Journal of Magnetic Resonance 261 (2015) 19–26

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

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

Deterministic multidimensional nonuniform gap sampling

Bradley Worley, Robert Powers *

Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588-0304, United States

ARTICLE INFO

Article history:
Received 5 May 2015
Revised 14 September 2015
Available online 23 October 2015

Keywords:
NMR
NUS
Poisson-gap
Deterministic sampling

ABSTRACT

Born from empirical observations in nonuniformly sampled multidimensional NMR data relating to gaps between sampled points, the Poisson-gap sampling method has enjoyed widespread use in biomolecular NMR. While the majority of nonuniform sampling schemes are fully randomly drawn from probability densities that vary over a Nyquist grid, the Poisson-gap scheme employs constrained random deviates to minimize the gaps between sampled grid points. We describe a deterministic gap sampling method, based on the average behavior of Poisson-gap sampling, which performs comparably to its random counterpart with the additional benefit of completely deterministic behavior. We also introduce a general algorithm for multidimensional nonuniform sampling based on a gap equation, and apply it to yield a deterministic sampling scheme that combines burst-mode sampling features with those of Poisson-gap schemes. Finally, we derive a relationship between stochastic gap equations and the expectation value of their sampling probability densities.

1. Introduction

The use of nonuniform sampling (NUS) in multidimensional NMR is rapidly becoming standard practice in most biomolecular solution-state experiments, thanks in large part to recent developments in fast reconstruction algorithms, novel sampling schemes, and the continually declining cost of computing power [1]. The potential benefits of collecting a subset of the full Nyquist grid – including increased sensitivity and signal-to-noise, improved resolution, and reduced experiment time – have received significant attention [2–6] in recent years as a consequence.

One intriguing result of recent investigations into the parameters of NUS experiments is the use of random deviates for generating sampling schedules [7]. In fully random sampling schemes, a subset of Nyquist grid points are drawn from a probability density function that varies over the grid, producing a sampling schedule with a desired distribution of points. Common fully random sampling schemes utilize uniform, exponential, Gaussian and envelope-matched probability densities [3,8]. While randomization is a simple means of reducing the artifacts due to aliasing of nonuniformly spaced samples, it turns the already complex task of schedule generation into that of selecting a schedule from an ensemble of possibilities, each of which performs differently in practice [9,10]. Several ad hoc metrics have been proposed to assess relative performance of sampling schedules, but no universally accepted metric exists to guide the selection of a stochastic schedule from its ensemble [10,11]. Without a priori knowledge of the frequency and decay rate distributions of the signals to be measured, it is difficult to reliably quantify sampling schedule performance [1,8]. As a result, numerous recent attempts have been made to reduce or remove pseudorandom seed-dependent variability from nonuniform sampling algorithms [9,10,12,13]. Such efforts are an important step towards increasing the practical utility of nonuniform sampling in everyday spectroscopic experiments.

A prominent method designed to reduce seed-dependent variability in pseudorandomly constructed schedules is Poisson-gap sampling. Through an empirical analysis of Forward Maximum Entropy (FM) reconstructions of randomly sampled data, Hyberts et al. proposed the use of constrained Poisson random deviates to define the *gaps* between sampled points in a Nyquist grid [9]. The FM reconstruction residuals of these so-named Poisson-gap schedules exhibited a markedly lower dependence on seed value than unconstrained random sampling methods. While Poisson-gap sampling yields high-quality reconstructions of NUS spectral data, its average behavior is not well-understood, its implementation for multidimensional Nyquist grids is unclear [14–16], and its relationship – if any – to fully random sampling is unknown. To meet this need, we describe in detail the deterministic generation of sinusoidally weighted multidimensional gap schedules that model the average behavior of stochastic Poisson-gap (PG)

* Corresponding author at: University of Nebraska-Lincoln, Department of Chemistry, 722 Hamilton Hall, Lincoln, NE 68588-0304, United States. Fax: +1 (402) 472 9402.
E-mail address: rpowers3@unl.edu (R. Powers).

http://dx.doi.org/10.1016/j.jmr.2015.09.016
1090-7807/

sampling. We also derive an expectation sampling probability distribution that describes the average weighting obtained using Poisson-gap sampling schedules.

Among the myriad of different sampling schemes proposed for NUS data collection [17], burst-mode sampling similarly concerns itself with gaps between sampled grid points. Unlike Poisson-gap sampling, which aims to minimize the *length* of gaps, burst-mode sampling aims to minimize the *number* of gaps while keeping the effective dwell time low [18]. We leverage the complementarity of burst-mode and Poisson-gap sampling in our deterministic gap sampling algorithm to describe a novel sampling scheme that simultaneously seeks to bias sample collection to early times, minimize the number of long gaps between densely sampled regions, and minimize the largest gap length in the schedule. Our new method, called sine-burst (SB) sampling, exhibits the high performance of Poisson-gap sampling while retaining the bijective mapping between inputs and outputs offered by deterministic methods.

2. Theory

2.1. Poisson-gap sequences

Gap schedules on a one-dimensional Nyquist grid are effectively finite integer sequences, computed from the following recurrence relation:

$$x_{i+1} = x_i + \lfloor g(x_i) \rfloor + 1 \tag{1}$$

where x_i is the grid index of the i-th term in the sequence and $g(x_i)$ is the "gap equation" that defines the distance between terms. The first term in the sequence, x_1, is set to 1, and subsequent terms are computed until their value exceeds N, the size of the grid. The gap equation $g(x)$ may be any non-negative function, and may be loosely interpreted as inversely related to the local sampling density at the grid index x_i. Thus, when the gap equation equals zero for all grid indices, gap sampling will yield a uniformly sampled grid.

Poisson-gap sequences treat the gap equation as a Poisson random deviate having a rate parameter that varies as either a quarter- or half-sinusoid over the grid indices:

$$g_{PG}(x_i) \sim \mathrm{Pois}\left\{\Lambda \sin\left(\frac{\pi}{2}\theta_i\right)\right\} \tag{2}$$

where Λ is a scaling factor that determines the global sampling density and θ_i is the fractional grid index that varies from 0 to 1 over the grid extents:

$$\theta_i = \frac{x_i}{N} \tag{3}$$

In all following descriptions of Poisson-gap methods, we shall restrict our attention to rate parameters which vary as quarter-sinusoids, where the fractional grid index is multiplied by a factor of one-half π. This choice of sinusoidal weight produces schedules which are heavily biased to earlier grid points. Using a factor of π produces half-sinusoidal rate parameters and schedules that are more densely sampled at both early and late grid points.

Because the expected value of a Poisson distribution is equal to its rate parameter, we may trivially construct a deterministic sinusoidally weighted gap sampler (sine-gap; SG) by setting the gap equation equal to the scaled quarter-sinusoid from Eq. (2), as follows:

$$g_{SG}(x_i) = \Lambda \sin\left(\frac{\pi}{2}\theta_i\right) \tag{4}$$

By construction, gap sampling schedules computed according to g_{SG} will describe the average behavior of g_{PG}. This is easily verified in one dimension by generating a sufficiently large set of stochastic Poisson-gap schedules and comparing the mean value of each sequence term to that of a sine-gap schedule (*cf.* Supplementary Methods and Fig. S-1). Sine-gap schedules lie centrally within the Poisson-gap ensemble, while other schedules unrelated to Poisson-gap deviate substantially from the confidence region of the ensemble.

2.2. Multidimensional gap sampling

Gap schedules on a Nyquist grid having at least two dimensions are generated by placing multiple one-dimensional sub-schedules onto the grid, each with a different direction and offset from the grid origin. In practice, this process is accomplished recursively, with planes built up from vectors, cubes built up from planes, and so forth. Initially, recursion begins on the entire grid. At each level of recursion, sub-grids are constructed by 'masking off' each available grid direction in turn and constructing the remaining unmasked directions. For example, a three-dimensional *xyz* cube will be constructed from repeated sequences of *yz*, *xz*, and *xy* planes, and each *xy* plane will be constructed from repeated sequences of *y* and *x* vectors. Once a round of sub-grid construction has been performed along each direction, the sub-grid offset is incremented and the process is repeated until no more sub-grids remain at the current level of recursion. For a more precise definition of the recursive algorithm, see Code Listing S-1 in the Supplementary Information.

Creation of multidimensional gap schedules requires a slight modification to the fractional index, which now assumes the following form:

$$\theta_i = \frac{x_i + \sum_{d=1}^{D} O_d}{\sum_{d=1}^{D} N_d} \tag{5}$$

where O_d and N_d are the origin and grid size along direction d, respectively. Eq. (5) is referred to as "ADD" mode in the context of Poisson-gap sampling, and effectively results in multidimensional schedules that exhibit triangular forms [14]. It is worthy of mention that, in the one-dimensional case, Eq. (5) reduces to Eq. (3).

Finally, whether the Nyquist grid is one- or many-dimensional, a value of the global scaling factor Λ must be determined that yields the desired number of sampled grid points. Our gap sampling implementation, like the existing Poisson-gap method, iteratively rebuilds new schedules until Λ has been suitably optimized. Our implementation uses a heuristic search method that adjusts Λ based on the relative difference between the desired and obtained global sampling density at each iteration.

2.3. Burst augmentation

Recent statistical descriptions of the discrete Fourier transform have shown that the bandwidth of a nonuniformly sampled signal is related to the greatest common factor of the gaps between sampled points [19]. One proposed method of increasing bandwidth and reducing artifacts in NUS data is to sample in multiple short bursts having zero gap length [18]. Using gap sampling, this may be accomplished by modulating the gap equation between zero and its maximum value several times over the Nyquist grid, like so:

$$g_{SB}(x_i; d) = \Lambda \sin\left(\frac{\pi}{2}\theta_i\right) \sin^2\left(\frac{\pi}{4} N_d \theta_i\right) \tag{6}$$

The sine-burst gap equation g_{SB} combines the sinusoidal forward-biasing and minimized gap lengths of Poisson-gap sampling with the minimized effective dwell time of burst-mode sampling, and does not require the use of random deviates to achieve reasonable artifact suppression.

2.4. Expectation sampling distributions

One disadvantage of stochastic gap equations is that they provide no direct measure of how likely each Nyquist grid point is to be sampled. While one may speculate on the approximate weighting obtained by a given gap equation, quantitation of the expectation of the sampling distribution requires the construction and averaging of a large number of sampling schedules (*cf.* Figs. S-2, S-3 and S-4). Fortunately, the expectation sampling distribution of a given gap equation may be analytically obtained by computing the probability of sampling each point on the grid using a recursive formula. We define an expectation sampling distribution $p(i)$ that varies over a one-dimensional Nyquist grid of N points as follows:

$$p(i) = \sum_{k=1}^{i-1} p(i|i-k) \cdot p(i-k) \tag{7}$$

where $p(i|i-k)$ is the probability of grid point i being emitted from grid point $i-k$, which requires a gap of size $k-1$:

$$p(i|i-k) = \Pr\{\lfloor g(i-k) \rfloor = k-1\} \tag{8}$$

In other words, the probability of sampling any given grid point is the weighted sum of the probabilities of arriving at that point from all prior points. By substituting Eq. (2) for the gap equation into Eqs. (7) and (8), we arrive at the sampling distribution of a one-dimensional Poisson-gap sequence:

$$p(i) = \sum_{k=1}^{i-1} \frac{\Lambda^{k-1}}{(k-1)!} \sin^{k-1}\left(\frac{\pi[i-k]}{2N}\right) \exp\left\{-\Lambda \sin\left(\frac{\pi[i-k]}{2N}\right)\right\} \cdot p(i-k) \tag{9}$$

As in the case of gap sampling, the sampling distribution produced by Eq. (9) is parameterized only by the scaling factor Λ, where larger values produce more forward-biased schedules. We refer to this equation as the "expectation" Poisson-gap sampling distribution because it describes the expected value of the probability of sampling any Nyquist grid point, and is not itself useful for generating schedules that obey g_{PG}. A more detailed derivation of Eq. (9) is provided in the Supplementary Information.

2.5. Multidimensional expectation sampling distributions

Extension of Eq. (7) to compute the expectation sampling distributions of stochastic gap equations in two or more dimensions follows from the fact that sampling along each direction is independent of other dimensions within our gap sampling framework. The probability of sampling any multidimensional grid point is therefore the sum of sampling that point along each grid direction. Supplementary Fig. S-3 illustrates the expectation Poisson-gap sampling distribution on two-dimensional Nyquist grids. It is important to note that the Poisson-gap sampler originally proposed by Hyberts et al. does not strictly follow our gap sampling algorithm, because its sampling of each dimension is dependent upon which points in other dimensions have been previously sampled. This divergence between multidimensional Poisson-gap and Poisson-gap constructed according to our algorithm is observed by comparison of Figs. S-3 and S-4, and is only truly apparent at very low sampling densities.

3. Materials and methods

3.1. Generation of deterministic schedules

Deterministic sine-gap and sine-burst schedules were constructed using a small C program which implements our recursive gap sampling algorithm described above. Schedules were generated at 30%, 10% and 5% sampling densities on one-dimensional grids having 1024 points and two-dimensional grids having 64×64 and 128×128 points. The first and third rows of Fig. 1 show the deterministic schedules resulting from g_{SG} and g_{SB} at 30% density on 128×128 grids, respectively, and Supplementary Fig. S-5 shows the schedules at 10% and 5% density.

3.2. Generation of stochastic schedules

Poisson-gap schedules were constructed using Java source code authored and provided by Hyberts et al. for generating multidimensional schedules (http://gwagner.med.harvard.edu/intranet/hmsIST/gensched_old.html). A small command-line wrapper was written to provide direct access to the core schedule generation functions without use of the graphical interface. Fifty thousand schedules were computed at each of the sampling densities and grid sizes listed above. Each schedule was generated with a unique, large, odd-valued seed number to ensure the broadest possible sampling of the PG ensemble. The second row of Fig. 1 shows an example two-dimensional Poisson-gap schedule at 30% sampling density. Supplementary Fig. S-5 additionally shows representative Poisson-gap schedules at 10% and 5% density.

3.3. Spectral data collection

Experiments were conducted on a Bruker Avance III HD 700 MHz spectrometer equipped with a 5 mm inverse quadruple-resonance (^{1}H, ^{13}C, ^{15}N, ^{31}P) cryoprobe with cooled ^{1}H and ^{13}C channels and a z-axis gradient. A high-resolution 2D ^{1}H–^{15}N HSQC NMR spectrum was collected at a temperature of 298.0 K on a sample of uniformly [^{15}N, ^{13}C]-labeled ubiquitin in aqueous phosphate buffer at pH 6.5. A 2D gradient-enhanced ^{1}H–^{15}N HSQC spectrum with improved sensitivity [20,21] was collected with 16 scans and 32 dummy scans over a uniform grid of 2048 and 1024 hypercomplex points along the ^{1}H and ^{15}N dimensions, respectively. Spectral windows were set to 3293 ± 4209 Hz along ^{1}H and 8514 ± 1419 Hz along ^{15}N. The spectrum was windowed with a squared-cosine function, Fourier-transformed and phase-corrected along ^{1}H to produce a half-transformed spectrum for IST reconstruction analysis (*vide infra*), and subsequently windowed and Fourier-transformed along ^{15}N to yield the "true" uniformly sampled 2D ^{1}H–^{15}N HSQC spectrum.

In addition, a 3D HNCA NMR spectrum was collected on the same uniformly [^{15}N, ^{13}C]-labeled ubiquitin sample. The spectrum was collected at 298.0 K with 16 scans and 32 dummy scans over a uniform grid of $1024 \times 64 \times 64$ hypercomplex points along the ^{1}H, ^{15}N and ^{13}C dimensions, respectively. Spectral windows were set to 3293 ± 4209 Hz along ^{1}H, 8514 ± 1419 Hz along ^{15}N, and 9508 ± 2818 Hz along ^{13}C. The spectrum was windowed with a squared-cosine function, Fourier-transformed and phase-corrected along ^{1}H to produce an F_3-transformed spectrum for IST reconstruction analysis, and subsequently windowed and Fourier-transformed along ^{15}N and ^{13}C to yield the "true" uniformly sampled 3D HNCA spectrum.

3.4. Computation of performance metrics

All computational analyses were performed using in-house developed C programs. An implementation of the hypercomplex algebra described by Schuyler et al. [22] was used to perform all spectral data processing. Iterative Soft Thresholding (IST) reconstructions of subsampled spectra were performed using the algorithm described by Stern et al. [23,24]. Impulse sets were generated for each constructed schedule by setting sampled grid points to one and skipped grid points to zero. At each sampling density and grid size for which schedules were created,

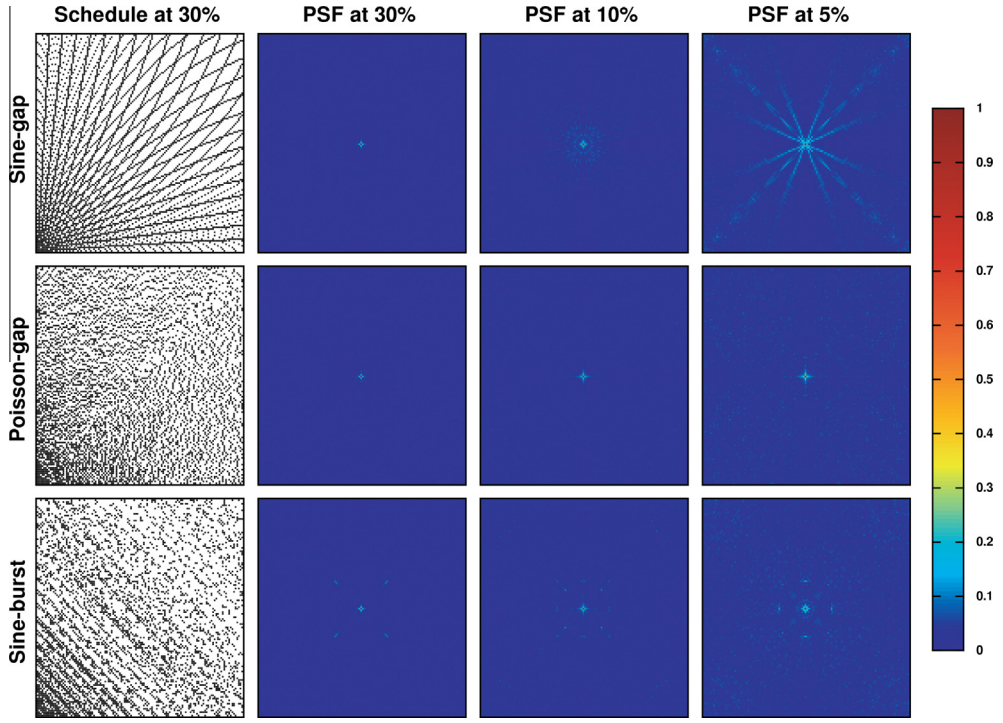

Fig. 1. Comparison of sine-gap, Poisson-gap and sine-burst sampling schedules and their resulting point spread functions at varying sampling densities, indicating close agreement between the sine-gap and Poisson-gap methods. The increased artifact intensity in the sine-gap schedule at 5% sampling density is due to slightly increased regularity of sampled grid points, which is reduced by Poisson-gap and sine-burst sampling. Grid sizes and point spread function colorings are the log-scaled versions of those found in Fig. 1 of [26] in order to emphasize low-intensity sampling artifacts. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

point-spread functions were calculated by hypercomplex discrete Fourier transformation of each schedule's impulse set. Point-spread functions for schedules built on two-dimensional grids are shown for each sampling density in Fig. 1. For one-dimensional schedules, reconstruction residuals were computed from a subset of 192 F_1 traces of the half-transformed HSQC spectrum. The traces were nonuniformly subsampled using sine-gap, sine-burst and Poisson-gap ($N = 10{,}000$) schedules and reconstructed with 400 iterations of IST at a threshold level of 98%. After reconstruction, the residual was calculated using the l_2-norm of the differences between the true and reconstructed signals. A convergence analysis was also performed (Supplementary Fig. S-6) to ensure convergence of IST to a stationary point, as measured by a lack of decrease in the l_2 error. Fig. 2A shows the distributions of IST reconstruction residuals from the HSQC traces, and example reconstructions from each sampling schedule at 5% density are illustrated in Fig. 3. See Fig. S-7 for a version of Fig. 3 with contours reduced ten-fold to expose low-intensity noise and reconstruction artifacts. Reconstructions of 10 F_2–F_1 planes of the F_3-transformed HNCA were also performed after nonuniformly subsampling using sine-gap schedules, sine-burst schedules, and a subset ($N = 10{,}000$) of the generated Poisson-gap schedules. Fig. 2B shows IST reconstruction residuals computed from the HNCA planes, and example reconstructions from each sampling schedule at 5% density are illustrated in Fig. 4.

3.5. Generation of peak-picking statistics

A summary of the relative HSQC peak-peaking performance for the IST reconstructions from each sampling schedule and at each

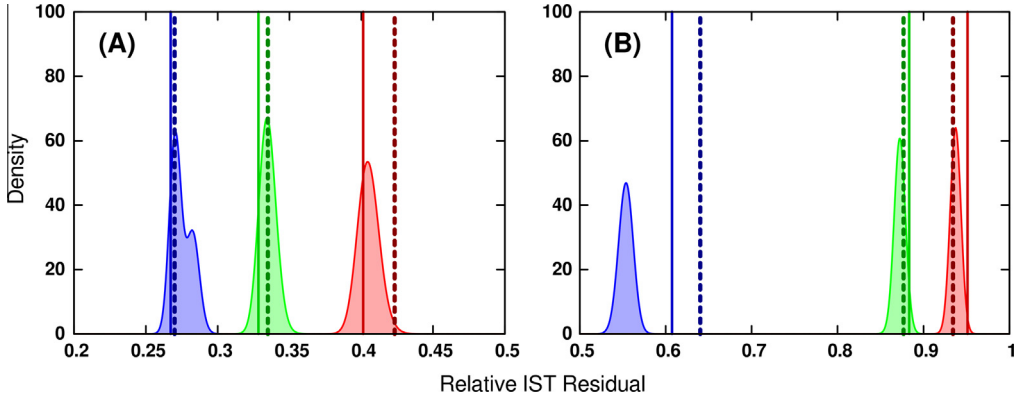

Fig. 2. Iterative Soft Thresholding reconstruction l_2 residuals of (A) 192 ^1H–^{15}N HSQC F_1 traces and (B) 10 HNCA F_2–F_1 planes from Poisson-gap schedules having sampling densities of 30% (blue), 10% (green) and 5% (red). Residuals of sine-gap and sine-burst schedules are shown as solid and dashed vertical lines, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

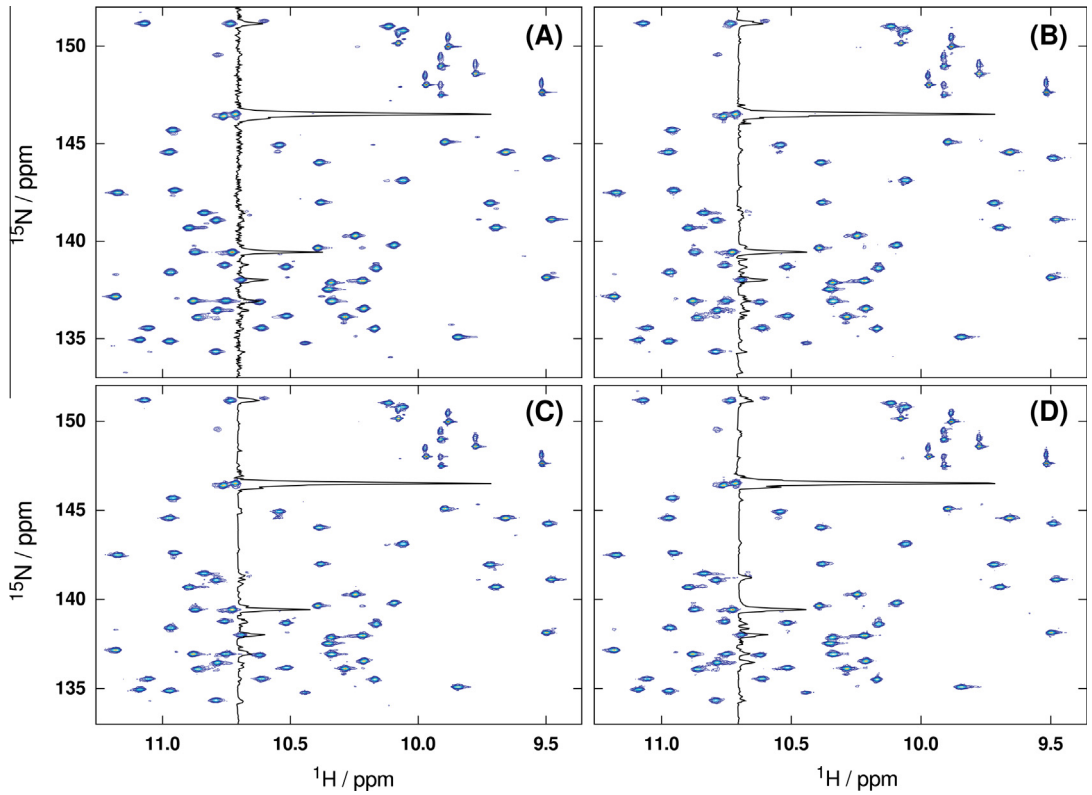

Fig. 3. Uniformly sampled (A) and IST reconstructed (B–D) 2D ^{1}H–^{15}N HSQC spectra of ubiquitin, indicating nearly equivalent performance of all three gap sampling methods at low (5%) sampling density. Spectra shown in (B) through (D) were reconstructed from nonuniformly subsampled copies of (A) using (B) Poisson-gap, (C) sine-gap and (D) sine-burst methods, respectively. All spectra are plotted with identical contour levels. For a rendering of the same spectra at very low contour levels, refer to Supplementary Fig. S-7.

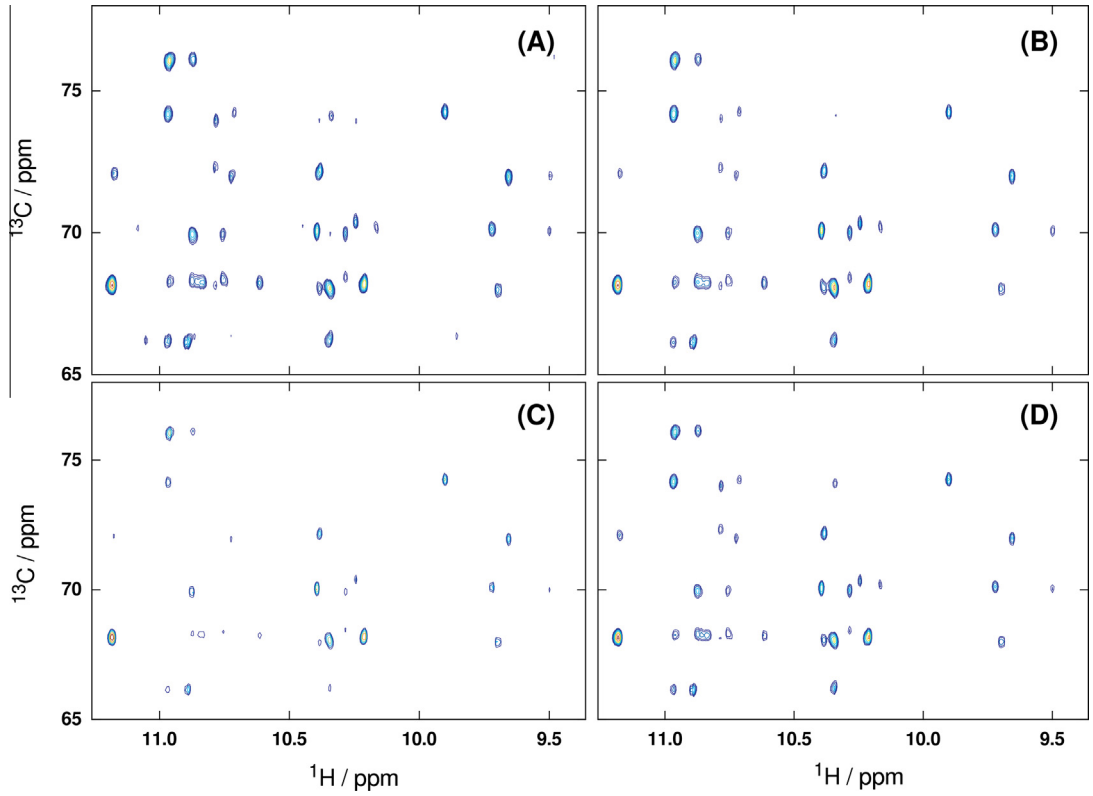

Fig. 4. Uniformly sampled (A) and IST reconstructed (B–D) 3D HNCA spectra of ubiquitin at low (5%) sampling density, projected along the ^{15}N dimension. Spectra shown in (B) through (D) were reconstructed from nonuniformly subsampled copies of (A) using (B) Poisson-gap, (C) sine-gap and (D) sine-burst methods, respectively. While sine-gap sampling (C) fails to adequately reproduce the spectrum due to its high sampling coherence, sine-burst sampling yields an essentially identical result to Poisson-gap sampling. All spectra are plotted with identical contour levels.

sampling density is listed in Table 1. For each 2D ^{1}H–^{15}N HSQC spectrum of ubiquitin reconstructed via Iterative Soft Thresholding (IST) at each sampling density and each sampling method, a set of quality statistics was computed. Peak lists were generated using the *peakHN.tcl* utility provided by NMRPipe [25], with a minimum intensity threshold of 3.0×10^7. Then, a greedy algorithm was used to generate maximum-cardinality bipartite matching between the peak list of each reconstructed spectrum and the peak list of the true spectrum. Chemical shift windows of 0.015 ppm and 0.08 ppm were used along the ^{1}H and ^{15}N dimensions, respectively, during matching. The number of peaks matched, lost and gained in the reconstructed spectra, relative to the true spectrum, were all counted. Lost peaks were any picked peaks in the true spectrum that had no match in the reconstruction. Gained peaks were any picked peaks in the reconstruction with no partner in the true spectrum. The intensities of all matched peaks in each reconstruction were then compared against their true intensities through the computation of a Pearson correlation coefficient, r_{int}, which effectively summarizes the linearity of the reconstruction algorithm as a function of sampling schedule. Finally, root-mean-square chemical shift deviations of all matched peaks along the ^{1}H dimension (d_H) and the ^{15}N dimension (d_N) were also computed. Identical procedures and parameters, with the exception of an intensity threshold of 6.0×10^8, were used to peak-pick ^{1}H–^{15}N projections of the uniform and reconstructed HNCA spectra (*cf.* Table 2).

3.6. Analysis of sampling distributions

Expectation sampling distributions were also generated from the set of Poisson-gap schedules by averaging their resulting impulse sets. Supplementary Fig. S-2 shows the expectation sampling distributions for one-dimensional schedules having different sampling densities, and Supplementary Figs. S-3 and S-4 show the distributions for two-dimensional schedules having the same densities. The heavy bias towards early time points in Poisson-gap sampling is reaffirmed in all figures. Sampling distributions were also computed *via* Eq. (9) for comparison to the distributions obtained by averaging multiple impulse sets (Figs. S-2 and S-3). To verify that fully random sampling from Eq. (9) and gap sampling from g_{PG} are not equivalent, 10,000,000 sampling schedules were generated by rejection sampling 51 grid points from Eq. (9) at $\Lambda = 62.9$ and $N = 1024$, and histograms of the gap lengths at each grid point were computed (Supplementary Fig. S-8). If the two methods were indeed equivalent, one would expect the histograms in Fig. S-8A to resemble Poisson distributions (S-8B).

Table 1
Summary of peak-picking performance figures produced from comparing IST-reconstructed subsampled 2D ^{1}H–^{15}N HSQC spectra of ubiquitin with their true original uniformly sampled spectrum.

Method		Matched	Lost	Gained	r_{int}	d_H (ppm)	d_N (ppm)
PG	30%	99/99	0/99	2	0.9994	0.000724	0.004459
	10%	99/99	0/99	4	0.9983	0.001208	0.008316
	5%	98/99	1/99	8	0.9920	0.001430	0.009398
SG	30%	99/99	0/99	0	0.9996	0.000580	0.005957
	10%	98/99	1/99	6	0.9983	0.001546	0.007809
	5%	98/99	1/99	7	0.9939	0.001660	0.011393
SB	30%	99/99	0/99	1	0.9996	0.000534	0.008977
	10%	98/99	1/99	5	0.9981	0.001071	0.010007
	5%	98/99	1/99	7	0.9699	0.001482	0.013357

Table 2
Summary of peak-picking performance figures produced from comparing IST-reconstructed subsampled 2D HNCA ^{1}H–^{15}N spectral projections of ubiquitin with their true original uniformly sampled spectral projection.

Method		Matched	Lost	Gained	r_{int}	d_H (ppm)	d_N (ppm)
PG	30%	73/74	1/74	0	0.9978	0.000532	0.007556
	10%	70/74	4/74	0	0.9905	0.001176	0.015378
	5%	66/74	8/74	0	0.9745	0.001488	0.015092
SG	30%	73/74	1/74	0	0.9955	0.000585	0.010554
	10%	66/74	8/74	1	0.9864	0.001793	0.016878
	5%	64/74	10/74	0	0.9638	0.001903	0.020252
SB	30%	73/74	1/74	0	0.9977	0.000560	0.010475
	10%	69/74	5/74	0	0.9883	0.001311	0.015739
	5%	66/74	8/74	1	0.9781	0.001852	0.017306

4. Results

While at first glance, the deterministic schedules constructed using g_{SG} in Figs. 1 and S-5 may appear unrelated to the Poisson-gap schedules, they are in fact realizations of Poisson-gap sampling in which all random draws from the underlying Poisson distribution have resulted in the expected value. This fact is corroborated by the corresponding point-spread functions, which closely resemble those of the stochastic example at 30% and 10% sampling density. Reconstruction residuals from IST (Fig. 2) also reveal a high similarity between the deterministic sine-gap and stochastic Poisson-gap schedules at 30% and 10% sampling density. However, the sine-gap PSF becomes less comparable to that of Poisson-gap at low sampling densities, where the benefits of incoherent sampling are more apparent. It is worth noting that the striking appearance of sampling artifacts in the sine-gap PSF is a consequence of the log-scaled color gradient used in Fig. 1, which was necessary in order to visually expose very low-intensity artifacts.

The addition of burst augmentation in the form of g_{SB} does not substantially alter IST reconstruction residuals relative to g_{SG} and g_{PG}. However, artifacts arising from regularity in g_{SG}-based schedules at low sampling densities are diminished by burst augmentation, resulting in point-spread functions that more closely resemble those from stochastic Poisson-gap sampling. This reduction of artifacts by burst augmentation comes at a small cost, as low-frequency spurs are introduced into the sine-burst point spread function (Supplementary Fig. S-9) by modulating the gap equation. However, these spurs are low in magnitude and only readily apparent at very low (5%) sampling density. These spurs could potentially be reduced by burst-modulating each dimension in the schedule by a different factor.

IST residuals of sine-burst schedules (Fig. 2, dashed lines) are slightly greater than those of one-dimensional sine-gap schedules and dense two-dimensional sine-gap schedules, but they improve relative to sine-gap as sampling density is decreased. Therefore, while sine-gap sampling is a valuable tool for understanding the nature of Poisson-gap sampling, it is clearly bested in performance by multidimensional sine-burst sampling as global sampling density is decreased. Burst augmentation re-introduces sampling incoherence into highly coherent sine-gap schedules to produce sine-burst schedules that more closely resemble Poisson-gap sampling schedules (Supplementary Fig. S-5). This added incoherence is clearly evident in the ^{1}H–^{13}C projections of reconstructed HNCA spectra (Fig. 4), where the more incoherent sine-burst schedule yields a more faithful spectral reconstructions than sine-gap schedule can.

5. Discussion and conclusions

We have shown that Poisson-gap sampling is a single instance in a class of gap sampling methods that may or may not be defined stochastically. Using a well-defined gap sampling algorithm, we

have described two new deterministic sampling methods: sine-gap and sine-burst sampling, which do not rely on random deviates and have comparable performance to stochastic Poisson-gap sampling according to IST reconstruction residuals. From a practical perspective, Poisson-gap, sine-gap and sine-burst sampling methods produced nearly equivalent HSQC spectra (Figs. 3 and S-7) that yielded essentially identical information (chemical shifts, peak intensities) as highlighted in Table 1. Poisson-gap and sine-burst sampling also produced nearly equivalent HNCA spectra (Fig. 4) after IST reconstruction, even at low sampling density. Table 2 also summarizes the peak-picking statistics collected on $^{1}H–^{15}N$ projections of the reconstructed HNCA spectra. For the practicing spectroscopist, this equates to the ability to nonuniformly sample at the performance level of Poisson-gap, without specifying a pseudorandom seed. We find gap sampling to be a flexible and attractive alternative to traditional probabilistic sampling methods that use probability densities to define the local sampling density over a Nyquist grid. In effect, gap sampling approaches the problem of local sampling density from the opposite direction of probabilistic sampling by defining the distances *between* samples on the grid. We have also derived the mathematical connection between stochastic gap equations and their expectation sampling distributions as a means of directly visualizing the grid-point weighting produced by a given gap equation. While these expectation sampling distributions are useful in describing the average sampling behavior of a stochastic gap equation, they do not provide a means of converting a gap-based sampling method into a fully random sampling method. In other words, we have shown that any method of constrained random sampling using a gap equation is inequivalent to fully random sampling from its corresponding expectation sampling distribution.

Finally, burst augmentation provides a concrete example of how deterministic gap sampling may be tuned to behave in a similar fashion to pseudorandom numbers. At first glance, the third rows of Figs. 1 and S-5 would appear to have been generated stochastically, but they are a consequence of the squared-sine modulation term in g_{SB}. It has historically been true that stochastically generated sampling schedules produced fewer prominent artifacts than deterministic methods such as radial or spiral sampling, due to high regularity (*i.e.* coherence) of the latter schemes. However, burst augmentation demonstrates that pseudorandom variates are not strictly required for producing incoherent sampling methods. Furthermore, while most pseudorandom number generators are indeed deterministic for a given seed value, this determinism is *inherently different* from the determinism offered by sine-gap and sine-burst sampling. By design, any parameter (*e.g.*, reconstruction residuals) measured from pseudorandomly generated sampling schedules will not be smoothly varying – and therefore optimizable – functions of their random seed value. As a consequence, no absolute guarantee of spectral quality is provided to the spectroscopist employing pseudorandom sampling schedules, even if the relative difference in quality between the best- and worst-performing Poisson-gap seed values is small at sampling densities above 30% (Fig. 2). This problem with seeds has already been recognized: Poisson-gap and jittered sampling methods are, in fact, two separate attempts at minimizing – but not removing – the effect of seed values on schedule performance [9,10,13]. Deterministic gap sampling completely frees the user from specifying an arbitrary seed value, and provides a highly general framework that enables further investigation into which features of NUS schedules yield higher-quality reconstruction results.

Our C implementations of Poisson-gap, sine-gap and sine-burst sampling are free and open source software, and are available for download at http://bionmr.unl.edu/dgs.php. The programs are highly portable and C99-compliant, so they may be compiled on any modern operating system. An online schedule generation tool is also provided at the same address for rapid generation of one-, two- and three-dimensional NUS schedules suitable for direct use on Bruker or Agilent spectrometers. As defined and implemented, our recursive schedule generation algorithm is not limited to any number of grid dimensions. However, we have limited the online tool to three-dimensional grids to minimize server load.

Acknowledgments

This manuscript was supported in part by funds from grant R01 CA163649 from the National Institute of Health. The research was performed in facilities renovated with support from the National Institutes of Health (RR015468-01).

Appendix A. Supplementary material

This material is available free of charge via the Internet at http://www.sciencedirect.com. Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmr.2015.09.016.

References

[1] M. Mobli, J.C. Hoch, Nonuniform sampling and non-Fourier signal processing methods in multidimensional NMR, Prog. Nucl. Magn. Reson. Spectrosc. 83C (2014) 21–41.
[2] S.G. Hyberts, S.A. Robson, G. Wagner, Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra, J. Biomol. NMR 55 (2013) 167–178.
[3] M.R. Palmer, B.R. Wenrich, P. Stahlfeld, D. Rovnyak, Performance tuning non-uniform sampling for sensitivity enhancement of signal-limited biological NMR, J. Biomol. NMR 58 (2014) 303–314.
[4] D. Rovnyak, D.P. Frueh, M. Sastry, Z.Y.J. Sun, A.S. Stern, J.C. Hoch, G. Wagner, Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction, J. Magn. Reson. 170 (2004) 15–21.
[5] D. Rovnyak, J.C. Hoch, A.S. Stern, G. Wagner, Resolution and sensitivity of high field nuclear magnetic resonance spectroscopy, J. Biomol. NMR 30 (2004) 1–10.
[6] K. Kazimierczuk, V.Y. Orekhov, Accelerated NMR spectroscopy by using compressed sensing, Angew. Chem., Int. Ed. 50 (2011) 5556–5559.
[7] J.C. Hoch, M.W. Maciejewski, B. Filipovic, Randomization improves sparse sampling in multidimensional NMR, J. Magn. Reson. 193 (2008) 317–320.
[8] A.D. Schuyler, M.W. Maciejewski, H. Arthanari, J.C. Hoch, Knowledge-based nonuniform sampling in multidimensional NMR, J. Biomol. NMR 50 (2011) 247–262.
[9] S.G. Hyberts, K. Takeuchi, G. Wagner, Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data, J. Am. Chem. Soc. 132 (2010) 2145–2147.
[10] M. Mobli, Reducing seed dependent variability of non-uniformly sampled multidimensional NMR data, J. Magn. Reson. 256 (2015) 60–69.
[11] P.C. Aoto, R.B. Fenwick, G.J.A. Kroon, P.E. Wright, Accurate scoring of non-uniform sampling schemes for quantitative NMR, J. Magn. Reson. 246 (2014) 31–35.
[12] M.T. Eddy, D. Ruben, R.G. Griffin, J. Herzfeld, Deterministic schedules for robust and reproducible non-uniform sampling in multidimensional NMR, J. Magn. Reson. 214 (2012) 296–301.
[13] K. Kazimierczuk, A. Zawadzka, W. Kozminski, I. Zhukov, Lineshapes and artifacts in multidimensional Fourier transform of arbitrary sampled NMR data sets, J. Magn. Reson. 188 (2007) 344–356.
[14] S.G. Hyberts, H. Arthanari, S.A. Robson, G. Wagner, Perspectives in magnetic resonance: NMR in the post-FFT era, J. Magn. Reson. 241 (2014) 60–73.
[15] S.G. Hyberts, H. Arthanari, G. Wagner, Applications of non-uniform sampling and processing, Top. Curr. Chem. 316 (2012) 125–148.
[16] S.G. Hyberts, A.G. Milbradt, A.B. Wagner, H. Arthanari, G. Wagner, Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson gap scheduling, J. Biomol. NMR 52 (2012) 315–327.
[17] M.W. Maciejewski, M. Mobli, A.D. Schuyler, A.S. Stern, J.C. Hoch, Data sampling in multidimensional NMR: fundamentals and strategies, Top. Curr. Chem. 316 (2012) 49–77.
[18] M.W. Maciejewski, H.Z. Qui, I. Rujan, M. Mobli, J.C. Hoch, Nonuniform sampling and spectral aliasing, J. Magn. Reson. 199 (2009) 88–93.
[19] G.L. Bretthorst, Nonuniform sampling: bandwidth and aliasing, Concepts Magn. Reson. A 32A (2008) 417–435.
[20] L.E. Kay, P. Keifer, T. Saarinen, Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity, J. Am. Chem. Soc. 114 (1992) 10663–10665.

[21] A.G. Palmer, J. Cavanagh, P.E. Wright, M. Rance, Sensitivity improvement in proton-detected 2-dimensional heteronuclear correlation Nmr-spectroscopy, J. Magn. Reson. 93 (1991) 151–170.
[22] A.D. Schuyler, M.W. Maciejewski, A.S. Stern, J.C. Hoch, Formalism for hypercomplex multidimensional NMR employing partial-component subsampling, J. Magn. Reson. 227 (2013) 20–24.
[23] A.S. Stern, D.L. Donoho, J.C. Hoch, NMR data processing using iterative thresholding and minimum l(1)-norm reconstruction, J. Magn. Reson. 188 (2007) 295–300.
[24] S.J. Sun, M. Gill, Y.F. Li, M. Huang, R.A. Byrd, Efficient and generalized processing of multidimensional NUS NMR data: the NESTA algorithm and comparison of regularization terms, J. Biomol. NMR 62 (2015) 105–117.
[25] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe – a multidimensional spectral processing system based on Unix pipes, J. Biomol. NMR 6 (1995) 277–293.
[26] J.C. Hoch, M.W. Maciejewski, M. Mobli, A.D. Schuyler, A.S. Stern, Nonuniform sampling and maximum entropy reconstruction in multidimensional NMR, Acc. Chem. Res. 47 (2014) 708–717.