 50 kDa protein, and using the sampling schedules in Figure 4c. When increasing the field from 14.1 to

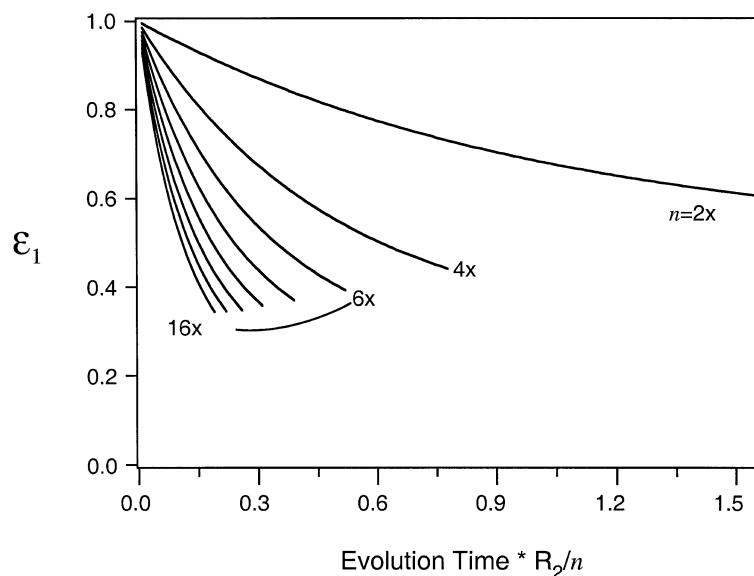


Figure 3. The relationship between signal to noise and resolution as a function of signal evolution time is illustrated with plots of Equation 9 showing the decrease in S/N from acquiring a signal further by a factor n versus acquiring n times more transients for $n = 2, 4, 6, 8, 10, 12, 14,$ and 16 . Each curve is plotted to a maximum time $t = \pi R_2^{-1}/n$.

21.1 Tesla, R_2 decreases due to the TROSY effect while Δt must be decreased to maintain a constant spectral window. Thus, the S/N decreases more slowly with t_{\max} at higher fields (Figure 4b). Generally it is evident from Figure 4 that the use of non-uniform sampling to increase t_{\max} (and thus the resolution) does not greatly diminish sensitivity, particularly when modern experiments at high fields are applicable to molecules with very small R_2 values.

The discrete Fourier transform (DFT) algorithm may only be applied to uniformly sampled data, so an alternative algorithm for spectral estimation is required to take advantage of the combined sensitivity and resolution offered by non-uniformly sampled data. In this communication we use maximum entropy (MaxEnt) reconstruction (Hoch and Stern, 1996), which has been shown to provide highly accurate and sensitive spectral estimations (Stern et al., 2002). The MaxEnt algorithm we use here reconstructs the (hyper) complex spectrum with the highest entropy that is consistent with the data, and is different from the Burg MEM algorithm (Burg, 1978), which yields the power spectrum and has been known to occasionally generate false positive peaks. MaxEnt reconstruction of non-uniformly sampled data is also available in the GIFA package (Pons et al., 1996). However, GIFA uses a different algorithm and a different entropy functional.

Thus, the results obtained using GIFA may not be the same as those using RNMRTK.

Another method that has been applied recently to non-uniformly acquired multi-dimensional NMR experiments, termed three-way decomposition, is implemented in the program MUNIN (Orekhov et al., 2001). This approach decomposes NMR spectra into sums of components and appears to be well suited for analysis of NOE-HSQC data sets, by exploiting simultaneously the high density of information in all three dimensions. A detailed comparison of methods is outside the scope of this communication, however we call attention to two important features shared by both MaxEnt reconstruction and three-way decomposition, that no assumptions are made regarding the number or shape of the signals in the data, and that the computations are readily accessible to desktop computing.

The higher resolution afforded by using non-uniform sampling to extend t_{\max} can significantly aid in the task of obtaining with automated assignment programs complete spectral assignments for heteronuclei (e.g., ^{13}C , ^{15}N) in large biomolecules (Hyberts and Wagner, 2003). To document the enhanced ‘assignability’ due to the use of non-uniform sampling and MaxEnt reconstruction we demonstrate the resolution gain obtainable with this approach. Figure 5a–c compares $^{15}\text{N}(^1\text{H})$ heteronuclear single quantum

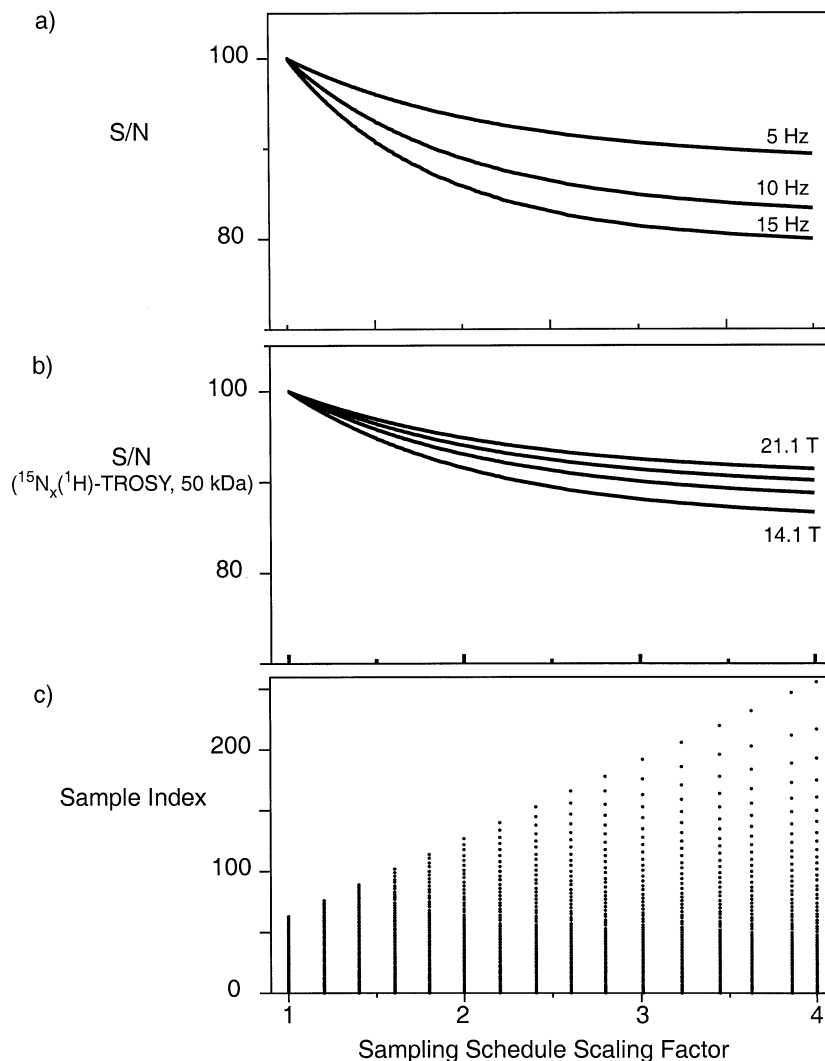


Figure 4. The dependence of S/N upon a range of sampling schedules is illustrated for (a) 5 Hz, 10 Hz, and 15 Hz ^{15}N signals assuming a spectral width of 36 ppm at 21.1 Tesla, and for (b) $^{15}\text{N}_x^1\text{H}_z$ TROSY signals corresponding to R_2 values computed in Table 1 for a 50 kDa protein and for static fields of 14.1 T, 16.4 T, 18.8 T, and 21.1 T respectively. Representative sampling schedules that were used to compute the S/N values in (a–b) are shown in (c) for scaling of t_{max} by factors up to 4 while keeping the overall number of samples constant at 64. In (a) and (b) S/N values are normalized to 1 and are not meant to indicate absolute sensitivities.

coherence spectra acquired uniformly and nonuniformly on a 15 kDa protein ($u\text{-}^{15}\text{N}$ enrichment) on a 17.6 Tesla spectrometer ($\nu(^1\text{H}) = 750$ MHz). The expected ^{15}N anti-phase line width for a general 15 kDa protein is estimated to be about 6 Hz, giving a predicted decay time constant for this sample $1/R_2 = (\pi * LW) = 1/0.053\text{s}$. Assuming a spectral width of 36 ppm, an acquisition with 32 uniform increments would span just $t_{\text{max}} = 0.0117$ s. To explore the resolution achieved using this sampling interval, we first acquired a spectrum with uniform sampling to

$t_{\text{max}} = 0.064$ s in the indirect evolution period using $N = 256$ samples. A crowded region from this spectrum is shown in Figure 5a. For Figure 5a, the signal was extended to 512 points with linear prediction, apodized with a single 90-degree shifted sine bell function, and extended with zeros to a final size of 1024 points before applying the discrete Fourier transform (DFT). Cross-sections at $\delta(^1\text{H}) = 8.41$ ppm (not shown) and $\delta(^1\text{H}) = 8.36$ ppm contain ^{15}N resonances separated by 18.5 Hz and 17.0 Hz, respectively, which are successfully resolved by this experiment

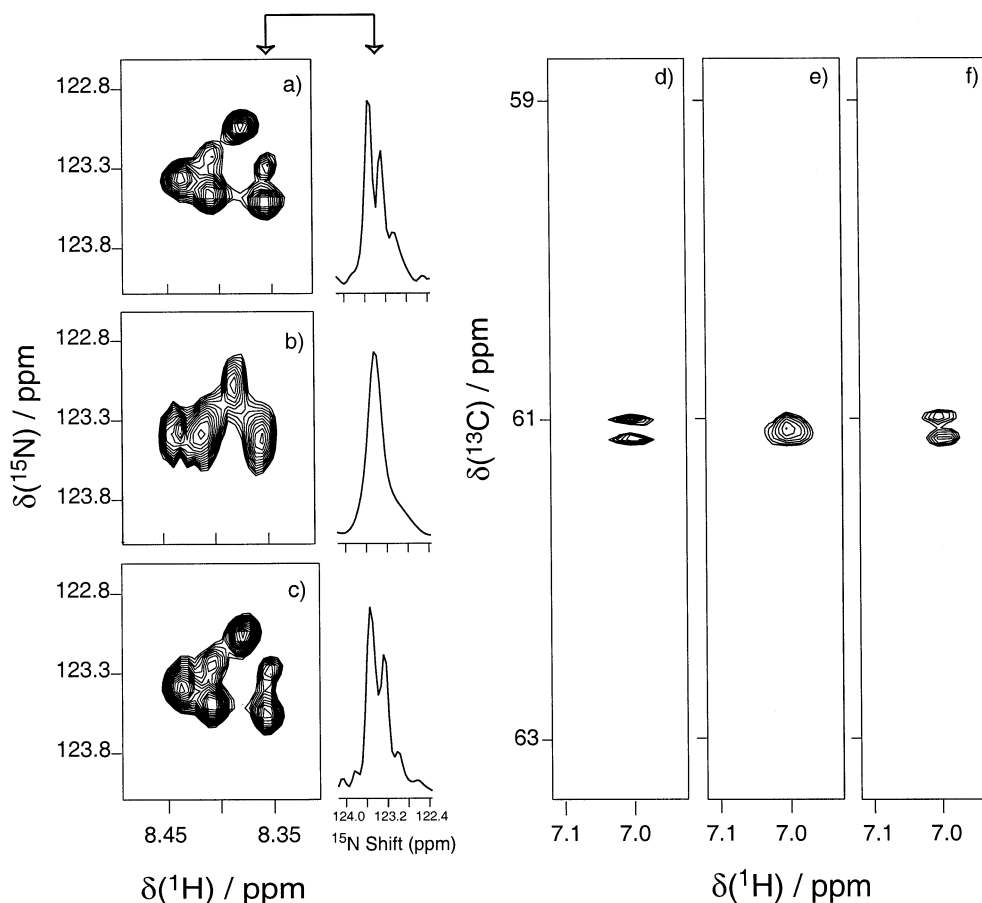


Figure 5. A comparison of $^{15}\text{N}(^1\text{H})$ heteronuclear single quantum correlation spectra acquired at 17.6 Tesla, $\nu(^1\text{H}) = 750$ MHz with: (a) 256 uniform samples up to $t_{\text{max}} = 0.064$ s, (b) 64 uniform samples up to $t_{\text{max}} = 0.016$ s and (c) 64 samples distributed up to $t_{\text{max}} = 0.064$ s. Cross-sections at $\delta(^1\text{H}) = 8.36$ ppm are also shown for (a), (b) and (c). The signals in (a) and (b) were extended to twice their respective sizes with linear prediction, apodized with a single 90° shifted sine bell function, zero-filled to 1024 total points, and processed with the fast Fourier transform. The signal in (c) was processed by row-wise maximum entropy reconstruction using a constant-lambda algorithm (Schmieder et al., 1997, Hoch and Stern, 1996), and required less than one minute of processing time using a 1.5 GHz Athlon (AMD) processor (Red Hat Linux 7.2). The list of samples for (c) is {1–7, 9–12, 15, 16, 18, 20, 21, 22, 24, 25, 27, 28, 30, 31, 33, 35, 36, 38, 40, 42, 44, 46, 47, 51, 52, 56, 59, 61, 64, 67, 68, 71, 74, 78, 82, 86, 89, 93, 98, 102, 108, 112, 120, 125, 132, 142, 153, 161, 178, 188, 192, 209, 218, 239, 256}; (d–f) Simulation of the resolution gain using nonuniform sampling and MaxEnt reconstruction in the ^{13}C dimension of an HNCA experiment assuming a 50 kDa protein and sample deuteration (see Table 1 for relaxation parameters). In (d) 256 uniform samples were extended with linear prediction to 512, treated with a 3 Hz exponential window function and zero-filled to 1024 samples prior to the DFT along the ^{13}C dimension. In (e) 64 uniform samples were doubled by linear prediction, treated with a 3 Hz exponential window and zero-filled to 1024 prior to DFT. In (f) MaxEnt reconstruction was applied to 64 samples distributed nonuniformly up to and including $N = 256$.

($\Delta f = 1/0.064 = 15.6$ Hz). Figure 5b shows the same spectral region when the first 64 samples of the uniform signal are extended by linear prediction to 128 samples, apodized with a single 90° -degree shifted sine bell function, and extended with zeros to 1024 points before applying the DFT. Clearly the data are no longer sufficient to resolve distinct signals in this region ($\Delta f = 1/0.016 = 62.5$ Hz). Figure 5c shows the spectrum obtained by MaxEnt reconstruction of a separately acquired data set where 64 samples

were distributed over the evolution period of 0.064 s. The spectrum in Figure 3c was acquired in the same elapsed time as that shown in Figure 5b, but provides high resolution comparable to that observed in Figure 5a since samples are retained up to and including $N_s = 256$. The sampling schedule used for Figure 5c, listed in the figure caption, follows an exponential distribution; for example just 15 samples are distributed in the range [102,256]. MaxEnt reconstruction is relatively insensitive to truncated or missing data so that

no window function is required prior to running the MaxEnt algorithm. In contrast, signal extrapolation of uniformly sampled data by linear prediction generally requires subsequent convolution of the signal with a window function to avoid artifacts in a DFT spectrum. Whereas some apodization functions can be applied to uniformly sampled data prior to a DFT to suppress artifacts while artificially improving the line widths of the resonances (Ernst et al., 1987), such convolutions can distort line shapes, enhance noise, and suggest a higher resolution in the data than may be justified. When faced with a spectral region that is poorly resolved, parametric fitting techniques may be applied to the data (Andrec and Prestegard, 1998).

To illustrate this, we also carried out simulations of a region of a ^{13}C - ^1H correlation spectrum, shown in Figure 5d–f, that would be analogous to a strip plot for an HNCA experiment. Two signals separated by 28 Hz were used in the indirect dimension, each having an intrinsic line width of ~ 3.2 Hz, and randomly distributed noise was injected such that each signal has an S/N of about 15:1 in the full length ($N = 256$) FID. Their relaxation rates in the indirect evolution period were taken to be the same and corresponded to $^{13}\text{C}_x$ - $^2\text{H}_z$ coherence for a 50 kDa protein at 21.1 Tesla (see Table 1). A 32 ppm spectral width (7243 Hz) was also used. Similar to Figure 5a–c, uniformly sampled data for $N = 256$ and $N = 64$ were processed using linear prediction to double the data length in the indirect dimension. For uniform sampling with $N_s = 256$, processed strips are shown in Figure 5d, and the resolution obtainable with $N_s = 64$ with uniform and non-uniform sampling is shown in strips e and f of Figure 5, respectively. It is clear from Figure 5e that, regardless of the use of linear prediction, the two signals are not resolvable due to the low experimental value of $t_{\text{max}} = 8.8$ ms. However by using nonuniform sampling to distribute 64 samples up to $t_{\text{max}} = 256$, the two signals are resolved (Figure 5f). The line widths in Figure 5f are not identical to those obtained in Figure 5d, however some broadening of lines in MaxEnt spectra can occur with noisy or truncated data due to the smaller number of experimental constraints on the true lineshape (Rovnyak et al., 2003).

Figure 5 demonstrates the use of non-uniform sampling schedules to achieve resolution comparable to extending a uniform acquisition by a factor of 4. It is clear that the assignability of such spectra is significantly enhanced by non-uniformly distributing samples in indirect dimensions to long evolution times. In our experience, non-uniform sampling can be used to ex-

tend the evolution time by a factor of 2–4 over the range spanned by a uniform distribution of the same number of samples, significantly improving the extent to which indirect evolution periods are acquired relative to $t_{\text{max}} = \pi R_2^{-1}$.

Application of non-uniform sampling to both indirect dimensions of 3D experiments will be described in a subsequent publication, however we call attention to the impracticality of using uniform sampling to recover high resolution at high fields. Given that typical measurement times for 3D experiments can be 1–5 days, if each indirect period were extended uniformly by a factor of three, then 9–45 days would be needed per 3D experiment. In contrast non-uniform sampling redistributes samples over the increased evolution periods without altering the measurement time. The limits for distributing samples non-uniformly depend largely on the intrinsic S/N ratio of the time domain data and the number of samples used (Rovnyak et al., 2003). For routine application of non-uniform sampling and MaxEnt reconstruction, we advise adhering to factors of 2–4 for the ratio of the non-uniformly sampled period to the uniformly sampled period with an equivalent number of samples.

NMR systems operating in the range 18.8–21.1 T are rapidly becoming available to most researchers as components of multi-user regional and national shared instrument facilities. We have shown that uniform incrementation cannot reasonably recover the resolution offered at high fields, but significant improvements in resolution can be obtained when MaxEnt reconstruction is applied to data that is non-uniformly acquired to include samples at long evolution times. We expect this methodology will be indispensable for increasing the size and complexity of biomacromolecules accessible for study by NMR, and for obtaining the greatest value from high field instruments.

Acknowledgements

We gratefully acknowledge the support of the Rowland Institute for Science (J.H. and A.S.), the National Institutes of Health grants GM47467 (D.R., G.W., J.H., A.S.), and CA89940 (D.R.), and RR 00995.

References

- Andrec, M. and Prestegard, J.H. (1998) *J. Magn. Reson.*, **130**, 217–32.

- Barna, J.C.J., Laue, E.D., Mayger, M.R., Skilling, J. and Worrall, S.J. P. (1987) *J. Magn. Reson.*, **73**, 69–77.
- Burg, J.P. (1978) In *Modern Spectrum Analysis*, Childers, D.J. (Ed.), I.E.E.E. Press, New York, pp. 42–48.
- Cavanagh, J., Fairbrother, W.J., Palmer, A.G. and Skelton, N.J. (1996) *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, Inc., San Diego, CA.
- Ernst, R.R. and Anderson, W.A. (1966) *Rev. Sci. Instr.*, **37**, 93–102.
- Ernst, R.R., Bodenhausen, G. and Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford.
- Hoch, J.C. and Stern, A.S. (1996) *NMR Data Processing*, Wiley-Liss, New York.
- Hyberts, S.G. and Wagner, G. (2003) *J. Biomol. NMR*, **26**, 335–44.
- Levitt, M.H., Bodenhausen, G. and Ernst, R.R. (1984) *J. Magn. Reson.*, **58**, 462–472.
- Orekhov, V.Y., Ibraghimov, I.V. and Billeter, M. (2001) *J. Biomol. NMR*, **20**, 49–60.
- Peng, J.W. and Wagner, G. (1992) *J. Magn. Reson.*, **98**, 308–332.
- Pervushin, K., Riek, R., Wider, G. and Wüthrich, K. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 12366–12371.
- Pons, J.-L., Malliavin, T.E. and Delsuc, M.A. (1996) *J. Biomol. NMR*, **8**, 445–452.
- Rovnyak, D., Filip, C., Itin, B., Stern, A.S., Wagner, G., Griffin, R. G. and Hoch, J.C. (2003) *J. Magn. Reson.*, **161**, 43–55.
- Schmieder, P., Stern, A.S., Wagner, G. and Hoch, J.C. (1997) *J. Magn. Reson.*, **125**, 332–339.
- Seavey, B.R., Farr, E.A., Westler, W.M. and Markley, J.L. (1991) *J. Biomol. NMR*, **1**, 217–236.
- Stern, A.S., Li, K.-B. and Hoch, J.C. (2002) *J. Am. Chem. Soc.*, **124**, 1982–1993.
- Venters, R.A., Thompson, R. and Cavanagh, J. (2002) *J. Mol. Struct.*, **602–603**, 275–292.
- Wagner, G. (1997) *Nat. Struct. Biol.*, **4**, 841–844.
- Yamazaki, T., Lee, W., Arrowsmith, C.H., Muhandiram, D.R. and Kay, L. E. (1994) *J. Am. Chem. Soc.*, **116**, 11655–11666.