

Enhanced spectral resolution by high-dimensional NMR using the filter diagonalization method and “hidden” dimensions

Xi Meng^a, Bao D. Nguyen^b, Clark Ridge^a, A.J. Shaka^{a,*}

^aChemistry Department, University of California, Irvine, Room 4134, Natural Science One, ZOT Code 2025, Irvine, CA 92697-2025, USA

^bPfizer Inc., 445 Eastern Point Road, MS 8118A-2011, Groton, CT 06340, USA

ARTICLE INFO

Article history:

Received 25 July 2008

Revised 20 September 2008

Available online 30 September 2008

Keywords:

Multidimensional NMR

Filter diagonalization

FDM

Resolution enhancement

High-dimensional

Spectra

Hidden dimensions

Protein NMR assignment

ABSTRACT

High-dimensional (HD) NMR spectra have poorer digital resolution than low-dimensional (LD) spectra, for a fixed amount of experiment time. This has led to “reduced-dimensionality” strategies, in which several LD projections of the HD NMR spectrum are acquired, each with higher digital resolution; an approximate HD spectrum is then inferred by some means. We propose a strategy that moves in the opposite direction, by adding *more* time dimensions to increase the *information content* of the data set, even if only a very sparse time grid is used in each dimension. The full HD time-domain data can be analyzed by the filter diagonalization method (FDM), yielding very narrow resonances along *all* of the frequency axes, even those with sparse sampling. Integrating over the added dimensions of HD FDM NMR spectra constitutes LD spectra with enhanced resolution, often more quickly than direct acquisition of the LD spectrum with a larger number of grid points in each of the fewer dimensions. If the extra-dimensions do not appear in the final spectrum, and are used solely to boost information content, we propose the moniker *hidden-dimension NMR*. This work shows that HD peaks have unmistakable frequency signatures that can be detected as single HD objects by an appropriate algorithm, even though their patterns would be tricky for a human operator to visualize or recognize, and even if digital resolution in an HD FT spectrum is very coarse compared with natural line widths.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

The resolving power of high-dimensional (HD) NMR spectra has been instrumental in solution protein structure elucidation [1–4], now allowing large, interesting proteins and complexes to be studied [5–8]. High-dimensional (HD) NMR data sets have time-domain data with 3–5 dimensions and 2–4 interferometric or “indirect” time variables, t_k . The direct acquisition time, t_{ACQ} , is for data collection, yielding a complex data set $C(t_1, t_2, t_3, \dots, t_{ACQ})$. Data are usually acquired over a regular time grid for each of the t_1, t_2 , etc., independently and then transformed into 3–5D spectra $S(F_1, F_2, F_3, \dots, F_{ACQ})$ in which each dimension has independent digital resolution $SW_1/N_1, SW_2/N_2, \dots, SW_{ACQ}/N_{ACQ}$. This sampling strategy suffers from a multiplicative time penalty arising from the need for adequate digital resolution in *each* spectral dimension *independently*. The total number of indirect points, N_{TOT} , and hence the minimum experiment time T_{MIN} (assuming adequate sensitivity) increases rapidly as the number of indirect dimensions is increased:

$$T_{MIN} \propto N_{TOT} \propto N_1 N_2 N_3, \dots \quad (1)$$

* Corresponding author. Fax: +1 949 824 9920.

E-mail address: ajshaka@uci.edu (A.J. Shaka).

A 4D experiment of size $64 \times 64 \times 64 \times N_{ACQ}$, with a minimum two-step phase cycle and acquisition of both N- and P-type data sets for each of the three indirect dimensions, would require $2^{21} \sim 2.1$ million transients, taking 24.3 days at 1 Hz repetition rate. Even 64 points may be marginal at high magnetic field, where the spectral widths for both $^{13}C/^{15}N$ expand to quite a few kHz, and $^{13}C/^{15}N$ line widths may be surprisingly narrow when optimized pulse sequences exploiting the TROSY effect [9–11] are used.

Strategies to speed up data collection have been proposed. Most involve reducing the dimensionality of the spectrum by taking various cross-sections or subsets of time-domain points [12–16], often corresponding to integral projections in the frequency spectrum. In FT spectra, extending acquisition in a k th orthogonal dimension cannot possibly improve the resolution of the integral projection in which this same dimension is integrated over. The integral projection along a frequency axis yields essentially the first point of the associated time-domain interferogram; the value of the first point does not depend on how many subsequent time points are acquired. This independence of the resolution of projections in FT spectra is shown in Fig. 1.

While the *digital resolution* in any orthogonal frequency dimension is not altered in an FT spectrum, there is clearly far more information in the higher-dimensional spectrum *provided it can be analyzed in the higher-dimensional space*: degeneracies are resolved, the number of resonances can be tallied more completely, and the

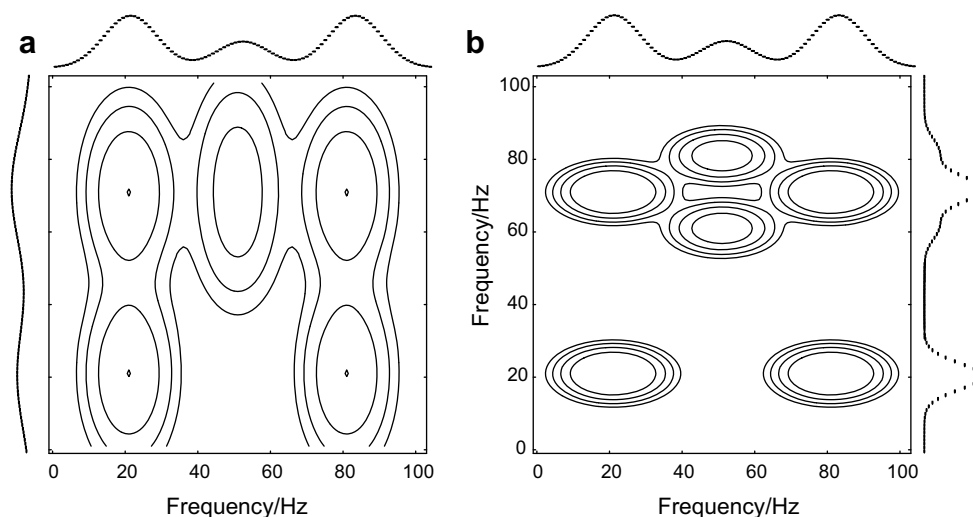


Fig. 1. (a) A noise-free 2D spectrum at low resolution, showing the integral projections at the margins. Neither of the orthogonal projections allows a full count of the number of peaks in the 2D plane. (b) The same spectrum as in (a) but with superior vertical resolution, allowing two peaks to be resolved. The horizontal spectrum obtained from the vertical integral projection remains identical to the previous case, because in FT spectra the two frequency dimensions are independent.

line shapes of the various peaks can yield a wealth of useful information. In short, the resolution of the HD spectrum with n dimensions is best assessed in n dimensions. It is mostly an artifact of display technology and limitations of the human mind that necessitates contouring planes or plotting traces to inspect the spectrum. The extra-dimensions unravel overlap, by separating and visualizing each resonance. Integrating over any dimension is clearly counter-productive to this aim.

With certain assumptions about the nature of the data, it may be possible to do far better than FT methods, and materially change the resolution in one orthogonal dimension— or *all* the other dimensions. For example, if one were to assume that peaks of interest were *already resolved* in a 2D spectrum derived from a projection of a 3D spectrum, then the additional frequency information in the 3rd dimension for each resolved peak could be retrieved with very few additional points in the 3rd time dimension, using parametric fitting of the data to a model with a single peak at unknown frequency [17]. Likewise, with aggressive non-linear methods, a small number of n D projections may allow more accurate inference of peak positions in $(n + 1)$ D spectrum than is possible by straightforward Fourier transformation. The utility of these more aggressive methods may also depend on whether simply peak position and total number of peaks, or detailed line shape, intensity and line widths are of interest. For some assignment tasks the exact intensity of each peak is relatively unimportant compared to the exact frequency coordinates, whereas in NOE experiments the peak intensities clearly matter.

Rather different altogether is the case in which the lower-dimensional data is poorly resolved, is somewhat noisy, or in which it is not clear whether or not there is overlap. In this case, aggressive processing may result in a possibly misleading HD spectrum. This is shown schematically in Fig. 2, where the simple noise-free 2D spectrum of Fig. 1 is reconstituted using the two coordinate projections, by taking the minimum ordinate at each position, the lower-value algorithm [14]. Obviously this result could be improved with more 1D projections, and this example is for illustration only, and certainly not as an example of the best practice of the method. However, even many such projections fail to reconstruct the true 2D spectrum. With more complex experimental data containing significant noise, genuine peaks may be lost, or artifacts introduced, limiting the utility of the rapid acquisition

for the task at hand, which is to gain rapid and accurate information about the sample under study.

2. Results

2.1. Theory

2.1.1. A simple case of one versus two dimensions

Suppose, with acquisition time T_k , that two peaks are unresolved in some dimension k of a conventional multidimensional FT spectrum. If the full width at half maximum, Dw , of each peak is much less than $1/T_k$, then their frequency spacing Df must be somewhat less than $1/T_k$, i.e., the limited acquisition is limiting the resolution. What is the chance of resolving these two peaks by increasing T_k ? Clearly, if the two peaks happen to be nearly degenerate ($\Delta f \ll \Delta w$) then increasing T_k is fruitless. However, if these same two peaks are displayed as a function of *two* dimensions, and the line positions in the second dimension are not tightly correlated with those in the first, then acquiring data in this new dimension, even over the entire 2D time grid, might be more efficient than acquiring an ever longer signal in the first dimension alone. This is shown schematically in Fig. 3, where two peaks are separated by 2.0 Hz in the horizontal dimension. With a spectral width of 100 Hz, about 64 points are required to resolve them using a 1D data set. However, if they were separated by 50 Hz in a second dimension, just a 3×3 time grid suffices to indicate clearly the presence of at least two peaks using a 100 Hz spectral width in each dimension, even by conventional FT analysis (with extensive zero-filling). Taking into account the necessity of using both N/P (or sin/cos) data sets to achieve absorption line shapes, a total of 18 points is the minimum for the 2D realization. This compares very favorably to 64, reducing the number of required points by a factor of almost four. If these data were to be *fit* to two peaks, rather than simply *transformed* into a spectrum, both could be extracted quite accurately, even in the presence of some noise, for the 2D case. In the 1D case, noise would necessitate more than 64 points to get anything like the same reliability for frequency, line width, and amplitude. The displacement of the peaks in the additional dimension facilitates their free and clear characterization, just as the reduced E. COSY

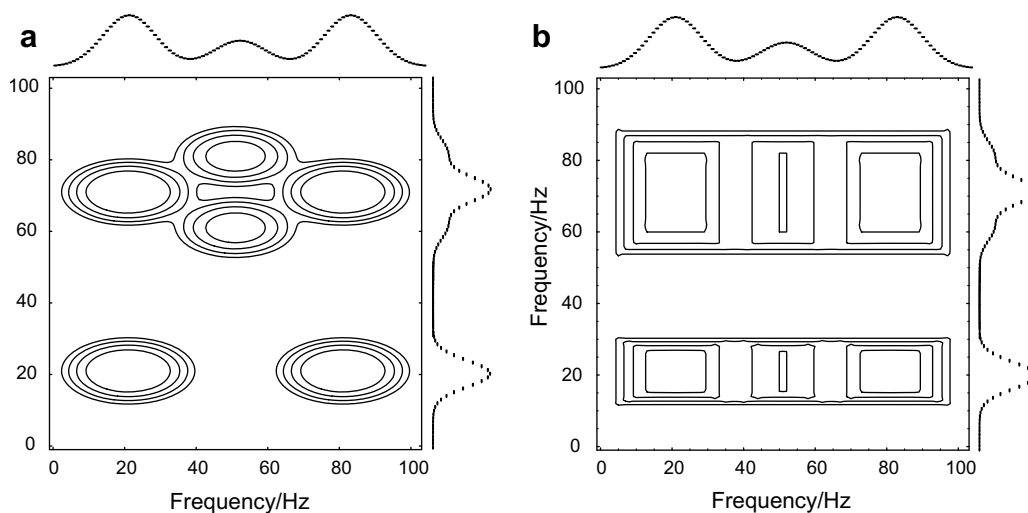


Fig. 2. (a) Integral projections obtained from a 2D spectrum, the high-resolution case of Fig. 1. (b) A 2D spectrum obtained from the two integral projections, by the “lower bound” algorithm, in which the 2D spectral grid points are set according to the formula $S(n, m) = \text{Min}\{I_h(m), I_v(n)\}$, where I_h is the horizontal projection and I_v the vertical projection. Clearly, two projections are insufficient to reconstruct the true 2D spectrum in this example.

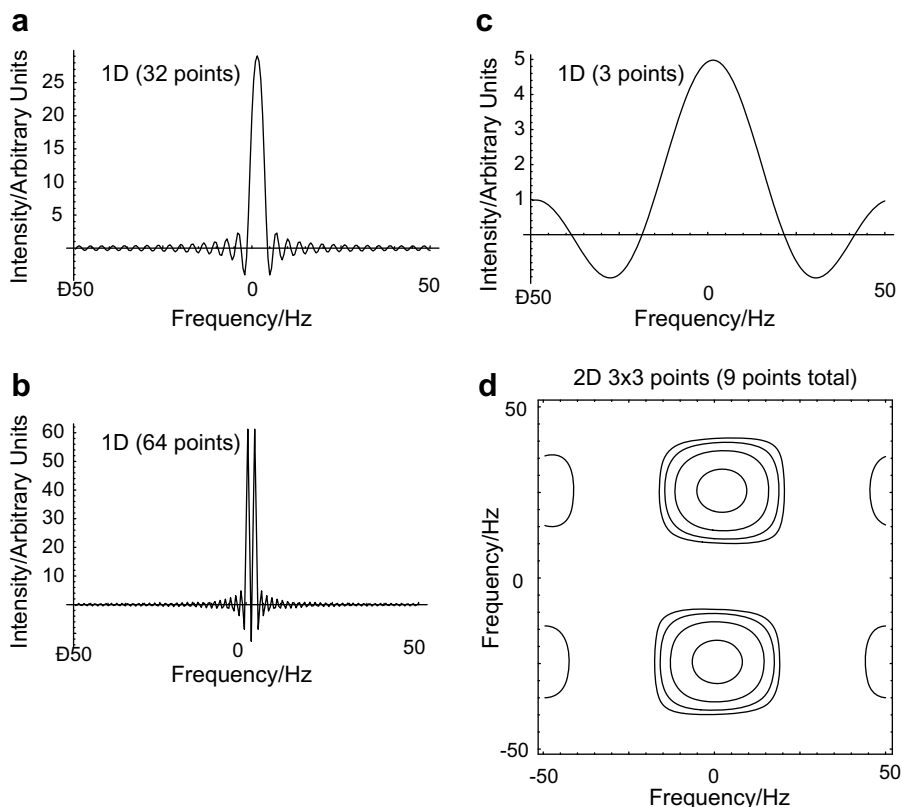


Fig. 3. A comparison of the amount of data needed to resolve two peaks that are very close in one frequency dimension and quite well-separated in a second orthogonal frequency dimension. (a) A 1D projection using 32 time-domain points, zero-filled extensively before FT. The two peaks are unresolved. (b) The same as (a) but with 64 data points. The two peaks are resolved. (c) The same as (a) but with only *three* data points (before zero-filling). The two peaks are *completely* unresolved. (d) A 3×3 2D FT spectrum, extensively zero-filled in both dimensions. Because the peaks are widely separated in the second dimension, they are resolved with only 9 data points total (18 points to obtain an absorption-mode spectrum), even with FT analysis only. The large sinc-wiggles explain the features at the horizontal edges. Using parametric methods that try to identify actual peaks, or otherwise avoid the transform-limited line widths, the well-separated 2D peaks are much easier to identify accurately than the overlapping 1D peaks, especially in the presence of noise.

multiplet patterns allow the measurement of J -couplings that are far too small to observe directly [18]. This example serves to illustrate an important point that bears emphasizing: HD NMR spectra have higher intrinsic resolution than their lower-dimensional pieces. Once the idea of independent 1D transformations

is abandoned, the HD paradigm becomes ever more compelling, as very short data sets can exploit the uncorrelated line positions in, e.g., protein backbone assignment experiments even though none of the constituent 1D projections show any resolved features at all.

2.1.2. The one-dimensional filter diagonalization method

The filter diagonalization method (FDM) is a data analysis technique that has been in the literature for a number of years, and employed for 1D [19], 2D [20–22], 3D [23] and 4D [24,25] NMR experiments of various kinds. The systematic development of the equations, numerical aspects, and performance issues of FDM have been reviewed [26]. Our brief, qualitative treatment here is to allow issues arising later on to be addressed clearly and to make the surprising results understandable.

In the one-dimensional case, FDM efficiently and reliably solves the problem of fitting a complex time signal of N points, sampled on a uniform time grid, in terms of a linear combination of M complex sinusoids:

$$\begin{aligned} c_n &= C(n\tau) = \sum_k^M d_k \exp(i\{2\pi f_k + i\lambda_k\}n\tau) \\ &\equiv \sum_k^M d_k \exp(i\tilde{\omega}_k n\tau), \quad n = 0, \dots, N-1 \end{aligned} \quad (2)$$

with complex amplitudes d_k (integral and phase) and complex frequencies $\tilde{\omega}_k$ (position and width) over a spectral width $SW = \tau^{-1}$. The line shape from Eq. (2) is Lorentzian; the discrete infinite Fourier transform can be computed analytically [26] by introducing $u_k = \exp(i\tilde{\omega}_k \tau)$ and $z = \exp(2\pi i f \tau)$ and summing a geometric series:

$$\begin{aligned} S(f) &= \tau \left\{ \frac{-c_0}{2} + \sum_{n=0}^{\infty} \sum_k^M d_k (u_k/z)^n \right\} \\ &= \tau \left\{ \frac{-c_0}{2} + \sum_k^M d_k \sum_{n=0}^{\infty} (u_k/z)^n \right\} = \tau \sum_k^M d_k \left\{ \frac{1}{1 - (u_k/z)} - \frac{1}{2} \right\} \end{aligned} \quad (3)$$

The first point correction and factor of t ensure the sum coverages to the continuous Fourier transform in the limit $\tau \rightarrow 0$. The number of peaks comprising the fit is $M = N/2$, half the number of frequency grid points that would occur in a non-zero-filled FT spectrum. The amplitudes and frequencies are obtained by an eigenvalue problem whose input matrix elements depend only on the discrete measured data points c_n . We assume that some initial state Φ_0 evolves in time and that its autocorrelation function generates the signal:

$$c_n = (\Phi_0 | \hat{U}^n \Phi_0) \quad (4)$$

where a complex symmetric inner product,

$$(\Phi | \Psi) = (\Psi | \Phi) \quad (5)$$

allows for arbitrary signal phase. Diagonalizing \hat{U} is then equivalent to determining both d_k and u_k for each peak:

$$\begin{aligned} \hat{U} &= \sum_k^M |Y_k\rangle u_k \langle Y_k| \& (Y_k | Y_j) = \delta_{kj} \Rightarrow c_n = \left(\Phi_0 | \cdot \left(\sum_k^M |\mathcal{Y}_k\rangle u_k \langle \mathcal{Y}_k| \right)^n | \Phi_0 \right) \\ &= \sum_k^M (\Phi_0 | \mathcal{Y}_k) (\mathcal{Y}_k | \Phi_0) u_k^n \end{aligned} \quad (6)$$

$$= \sum_k^M (\Phi_0 | \mathcal{Y}_k)^2 u_k^n \equiv \sum_k^M d_k u_k^n, \quad (7)$$

and \hat{U} itself, by translating the initial state in discrete time steps, can generate other (non-orthogonal, unnormalized) states that can be used to make up a basis $\{\Phi_n\}$:

$$\Phi_0; \quad \Phi_1 = \hat{U}\Phi_0; \quad \Phi_2 = \hat{U}\Phi_1 = \hat{U}^2\Phi_0, \text{ etc.} \quad (8)$$

Using the evolution operator itself to construct the basis is attractive because it should yield exactly the states that are needed to describe the dynamics. Simple representations of the unit matrix, $\hat{\mathbf{1}}$ or $\mathbf{U}^{(0)}$ and \hat{U} or $\mathbf{U}^{(1)}$ are obtained from the measured data [26] using this primitive basis $\{\Phi_n\}$:

$$\mathbf{U}_{mn}^{(0)} = (\Phi_m | \Phi_n) = (\hat{U}^m \Phi_0 | \hat{U}^n \Phi_0) = (\Phi_0 | \hat{U}^{n+m} \Phi_0) = c_{n+m} \quad (9)$$

$$\mathbf{U}^{(0)} = \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{M-1} \\ c_1 & c_2 & & & \\ c_2 & & & & \\ \cdots & & & & \\ c_{M-1} & \cdots & & & c_{2M-2} \end{bmatrix} \quad (10)$$

$$\mathbf{U}_{mn}^{(1)} = (\Phi_m | \hat{U} \Phi_n) = (\hat{U}^m \Phi_0 | \hat{U}^{n+1} \Phi_0) = (\Phi_0 | \hat{U}^{n+m+1} \Phi_0) = c_{n+m+1} \quad (11)$$

$$\mathbf{U}^{(1)} = \begin{bmatrix} c_1 & c_2 & c_3 & \cdots & c_M \\ c_2 & c_3 & & & \\ c_3 & & & & \\ \cdots & & & & \\ c_M & & & & c_{2M-1} \end{bmatrix} \quad (12)$$

These data matrices also arise in the formulation of parametric Linear Prediction (LP) [27], which is closely related to FDM. The eigenvalue problem

$$\hat{U}|\mathcal{Y}_k\rangle = u_k \hat{\mathbf{1}}|\mathcal{Y}_k\rangle \quad (13)$$

is then written

$$\mathbf{U}^{(1)} \mathbf{B}_k = u_k \mathbf{U}^{(0)} \mathbf{B}_k \quad (14)$$

It is essential to include the matrix representation for the identity operator because the basis $\{\Phi_n\}$ is not orthonormal. With the eigenvalues and eigenvectors in hand, the spectrum is then computed by Eqs. (3) and (7). Equation (14) is a generalized eigenvalue problem that has a number of library routines available to solve it. Matrix diagonalization is one of the best-studied algorithms and has been optimized extensively, making this approach efficient compared to any attempt to directly fit the FID as damped sinusoids using a non-linear least squares routine. Nevertheless, for realistic data sets, the \mathbf{U} matrices are (a) dense; (b) huge; and (c) possibly ill-conditioned (Hankel structure). FDM overcomes some of these problems by making linear combinations of the M time-like basis functions Φ_k to form frequency-like basis functions $\Psi(f_k)$ at M equally-spaced frequency grid points across the spectral width, $f_k = k(SW/M) - SW/2$, $k = 0, \dots, M-1$:

$$\Psi(f_k) = \sum_{n=0}^{M-1} \exp(-2\pi i f_k n \tau) \Phi_n = \sum_{n=0}^{M-1} \exp(-2\pi i f_k n \tau) \hat{U}^n \Phi_0 \quad (15)$$

When we evaluate matrix elements of the \mathbf{U} operators between basis functions $\Psi(f_k)$ and $\Psi(f_j)$ at widely different frequencies, the values are small compared to those near the diagonal—much smaller than some intermediate point c_j in the FID, which would be the corresponding value between the Φ basis functions. Indeed, the Hankel structure of Eqs. (9) and (11) ensure that very large matrix elements can occur far from the diagonal, as values along the anti-diagonal direction are identical. The Fourier basis thus disentangles the structure of the matrices to make them easier to diagonalize, in the same way the FT separates out peaks in the spectrum, making them easier to recognize than the interfering sinusoids and noise comprising the raw FID. Small off-diagonal matrix elements, which often refer mostly to noise, can be neglected according to time-independent perturbation theory when there is a large frequency difference between the states involved. This allows the giant matrix problem to yield to a divide-and-conquer strategy in which smaller submatrices, near the diagonal, are diagonalized sequentially. Each submatrix corresponds to a small frequency window. The spectra are calculated over each window by Eq. (3) or a number of other similar formulas [26] and then smoothly summed up. In our implementation of multi-window FDM we overlap adjacent windows by 50% and use a \cos^2 weighting function to stitch the spectral estimates together into a full spectrum [22], but this is somewhat arbitrary.

trary. The spectra so obtained are usually surprisingly insensitive to the details of the window size. This insensitivity is a prerequisite for reliable results.

This background is important to facilitate understanding the resolving power of FDM. Clearly, there are at most M peaks. If the true spectrum contains more than M distinct peaks then the fit cannot possibly be perfect. However, when there is no noise, no degenerate frequencies in any of the dimension, and the data conforms exactly to the model, then the fit obtained by FDM is perfect whenever the true number of peaks K is less than M . There is a sudden “phase transition” in which the spectrum crystallizes into clear view and then remains constant as more data is acquired. When the calculation is over one window, the number of local basis functions in the window should be greater than the number of signal peaks in the same window. These simple arguments get more complex when there are very large peaks outside the window, but there are ways to handle this situation smoothly.

Sadly, in the presence of even small noise amplitudes, the number of peaks is always essentially M or more. Noise does not repeat, does not follow any pattern, and cannot be cast as a certain fixed number of Lorentzian peaks, so that real data must appear as some number of “peaks” that is always at least M , frustrating FDM convergence somewhat. What effect will this have on the attainable resolution? The short answer is that if the number of *signal peaks* K is less than M , and the noise amplitude is much smaller than any of these peaks, then both the fit and the resolution are excellent. As noise increases, it may tend to dominate, forcing precious potential peaks to be “squandered” fitting the local noise instead. There appears to be no ironclad guarantee, in this case, that the output of the generalized eigenvalue problem is the *best* fit, in terms of Lorentzian peaks and, even if it were, that the best fit would result in the most useful spectral estimate. Even with a *perfect fit* the spectral estimate may not be accurate in the information we seek: line positions, intensities, etc. of *signal peaks*, because the perfect fit includes the noise. This makes the quality of the fit itself a potentially distracting criterion. It is well known in the LP literature that the case $M \gg K$ mitigates the influence of the noise. In FDM, this can be seen clearly by visualizing the grid of basis functions in frequency. Qualitatively, a reasonable guideline might be that the density of basis functions should be high enough that there is at least one grid basis function $\Psi(f_k)$ between the frequency centers of any pair of peaks we wish to resolve. The sampling must be of the order of the inverse frequency separation to achieve this density. The same is true in FT spectroscopy, although in the FT case the sampling must be long enough whether or not there is significant noise, and the line shape may still be transform limited, broad, and show sinc-wiggles. In FDM the line shape is always Lorentzian, and the accuracy of the width depends on the noise; when the latter is low, the former can be accurate. Sharp, narrow

resonance lines with good resolution and accurate intensities will result.

It might seem possible to “zero fill” the basis $\Psi(f)$ by adding more frequency coordinates between the existing f_k , using Eq. (14) to calculate additional functions, and thereby increasing the potential to pick up and ultimately discount any noise features. However, the phrase “zero fill” here suggests the problem: as the frequencies become substantially closer than the natural grid f_k , the functions $\Psi(f)$ just become more linearly dependent, making $\mathbf{U}^{(0)}$ closer to singular. The way the generalized eigenvalue problem is numerically handled means that linear dependence is not fatal to the algorithm, but artificially increasing the basis density without increasing the information content of the signal does not yield any notable improvement in resolution, either.

Fig. 4 shows a simple but illustrative 1D benchmark problem in which the object is to resolve the two sharp Lorentzian lines in the presence of varying amounts of noise and with varying numbers of points in the FID. When the noise level is extremely low, just two Fourier basis functions (one at the edge of the spectrum, and one in the center) from only four time-domain data points can still detect the small phase/amplitude shifts that occur when the FID is left-shifted by one point, allowing the lines to be calculated fairly accurately by Eq. (10). This case is analogous to being able to open a weightless door by pushing very near the hinge. Although each basis function has a frequency center, they are fairly delocalized when only four data points comprise the FID. Noise quickly destroys the ability to resolve the two peaks as quickly, as shown in the third panel.

As the noise increases it pays to have a higher density of basis functions, *i.e.*, sample more data and, in particular, having the local basis function density be of the order of the peak spacing can be quite important. (If the aforementioned door is heavy, using the handle is imperative.) The FDM behavior, shown in Fig. 5, is quite different from that of the FT spectrum. The FT spectrum requires a data length of the order of the inverse *width* of the peaks to really deliver all the potential resolution, and this requirement is independent of the noise, or peak spacing. It also necessitates a much longer acquisition time whenever the peaks are narrow.

2.1.3. Issues in multidimensional FDM

Suppose that a two-dimensional signal $C(t_1, t_2)$ is obtained, in which each of the one-dimensional projections shows substantial overlap. What is the possibility of obtaining a fully resolved 2D spectrum using FDM? Following the earlier discussion, the number of 2D *basis functions* must exceed the number of 2D *peaks*. However, the former is proportional to the *product* of the number of recorded data points:

$$M = \frac{N_1 N_2}{2} \quad (16)$$

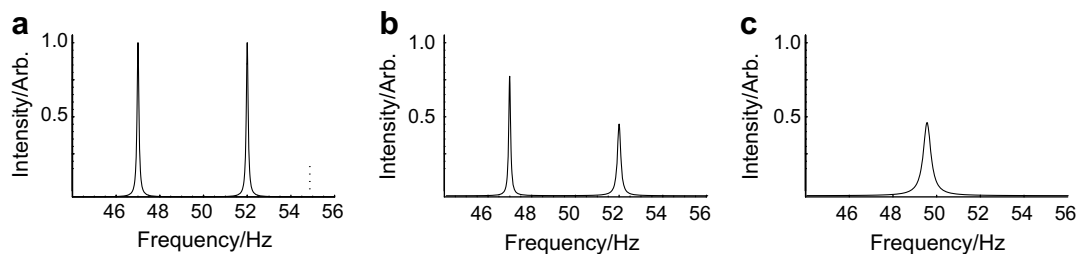


Fig. 4. Deterioration of a high-resolution FDM spectrum as a function of the percentage of noise added. Two narrow Lorentzian lines are located at 47 and 52 Hz, with identical amplitude, phase, and line width. The Nyquist range is 200 Hz and only four points of the synthetic FID were generated. (a) No noise. The two peaks are recovered quantitatively with only four points. (b) 0.1% random time-domain noise. Both lines have appreciable errors in line width, but not phase, frequency or amplitude (integral). (c) 1.0% random time-domain noise. The sparse data is now insufficient to resolve the two lines. A broadened peak near the mean frequency, and with line width much narrower than the original frequency separation, is the result. Only the amplitude of the noise is altered. The particular realization is otherwise identical to that in (b). With other noise realizations, the results may show narrower or broader lines, and slight errors in the integral as well.

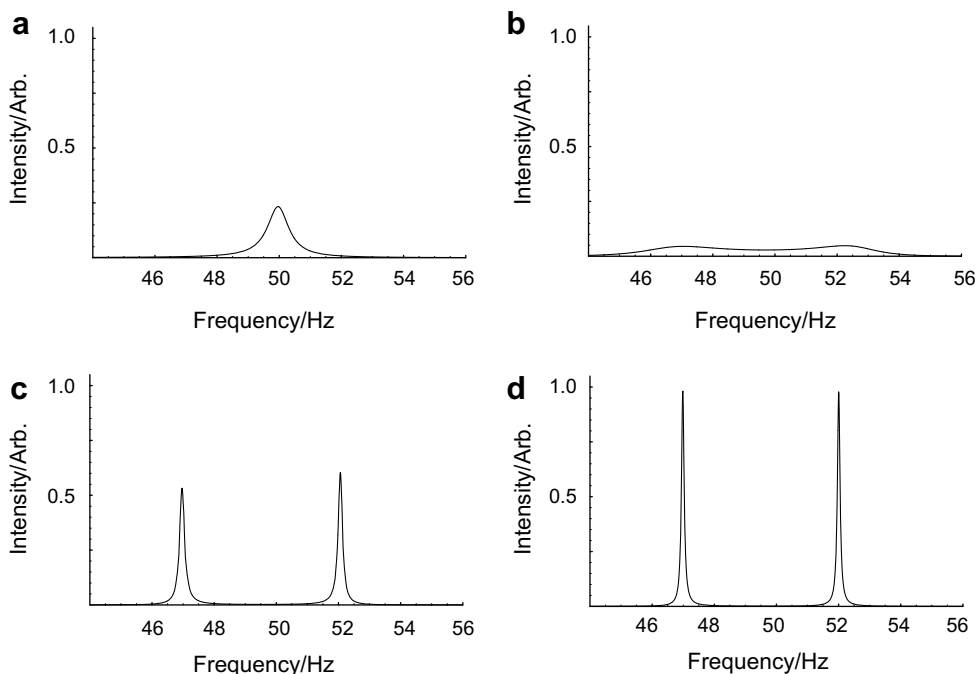


Fig. 5. Behavior of the high-resolution FDM spectrum as a function of the number of data points acquired. Two narrow Lorentzian lines are located at 47 and 52 Hz, with identical amplitude, phase, and line width. The Nyquist range is 200 Hz but only the relevant region is displayed. (a) 8 points. (b) 16 points. (c) 32 points. (d) 64 points. The noise amplitude was about 1% in the time domain. For a peak spacing of 5 Hz in a 200 Hz spectral width, ~ 40 Fourier basis functions would be an estimate of the density needed to resolve the two peaks. This in turn would necessitate ~ 80 time-domain data points. In fact two features emerge with only 8 Fourier basis functions (b), although the result is only convincing with 16 Fourier basis functions (c) and the line widths are only correct when 32 Fourier basis functions can be used (d). This is close to the estimate of 40 based on the argument that the local density be of the order of the frequency separation of the features to be resolved. The actual details depend on the percentage of noise added.

so that the 2D basis set can be very large if just one dimension has good digitization, $N_2 \gg N_1$. In most NMR experiments this can be the direct detection dimension, that is, $N_2 = N_{\text{ACQ}}$. By contrast, Fourier transformation along the direct dimension gives N_2 uncorrelated 1D FDM calculations, each with 1D basis size at most $N_1/2$. If N_1 is not large, then the result may be poor for any F_1 trace that happens to contain more than $N_1/2$ peaks. Peaks are not confined to a single data point in F_2 . A given 2D peak will appear as a 1D peak multiple times, on adjacent F_2 traces along the F_1 dimension. Large peaks may have significant intensity on many points in F_2 , limiting the number of weaker 1D peaks that can be identified accurately. The 1D peaks in F_1 for adjacent F_2 points also may not align precisely. By contrast, a 2D FDM calculation of a 2D peak, wherever it is, appears only *once* in the 2D plane, requiring only *one* 2D basis function to pin it down, and it cannot look like anything except a 2D peak. This, in a nutshell, is why partial Fourier transformation can be counterproductive compared to an up-front multidimensional method, and why it may be possible to increase the dimensionality of the data set and still reduce overall acquisition time. The total n D basis size can become enormous as the number of dimensions increases: an $8 \times 8 \times 8 \times 4096$ 4D data set requires the solution of four generalized eigenvalue problems of the form of Eq. (14), each with $128k \times 128k$ matrices containing ~ 17.2 billion matrix elements. It is only with the advent of FDM that it is possible to entertain problems of this enormous numerical scale.

2.1.4. Block sampling and deliberate aliasing

Non-uniform sampling (NUS) is a strategy to acquire a longer interferogram, using the same number of points, by relaxing the requirement to sample time points on a uniform grid at the Nyquist rate. Instead, some points are effectively skipped in favor of others at greater evolution times. There must be a trade off, of course, as very long evolution times have vanishing signal intensity

and hence little useful information. “Randomly” sampled data cannot be processed by standard FT methods, but can be handled using maximum entropy (MaxEnt) processing [27–29]. But consider a *block sampling* strategy, in which a certain number of points N_2 are sampled uniformly at dwell time t_2 , and then a gap Mt is introduced (usually $M \gg N_2$) a further N_2 points are sampled, and the process repeated N_1 times. A 1D signal $C(n_2 t_2 + n_1 M t_2)$, $n_2 = 0, \dots, N_2 - 1$, $n_1 = 0, \dots, N_1 - 1$, is obtained. We can, however, consider the signal instead as a pseudo-2D signal:

$$C(n_1 \tau_1, n_2 \tau_2) = C(n_1 M \tau_2, n_2 \tau_2) \equiv C(n_2 \tau_2 + n_1 M \tau_2) \quad (17)$$

Fig. 6 shows a comparison between uniform sampling, non-uniform sampling, block sampling, and a 1D block-sampled data set considered as a pseudo-2D data set. Processed as a two-dimensional signal, which is now sampled *uniformly* on a two-dimensional time grid, conventional 2D FDM can be used for analysis; there are no issues with unwanted artifacts, and the smaller spectral width in the new, hidden, dimension means that the basis function density in the new dimension will be higher than it was in the first dimension. This strategy is potentially useful when the peaks cluster into groups, *i.e.*, a number of closely spaced peaks and then areas without peaks. Many kinds of NMR spectra show these features. The first dimension then establishes the true chemical shift values over the true spectral width, and the second dimension boosts the resolution so that all peaks are distinct and resolved. Once the narrow 2D resonances are identified, the second dimension can be integrated over, giving back the original 1D spectrum with the correct frequencies, and higher resolution. This block sampling strategy is thus adapted to the different frequency scales in the spectrum.

Maximum entropy has also been used to create “virtual dimensions” in which the width of lines are plotted in a second dimension, with the line position plotted in the first one, thereby apparently increasing resolution [30]. The difference with the ap-

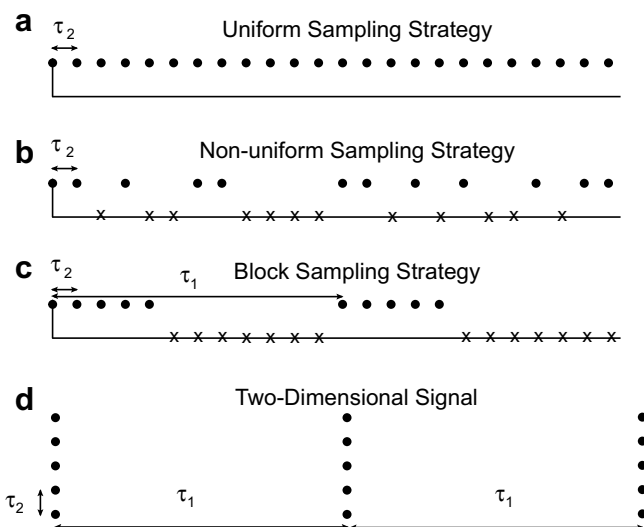


Fig. 6. Various sampling strategies. (a) Uniform sampling at the Nyquist rate. (b) Non-uniform sampling where, for simplicity, the points are still obtained at integer multiples of the dwell time used in (a). (c) A block sampling strategy, in which data are acquired in a repeating pattern in which multiple data points are skipped. (d) Rearrangement of the block-sampled 1D data into a 2D data set in which there are no missing data points but the spectral width in the new dimension is M times smaller than the Nyquist range.

proach here is that the data itself is interpreted as a higher-dimensional signal, rather than simply displaying the data in a higher-dimensional plot in which different parameters are shown in the orthogonal dimensions.

It should be clear that sampling points further out in the data set can markedly improve the resolution, much as the NUS strategy. It is also possible to “fill in” the missing data points. The samples are all on the same time grid, so that if all points are sampled a conventional signal results. This option is not usually exploited in the NUS case, where samples are often obtained with a random sampling schedule weighted by the expected properties of the FID. The fully sampled 1D signal will have the best attainable resolution, as it contains the most information. However, it may be possible to obtain a similar result more rapidly, and check its reliability, by incrementally sampling another set of points that gradually increases the block size N_2 . If the spectral features do not change much when additional points are taken, then they have converged to within the prevailing signal to noise ratio (SNR), and acquisition can be terminated.

The strategy shown in Fig. 6 has a couple of subtleties. First, complementary N- and P-type data are not available in the hidden dimension when the latter results from omitting blocks of data. A conventional FT spectrum would thus show “phase-twist” [34] line shapes in the hidden dimension plane. Using multidimensional FDM, only *one* data set is required: the fit is to a single phase-modulated multidimensional signal. In this case, only the P-type data is available by simply skipping intervening points, but this apparent deficiency is largely irrelevant as long as methods other than FT are used to analyze the data.

We should note that the analysis of incomplete or “gapped” data has a long history in both engineering [32] and astrophysics [33]. The approach here, in which a particular kind of block-sampled data is recast as a higher-dimensional signal has not, to our knowledge, been described elsewhere.

The one disadvantage to expanding the dimensionality lies with the factor of 2 in the denominator in Eq. (16). The first set of block-sampled points gives a 1D basis of size M_1 . Adding a second block at a much later time gives a 2D basis size that is also just M_1 . The

second block establishes the peaks as 2D, rather than 1D, but will not necessarily contribute to improved resolution because both the size of the basis and the number of peaks have remained fixed. The noise will have been reduced, but the noise would also have been reduced by simply sampling twice as long, without skipping any data points, and doing so would give a 1D basis of size $2M_1$ which, in principle, would have the capability to characterize *twice* as many signal peaks. In benchmark problems where peaks are nearly uniformly dispersed across a known spectral width, the 1D sampling strategy is the most efficient. However, when peaks cluster markedly, the block sampling strategy, with the introduction of a hidden dimension, outperforms. This should not be any great surprise: the more that is known about the data, the better the sampling strategy can be. In particular, some protein experiments can exploit the known position of resonances in the ^{15}N - ^1H plane, obtained from an initial 2D HSQC spectrum [34], to resolve closely spaced peaks in other dimensions much more rapidly and with better sensitivity, compared to extending the sampling along these other dimensions in the usual way.

2.2. Examples of hidden dimension spectra

One common example of spectral features with different scales is that of a 1D proton spectrum of a small molecule. At high field, the chemical shift separation is much larger than the scalar coupling, and resolved multiplets are observed. The spectrum therefore has signal peaks that are clustered near each other, and then larger frequency regions devoid of signals. Although there is no application for incomplete sampling of a 1D FID, it provides a straightforward way to demonstrate the potential advantage of a hidden dimension.

Fig. 7 shows the upfield region of the 1D spectrum of ethylbenzene. With 4096 data points in the FID, 7(a), individual peaks in each multiplet are resolved. Using only the first 22 points, 7(b) leads to an FT spectrum, after zero-filling, in which peaks are very wide, and show sinc-function oscillations in the baseline. Roughly speaking, if there are rather less than eleven *groups* of peaks, then a 2D spectrum with just 22 points along the first dimension, and enough points in the hidden dimension to characterize the number of peaks *within* each group, should suffice. This is indeed the case, as shown in Fig. 7(c). A 2D FDM calculation using a 22×8 data set, followed by integration over the hidden dimension, is able to recover almost all the original information in the full data set. In this case, only 176 points out of 4096 are actually needed to obtain a decent spectrum, a potential times savings of nearly 23-fold if this were a multidimensional spectrum.

Another case in which the hidden dimension idea can be employed is in a conventional ^{13}C - ^1H HSQC. In some compounds the carbon-13 lines can be very close compared to the entire survey spectral width. Using conventional sampling, it may be time-consuming to resolve the individual ^{13}C resonances if the corresponding multiplets in the proton dimension are also close, or overlapping. In this case, a viable strategy is to collect a block of data using the short dwell time corresponding to the full ^{13}C spectral width, and process it to observe whether signals appear clustered or uniformly distributed, or perhaps whether there is some ambiguity in some crowded parts of the 2D spectrum. Knowing the approximate maximum peak frequency difference, an appropriate gap can be introduced (of the order of the inverse of this frequency difference) and block sampling initiated. After four such blocks the 3D FDM spectrum can be calculated, and the additional dimension integrated out. It is usually a quick matter to decide whether one, two, or more ^{13}C resonances are present.

Fig. 8 shows an example from a zoomed region of the ^{13}C - ^1H HSQC of cholesteryl acetate. The assignment is indicated. There are two close diastereotopic methyl groups, Me_{26} and Me_{27} , with

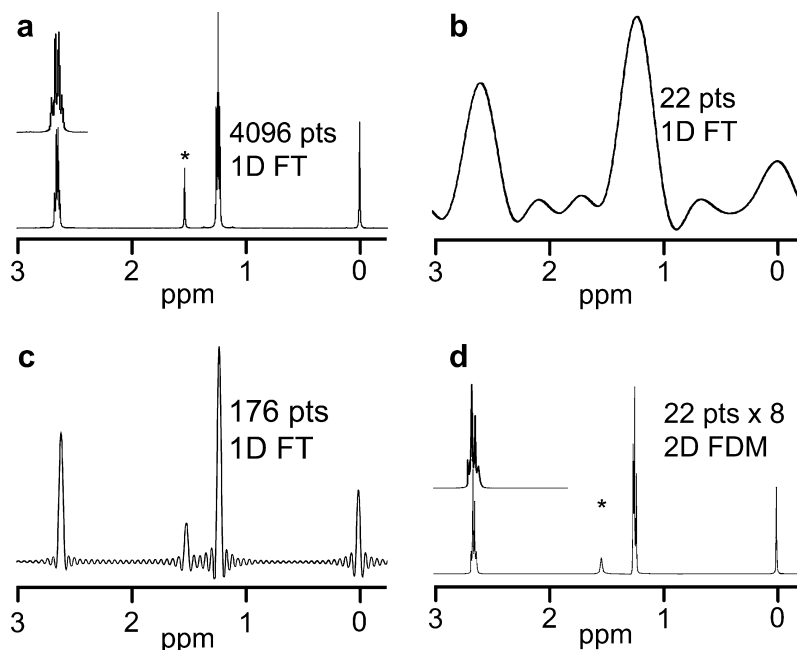


Fig. 7. A simple example showing how to boost the resolution by using a hidden dimension. (a) Conventional upfield ^1H FT spectrum of a standard ethylbenzene sample using 4096 data points. The multiplets are resolved. (b) The same FID as in (a) but using only the first 22 points, replacing the others with zeroes. Very wide transform-limited line shapes are apparent. (c) The result of a 2D FDM calculation using 8 blocks of 22 points (from the same FID again), each block separated by a gap of 490 points so that the last block is near the end of the 4096 point data record. The 2D FDM spectrum was calculated and then the hidden dimension, in which the true frequencies are aliased many times, was then integrated out. Even though only 176 points were used in the calculation, the resolution and the relative intensities are quite good, and the multiplet structure is again visible.

^{13}C resonances near 23 ppm, and with partially overlapping proton resonances (the proton couplings are not resolved). Resolving these two correlations in a single survey-mode 2DFT spectrum is a time-consuming challenge. With 256 points in the ^{13}C dimension, the resolution is too poor to detect anything but a slight hint of two ^{13}C frequencies, but with 512 points the expected two peaks emerge. Using 6 blocks of just 16 points (96 increments total), the 2D projection of the 3D FDM calculation is able to definitely resolve the two close peaks.

Of course it is not necessary that the hidden dimension simply be one of the existing indirect dimensions over a smaller spectral width. It could equally well be an entirely different, independent dimension in which the resolution is better than the plane under consideration. Once again, the multidimensional nature of FDM may make a 3D experiment, in which the extra-dimension boosts the information content, a superior strategy to a 2D experiment in which many increments may be needed to resolve close peaks. When the pulse sequence already contains fixed delays for magnetization transfer, the hidden dimension can be a constant-time evolution during one of these fixed delays, thus optimizing both sensitivity and resolution simultaneously. If the peak positions in the hidden dimension are known beforehand, from a preliminary experiment, then this information can be exploited to narrow the spectral width while still avoiding the possibility of aliasing one peak directly upon another one. If the peak positions are unknown, the full Nyquist range should be used in the hidden dimension.

As an illustrative example, suppose a high-resolution rendition of the $^1\text{H}_{\text{N}(i)}-^{13}\text{C}_{(i-1)}$ plane of labeled protein were of interest. As this is one of the planes of a 3D HNC0 [31] experiment, the most natural course would be to fix the ^{15}N dimension, at $t_{\text{N}}=0$, and increment the ^{13}C dimension until sufficient resolution is achieved. This 2D experiment should not be too time-consuming. However, as the pulse sequence must have fixed delays to allow ^{15}N magnetization to be transferred to and from the directly bonded carbonyl carbon, incrementing the ^{13}C dimension farther

and farther out can cause more magnetization losses due to transverse relaxation. In the 3D experiment, the necessary fixed delays are used to encode the ^{15}N chemical shift in a so-called “constant-time” evolution [35–37], avoiding any additional delays that would be necessary in the alternative “real-time” experiment [38]. It thus may make sense to use the ^{15}N fixed delay, which must be present in any event, as a hidden dimension to resolve peaks in the $^1\text{H}_{\text{N}(i)}-^{13}\text{C}_{(i-1)}$ plane rather than using ^{13}C evolution to do so. If, in addition, the ^{15}N resonance frequencies are known, a smaller spectral width can be chosen in the ^{15}N dimension, making sure that peaks with nearly degenerate ^1H frequencies are not accidentally aliased on top of each other in the nitrogen dimension. If no information on the $^{15}\text{N}-^1\text{H}$ plane is available, then the full Nyquist range must be used. However, the number of increments in the hidden dimension can still be very small.

Figs. 9 and 10 shows that this strategy bears fruit. Obtaining the $^1\text{H}_{\text{N}(i)}-^{13}\text{C}_{(i-1)}$ plane by a conventional 2D strategy, using 24 C' increments, results in a reasonable 2D FT spectrum that has a couple of congested regions near 8.5 ppm in the proton dimension. By comparison, a 4×6 ($^{15}\text{N} \times ^{13}\text{C}'$) 3D data set, in which the ^{15}N hidden dimension is integrated over, clearly reveals resolved peaks in the same region. Increasing the 2D acquisition to 128 C' increments reveals similar information but is a much longer experiment. Applying 2D FDM to the 24-increment data set likewise fails to completely resolve the region in question.

Fig. 10 shows zoomed views of the boxed regions of Fig. 9, to show the details of the improvement in resolution.

3. Experimental

All spectra were obtained at 500 MHz on a Varian UnityPlus NMR spectrometer with a conventional (room temperature) 5 mm HCN triax gradient Varian probe. Sample volumes were approximately 800 μl . The ethylbenzene sample was the 0.1%

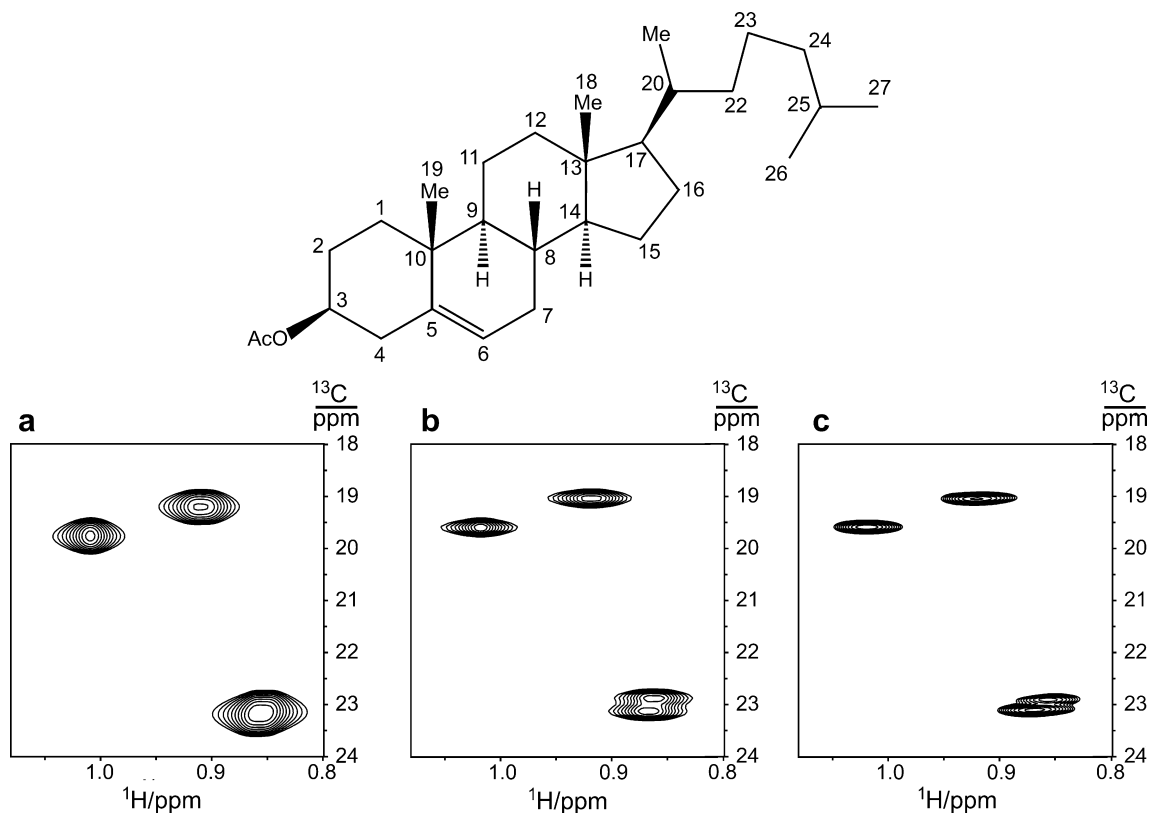


Fig. 8. *Top:* The structure of cholesteryl acetate, with conventional numbering of the carbons. *Bottom:* A zoomed region of a survey HSQC in the upfield methyl region. The carbon-13 spectral width was 16,340 Hz. (a) Conventional 2D FT spectrum using 256 increments in the ^{13}C dimension (~ 64 Hz resolution). The two close peaks near 23 ppm are unresolved. (b) Same as (a) but with 512 increments (~ 32 Hz resolution). The two peaks near 23 ppm are resolved. (c) 2D projections of a 3D FDM calculation using 16 points over the 16,340 Hz spectral width, and 6 increments over a hidden dimension of 100 Hz. In the hidden dimension, the three basis functions per 100 Hz are enough to boost the resolution so that two peaks are observed—with only 96 increments total. For the FT spectra both N- and P-type data were needed to obtain absorption-mode line shapes. The 3D FDM data set used only P-type data in each of the ^{13}C dimensions. The time taken to acquire these data sets, based on the same minimum phase cycling and identical relaxation delays was (a) 158 min, (b) 316 min, and (c) 30 min, showing a factor of ten improvement in throughput by using a hidden dimension and FDM.

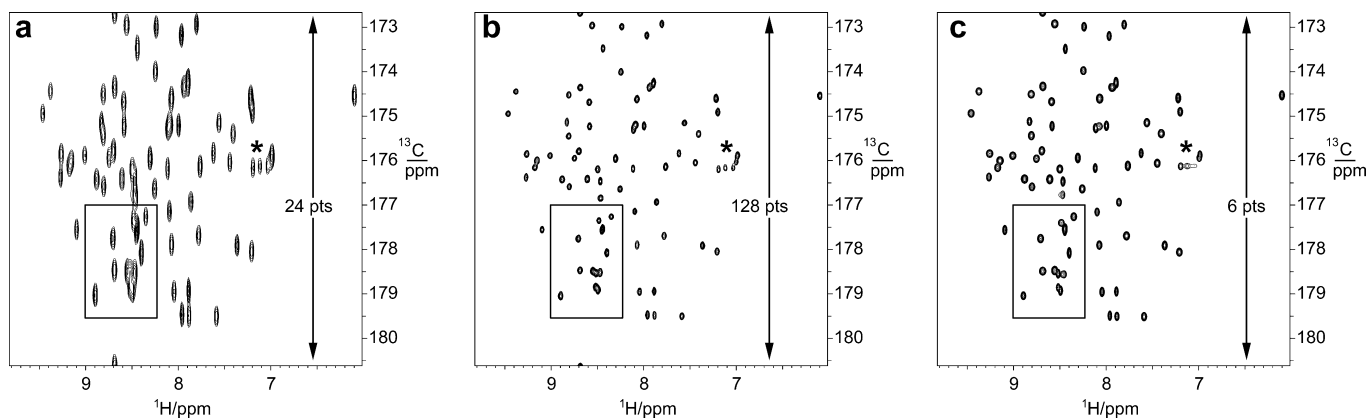


Fig. 9. The $^1\text{H}_{\text{N}(i)}-^{13}\text{C}_{(i-1)}$ plane of human ubiquitin obtained using various numbers of ^{13}C increments. See Section 3 for details. In each panel, the starred group of peaks are from arginine sidechains, and are folded into the spectral region in the ^{13}C dimension. (a) Conventional 2D FT spectrum using 24 increments in the ^{13}C dimension. The boxed region is unresolved. Only the zero-time increment in the ^{15}N dimension, which was fixed, was used. (b) Same as (a) but with 128 ^{13}C increments. The boxed region shows much-enhanced resolution. (c) 2D projections of a 3D FDM calculation using 6 ^{13}C increments and 4 ^{15}N increments. The data were obtained using constant-time evolution in both indirect dimensions. The ^{15}N constant time ($2T$ in the literature) was the usual 12.4 ms required for magnetization transfer. Using the known positions of peaks in the $^1\text{H}_{\text{N}}-^{15}\text{N}$ plane, a spectral width of 659 Hz (13.2 ppm) was employed in the ^{15}N dimension. The ^{13}C constant time $2T$ was 8.0 ms. Although only 24 joint increments were taken, the resolution rivals that of the much longer experiment in (b). Using the full Nyquist range (36 ppm) in the ^{15}N dimension gives somewhat inferior results, although using 6 ^{15}N increments, rather than 4, again allows the full 3D spectrum, with high-resolution along each dimension, to be obtained.

ASTM standard proton sensitivity sample. The cholesteryl acetate sample was approximately 50 mM in CDCl_3 . The ubiquitin sample was 1 mM in 90%/10% $\text{H}_2\text{O}/\text{D}_2\text{O}$ at pH 4.7. Standard pulse sequences were used for the experiments, except that BIP [39] inver-

sion pulses were substituted in pairs for conventional 180° pulses and care was taken to suppress artifacts that could arise from unwanted magnetization transfers, as described previously [40]. The spectra are completely representative of what could be obtained in

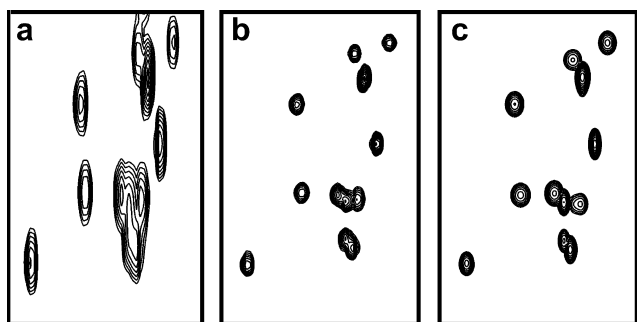


Fig. 10. A more detailed view of the boxed regions in Fig. 9, showing the comparison between (a) 24 C' increments only, (b) 128 C' increments only, and (c) 6 C' increments and 4 ^{15}N increments. The first two panels are FT spectra, and the third is a 3D FDM spectrum in which the ^{15}N dimension has been integrated out to give the 2D spectrum shown.

any modern facility. At higher field and with superior sensitivity, our approach will perform even better than shown herein. Further details of all the experimental conditions, including files for the BIP pulses that were employed, can be found in [Supplementary Material](#).

4. Discussion

We have shown that there are cases in which there is much to gain by exploiting the full HD spectrum. Hidden dimensions, either employing severe aliasing over a smaller spectral width, or a different nuclear spin species altogether, can boost the content of the HD data set. While none of the 1D or 2D FT projections of the data shown in Fig. 9(c) show resolved features with so few increments in the ^{15}N and ^{13}C dimensions, the full 3D object is fully resolved by FDM, allowing *post hoc* high-resolution 1D or 2D projections to be calculated if they are of interest. The flexibility of FDM, in which all the data work together to determine the HD peak positions, makes it possible to sample fewer points along dimensions that are unfavorable for sensitivity reasons, and yet still obtain good resolution. The onset of the sudden convergence of FDM, once the number of multidimensional basis functions and the underlying SNR are adequate, gives this approach a measure of reliability that is not addressed as easily by a disjoint collection of lower-dimensional projections.

Of course, there are natural limits to employing HD spectra. If too many magnetization transfers, all of which may require delays to effect, are needed, then sensitivity may suffer even if resolution is boosted. In the extreme case of many dimensions there may be too little SNR to employ FDM analysis at all. However, when an unused dimension is *already* present along with the associated delays, as in the HN(CO)CA experiment [36], the methods described here are quite applicable and may lead to superior resolution for the same instrument time.

There is an interesting connection between orderly skipping of data points and the resulting density of basis functions in a higher-dimensional space. Our results seem to indicate that, for unknown peak dispositions within a given Nyquist range, simple uniform sampling is usually the best option for FDM. When peaks cluster narrowly or appear in groups, or if some specific question regarding close unresolved features is to be settled, then block sampling may offer some distinct advantages.

5. Conclusions

Filter diagonalization can be a speedy and effective way to analyze NMR data that has sufficient SNR and is well matched to the

assumption of a finite number of multidimensional Lorentzian peaks. The ability of FDM to process just P-type data and still obtain absorption-mode line shapes is useful in hidden dimension spectra, where in some cases the corresponding N-type data set is not readily obtained. Although the FDM algorithm does not use prior knowledge *per se*, if peak positions are thought to be known from a previous experiment, they can be exploited to narrow down the spectral range, allowing aliasing to occur as long as no accidental degeneracies result. Optimizing data acquisition in this way allows surprisingly short time-domain data sets to deliver very high-resolution spectra, and will be generally useful for a large class of commonly-used NMR experiments.

Acknowledgments

This research was made possible by support from the National Science Foundation, CHE-0703182, and by a UC Discovery Grant BIO05-10533. B.D.N. thanks the UCI School of Physical Sciences for partial support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmr.2008.09.021](https://doi.org/10.1016/j.jmr.2008.09.021).

References

- [1] H. Oschkinat, C. Griesinger, P.J. Kraulis, O.W. Sørensen, R.R. Ernst, A.M. Gronenborn, G.M. Clore, Three-dimensional NMR spectroscopy of a protein in solution, *Nature* 332 (1988) 374–376.
- [2] E.R. Zuiderweg, S.W. Fesik, Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a, *Biochemistry* 28 (1989) 2387–2391.
- [3] D. Marion, P.C. Driscoll, L.E. Kay, P.T. Wingfield, A. Bax, A.M. Gronenborn, G.M. Clore, Overcoming the overlap problem in the assignment of ^1H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^1H – ^{15}N Hartmann–Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta, *Biochemistry* 28 (1989) 6150–6156.
- [4] M. Ikura, L.E. Kay, A. Bax, A novel approach for sequential assignment of ^1H – ^{13}C , and ^{15}N , spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin, *Biochemistry* 29 (1990) 4659–4667.
- [5] L.E. Kay, NMR studies of protein structure and dynamics, *J. Magn. Reson.* 173 (2005) 193–207.
- [6] A. Grishaev, V. Tugarinov, L.E. Kay, J. Trewthella, A. Bax, Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints, *J. Biomol. NMR* 40 (2008) 95–106.
- [7] W.J. Placzek, M.S. Almeida, K. Wüthrich, NMR structure and functional characterization of a human cancer-related nucleoside triphosphatase, *J. Mol. Biol.* 367 (2007) 788–801.
- [8] O.V. Tsodikov, D. Ivanov, B. Orelli, L. Staresincic, I. Shoshani, R. Oberman, O.D. Schärer, G. Wagner, T. Ellenberger, Structural basis for the recruitment of ERCC1–XPF to nucleotide excision repair complexes by XPA, *EMBO J.* 26 (2007) 4768–4776.
- [9] K. Pervushin, R. Riek, G. Wider, K. Wüthrich, Attenuated T_2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicated and avenue to NMR structures of very large biological macromolecules in solution, *Proc. Natl. Acad. Sci. USA* 94 (1997) 12366–12371.
- [10] K. Pervushin, R. Riek, G. Wider, K. Wüthrich, Transverse relaxation-optimized spectroscopy (TROSY) for NMR studies of aromatic spin systems in C-13-labeled proteins, *J. Am. Chem. Soc.* 120 (1998) 6394–6400.
- [11] G. Wider, NMR techniques used with very large biological macromolecules in solution, *Methods Enzymol.* 394 (2005) 382–398.
- [12] S. Kim, T. Szyperski, GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information, *J. Am. Chem. Soc.* 125 (2003) 1385–1393.
- [13] E. Kupce, R. Freeman, New methods for fast multidimensional NMR, *J. Biomol. NMR* 27 (2003) 101–113.
- [14] E. Kupce, R. Freeman, Projection–reconstruction technique for speeding up multidimensional NMR spectroscopy, *J. Am. Chem. Soc.* 126 (2004) 6429–6440.
- [15] M. Mobli, A.S. Stern, J.C. Hoch, Spectral reconstruction methods in fast NMR: reduced dimensionality, random sampling and maximum entropy, *J. Magn. Reson.* 182 (2006) 96–105.
- [16] B.E. Coggins, P. Zhou, PR-CALC: a program for the reconstruction of NMR spectra from projections, *J. Biomol. NMR* 34 (2006) 179–195.

- [17] G. Lippens, P.R. Bodart, F. Taulelle, J.-P. Amoureux, Restricted linear least square treatment processing of heteronuclear spectra of biomolecules using the ANAFOR strategy, *J. Magn. Reson.* 161 (2003) 174–182.
- [18] C. Griesinger, O.W. Sørensen, R.R. Ernst, Practical aspects of the E-COSY technique. Measurement of scalar spin–spin coupling constants in peptides, *J. Magn. Reson.* 75 (1987) 474–492.
- [19] H. Hu, Q.N. Van, V.A. Mandelshtam, A.J. Shaka, Reference deconvolution, phase correction, and line listing of NMR spectra by the 1D filter diagonalization method, *J. Magn. Reson.* 134 (1998) 76–87.
- [20] V.A. Mandelshtam, H.S. Taylor, A.J. Shaka, Application of the filter diagonalization method to one- and two-dimensional NMR spectra, *J. Magn. Reson.* 133 (1998) 304–312.
- [21] J. Chen, V.A. Mandelshtam, A.J. Shaka, Regularization of the two-dimensional filter diagonalization method: FDM2K, *J. Magn. Reson.* 146 (2000) 363–368.
- [22] J. Chen, A.A. De Angelis, V.A. Mandelshtam, A.J. Shaka, Progress on the two-dimensional filter diagonalization method. An efficient doubling scheme for two-dimensional constant-time NMR, *J. Magn. Reson.* 162 (2003) 74–89.
- [23] J. Chen, D.N. Nietispach, A.J. Shaka, V.A. Mandelshtam, Ultra-high resolution 3D NMR spectra from limited-size data sets, *J. Magn. Reson.* 169 (2004) 215–224.
- [24] G.S. Armstrong, V.A. Mandelshtam, A.J. Shaka, B. Bendiak, Rapid high-resolution four-dimensional NMR spectroscopy using the filter diagonalization method and its advantages for detailed structural elucidation of oligosaccharides, *J. Magn. Reson.* 173 (2005) 160–168.
- [25] G.S. Armstrong, B. Bendiak, High-resolution four-dimensional carbon-correlated H-1–H-1 ROESY experiments employing isotags and the filter diagonalization method for effective assignment of glycosidic linkages in oligosaccharides, *J. Magn. Reson.* 181 (2006) 79–88.
- [26] V.A. Mandelshtam, FDM: the filter diagonalization method for data processing in NMR experiments, *Prog. NMR Spectrosc.* 38 (2001) 159–196.
- [27] J.C. Hoch, A.S. Stern, *NMR Data Processing*, Wiley-Liss, New York, 1996. Chapter 4.
- [28] M. Mobli, M.W. Maciejewski, M.R. Gryk, J.C. Hoch, Automatic maximum entropy spectral reconstruction in NMR, *J. Biomol. NMR* 39 (2007) 133–139.
- [29] D.P. Frueh, Z.Y. Sun, D.A. Vosburg, C.T. Walsh, J.C. Hoch, G. Wagner, Non-uniformly sampled double-TROSY hNcaNH experiments for NMR sequential assignments of large proteins, *J. Am. Chem. Soc.* 128 (2006) 5757–5763.
- [30] G. Bodenhausen, R. Freeman, R. Niedermeyer, D.L. Turner, Double fourier transformation in high-resolution NMR, *J. Magn. Reson.* 26 (1977) 133–164.
- [31] S. Sibisi, Two-dimensional reconstructions from one-dimensional data by maximum entropy, *Nature* 301 (1983) 134–136.
- [32] P.D. Welch, The use of fast Fourier transforms for the estimation of power spectra: a method based on time averaging over short modified periodograms, *IEEE Trans. Audio Electroacoustics* 15 (1967) 70–73.
- [33] J.A. Högbom, Aperture synthesis with a nonregular distribution of interferometer baselines, *Astron. Astrophys. Suppl.* 15 (1974) 417–424.
- [34] D. Rubin, G. Bodenhausen, Natural abundance N-15 NMR by enhanced heteronuclear spectroscopy, *Chem. Phys. Lett.* 69 (1980) 185–189.
- [35] L.E. Kay, M. Ikura, R. Tschudin, A. Bax, Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins, *J. Magn. Reson.* 89 (1990) 496–514.
- [36] A. Bax, R. Freeman, Investigation of complex networks of spin–spin coupling by two-dimensional NMR, *J. Magn. Reson.* 44 (1981) 542–561.
- [37] R. Powers, A.M. Gronenborn, G.M. Clore, A. Bax, Three-dimensional triple-resonance NMR of C-13/N-15-enriched proteins using constant-time evolution, *J. Magn. Reson.* 94 (1991) 209–213.
- [38] M. Sattler, J. Schleucher, C. Griesinger, Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, *Prog. Nucl. Mag. Res. Sp.* 34 (1999) 93–158.
- [39] M.A. Smith, H. Hu, A.J. Shaka, Improved broadband inversion performance for NMR in liquids, *J. Magn. Reson.* 151 (2001) 269–283.
- [40] S. Keppetipola, W. Kudlicki, B.D. Nguyen, X. Meng, K. Donovan, A.J. Shaka, From gene to HSQC in under five hours: high-throughput NMR proteomics, *J. Am. Chem. Soc.* 128 (2006) 4508–4509.