JOURNAL OF MAGNETIC RESONANCE 73, 69–77 (1987)

Exponential Sampling, an Alternative Method for Sampling in Two-Dimensional NMR Experiments

J. C. J. BARNA, E. D. LAUE, M. R. MAYGER,* J. SKILLING,* AND S. J. P. WORRALL

*Department of Biochemistry, Tennis Court Road, Cambridge CB2 1QW, United Kingdom, and *Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge CB3 9EW, United Kingdom*

Received September 26, 1986

A new method for sampling in two-dimensional nuclear magnetic resonance experiments is proposed and tested using one-dimensional spectra as models. The free induction decays are sampled exponentially, using many points where the signal-to-noise ratio (*S/N*) is high and a few where it is low. Using the maximum entropy method to reconstruct spectra, much higher resolution can be obtained than by using conventional sampling (for a given number of data points). The method is shown to work for FIDs having even very poor *S/N*. It should prove valuable in the future for 2D NMR experiments where at present valuable high-resolution information is lost as a result of the necessity for truncation of data sets in t_1 in order to optimize sensitivity.

INTRODUCTION

Strategies which combine the use of phase-sensitive homonuclear proton two-dimensional nuclear magnetic resonance techniques with subsequent computer-aided calculation of structures are well developed for the study of small proteins and other biomolecules (*1*). There are, however, various limitations which continue to prompt a search for improvements in experimental design and in methods of data processing. For convenience we may roughly divide the problems according to the molecular weight of the molecules under study. We define "small" as $M_r < 10{,}000$, "medium" as $M_r \sim 10{,}000$–$25{,}000$, with all others classed as "large." We do not consider the case of "large" molecules here.

For small proteins, DNA fragments, etc., the available homonuclear proton 2D NMR techniques are in principle capable of yielding a wealth of information and the limitations are mainly practical. We have discussed these problems in a recent paper on the use of the maximum entropy method (MEM) to reconstruct phase-sensitive 2D spectra (*2*). A major problem in practice is that data sets are usually recorded so that the signal is truncated in t_1 in order to optimize sensitivity (*3*). We (*2*) and others (*4, 5*) have shown how MEM avoids, to a large extent, the resulting artifacts and the loss of resolution found in spectra processed using conventional data processing techniques.

Most molecules of biological interest do not, unfortunately, belong to the "small" class. For "medium"-sized molecules, spectra are inherently very complex and signals

0022-2364/87 $3.00

are broad. Methods have been proposed to counter these limitations. Such methods can obtain better resolution in ω_1 by utilizing the greater chemical-shift dispersion of another nucleus such as ^{13}C, ^{15}N, or ^{31}P (*6*). They are often used in conjunction with isotopic substitution or enrichment. Various 2D experiments of the ^1H (t_2)–X nucleus (t_1) type will continue to be developed. Alternatively isotopic substitution can be employed to simplify the spectra obtained using a modified homonuclear 2D experiment (*7*).

With both classes of molecular size and both homonuclear and heteronuclear experiments in mind we have sought a general method to further improve the resolution available in ω_1. We describe here an alternative method of sampling for 2D NMR (*8*).

EXPONENTIAL SAMPLING OF A DECAYING SIGNAL

In conventional Fourier transform NMR one always samples the signal at fixed intervals in both t_2 and t_1. One also samples for equal amounts of time at points where the signal-to-noise ratio (S/N) is high and where it is lower. When optimizing sensitivity by truncating the data, one necessarily sacrifices resolution, because the information about high-spectral frequencies, which gives good resolution, is in the latter part of the free induction decay in both t_2 or t_1. MEM can "reconstruct" some of the lost resolution, in effect by extrapolating from a given data set but there are limits to this. It seemed that a more logical way to sample an exponentially decaying signal would be to do so in an exponential manner, thus giving a better compromise between S/N and resolution. Many points would be sampled where S/N is high but a few would be sampled where S/N is very low, to aid the MEM reconstruction of high-resolution information. This idea is related to a proposal by Levitt and Freeman who suggested that one could exponentially decrease the amount of signal averaging throughout t_1 (*9*). In their method, the spacing between t_1 points would be equal and the total number of points would be large to obtain good resolution. The rationale for their method is solely that S/N decreases exponentially in t_1. By contrast, our proposal has a second basis quite separate from S/N arguments. If only a given (small) number of points can be sampled, then conventional sampling will give poor resolution in comparison to exponential sampling. This depends purely on the fact that the data are Fourier coefficients and not on the S/N. Thus even a conventional FT of an exponentially sampled data set will show improved resolution though at the expense of a new type of "truncation artifact" caused by missing data points; MEM minimizes such artifacts.

We have tested our proposal on various one-dimensional spectra (either simulated or conventionally acquired) which can act as models for the second dimension (t_1) of future 2D NMR experiments. Exponential sampling was produced before reconstruction using MEM, as described below. Eventually, the spectrometer would be programmed to sample in this manner during 2D experiments.

The principles are illustrated in Fig. 1. The conventional Fourier transform of a simulated FID (128 points) produces the spectrum shown in Fig. 1a. However, if the FID has only 17 measured points, sampled in an exponential manner from the first 64 points, its (128 point) conventional Fourier transform (setting unmeasured points to zero for lack of better values) contains such severe artifacts that the frequency information is ambiguous (Fig. 1b).

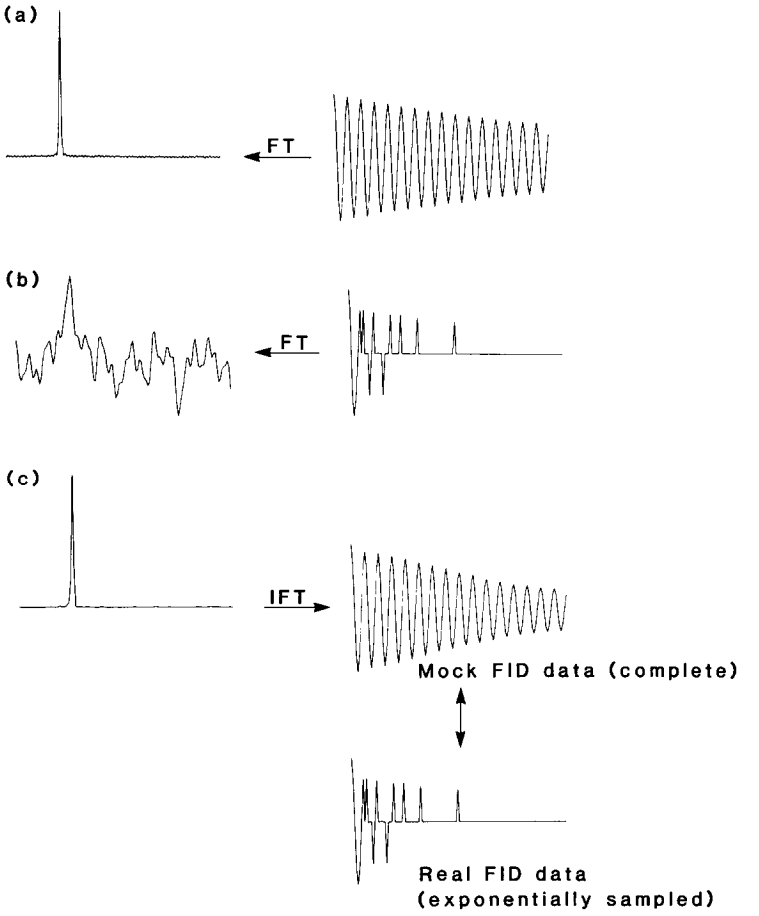

FIG. 1. (a) A simulated 128 point FID and its conventional Fourier transform. (b) A simulated FID where 17 points were exponentially sampled out of the first 64 points and its (128 point) conventional Fourier transform. (c) Flowchart showing a MEM reconstruction from the same data set as in Fig. 1b.

In a MEM reconstruction (Fig. 1c), a well digitized (128 point) trial spectrum is inverse Fourier transformed to give a mock data set (FID). The resulting mock FID data are then checked for consistency with the real FID (where 17 points were sampled in an exponential manner out of 64 points), though of course only the points corresponding to the measured points can be used. In successive iterations the trial spectrum is modified until its entropy is maximized, subject to the constraint that the corresponding mock FID data agree with the experimental data to within the noise (*10*). The trial spectrum is then the maximum entropy spectrum. The missing unmeasured points in the FID are allowed to take whatever values maximize the entropy of the spectrum, and in this way truncation artifacts are minimized. In a real situation exponential sampling would result in a great saving of time in t_1, or more importantly, a much better result for a given time.

The procedure used to find n exponentially sampled points out of N uniformly sampled points, ($n < N$), is described as follows. First we define a continuous sampling density function

$$D(x) = \exp(-kx), \qquad x > 0,$$

where k is calculated to make $\int_0^N D(x)dx = n - 1$. From $D(x)$ we then calculate the successive rounded integrals

$$I_j = \text{integer part}\left(\int_0^j D(x)dx\right) \qquad \text{for} \qquad j = 1, \ldots, N.$$

For each integer $l = 0, 1, \ldots, n - 1$ we sample at the first data point j which satisfies $I_j = l$. In this way we sample more points when the density, $D(x)$, is large and less often when it is small, producing an exponentially sampled data set.

In the following sections we show the application of exponential sampling to real data sets and discuss practical considerations related to S/N.

PROTON SPECTRA; A MODEL FOR ^{1}H-^{1}H CORRELATION

Proton (^{1}H) spectra of 2-furoic acid were obtained on a Bruker WP80 spectrometer from a sample in dimethyl sulfoxide. The AMX spin system was used to investigate the ability of MEM to reconstruct the high-resolution information (needed for determination of coupling constants), in conjunction with exponential sampling. It is this type of information which may be lost in 2D NMR experiments because the data set is truncated. We needed to ensure also that exponential sampling does not increase resolution at the expense of degrading the S/N too seriously.

FIDs of 1024 points were collected. The signal decayed into the noise in 1024 points. A comparison was made between spectra produced by conventional FTs, "conventional" MEM, and MEM reconstructions of exponentially sampled data, produced from either 512, 256, or 128 points. As the conventional FT of 512 points retained virtually all the resolution of the 1024 point spectrum but had improved S/N, this spectrum represented the best available Fourier result (Fig. 2a). Figure 2 shows how, for a given number of sampled points (i.e., 128) MEM (c) gives a better result than does conventional FT (b) from a conventionally sampled FID. Exponential sampling of 128 points out of 512 points followed by reconstruction of the spectrum with MEM can, however, reproduce the full splitting pattern with some changes in intensity but no serious loss of S/N (d). The results of a systematic investigation of the effect of noise are discussed below.

We plan next to apply exponential sampling to proton correlation spectra of small proteins. For a small protein, studied at 400 or 500 MHz, optimal resolution would require 2048–4096 points. Conventionally one might hope to obtain, e.g., a data matrix of 1024 × 4096 points ($t_1 \times t_2$), thus obtaining much worse resolution in t_1 than in t_2. If the 1024 points are exponentially sampled out of 4096 points, resolution will be comparable in both dimensions after reconstruction of the spectrum using MEM. For some phase-sensitive 2D spectra this could help avoid mutual cancellation of positive and negative components within and between cross peaks.

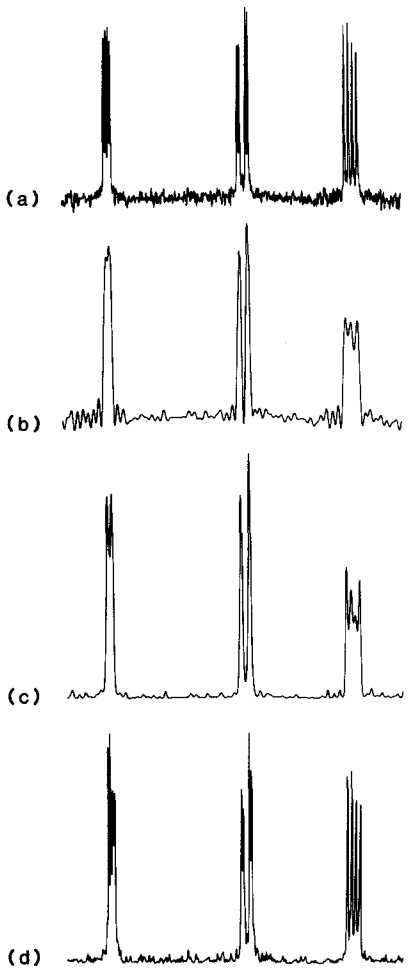

FIG. 2. ^{1}H spectra of the AMX spin system in 2-furoic acid in d_6-dimethyl sulfoxide produced by (a) a conventional Fourier transform (FT) of the first 512 points of the 1024 point FID, (b) a conventional FT of the first 128 points, (c) a MEM reconstruction using the first 128 points, and (d) a MEM reconstruction using 128 points exponentially sampled out of the first 512 points. All spectra have equal digital resolution.

CARBON SPECTRA; A MODEL FOR X-^{1}H CORRELATION

Carbon-13 spectra of a peptide in water were obtained on a Bruker AM400 spectrometer. These spectra are considerably noisier and more complex than the ^{1}H spectra and thus pose a more severe test of exponential sampling. Our illustrations here show the carbonyl region which is the most complex. Of course in practice, for correlations using one-bond couplings only protonated X nuclei would be involved; carbonyl signals could be used for correlations involving longer range couplings. The FIDs were of 32K points. Conventional or exponential samples were taken from the first 16K points, because the resolution was only marginally better when the full FIDs were used. The effect of taking exponential samples from either 16K, 8K, 4K, or 2K points (where

possible) was investigated. In this case the problem is to resolve crowded resonances, not coupling patterns, but the principles are the same. As for the ^1H spectra, it was found that exponential sampling produced an improvement in resolution but it appeared that the poorer S/N was very relevant. Better results were obtained by sampling, e.g., 2K points out of 8K than out of 16K points. Sampling too sparsely gave artifacts, e.g., spurious resolution, loss in resolution, and distortions in intensity (Fig. 3). This aspect was investigated systematically using simulated data.

For the "medium"-sized molecules for which exponential sampling of the ^{13}C FID in t_1 is designed, natural linewidths of protonated carbon signals are much greater than are those in the model spectra so optimal resolution would require fewer points

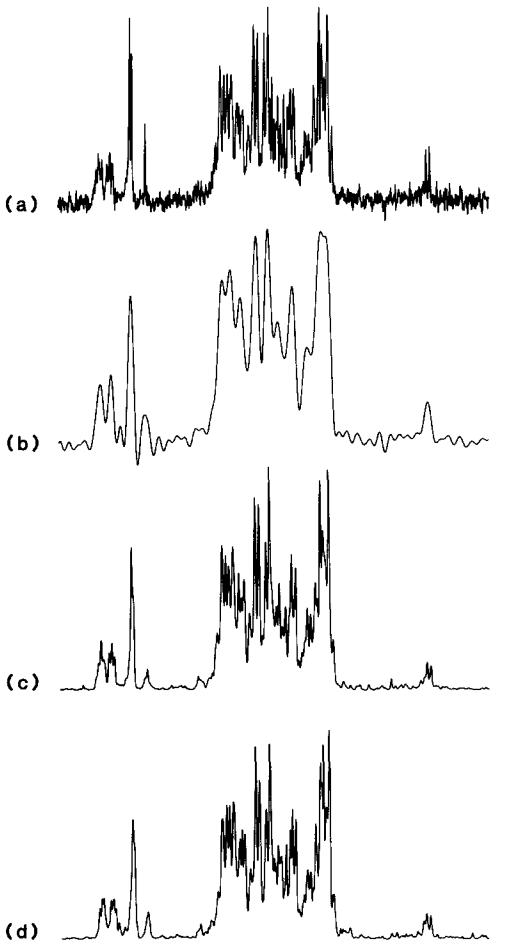

FIG. 3. ^{13}C spectra of the carbonyl signals of a synthetic peptide ($M_r \sim 3000$) in water produced by (a) a conventional FT of the first 16K points of the 32K point FID, (b) a conventional FT of the first 2K points, (c) a MEM reconstruction using 2K points exponentially sampled from the first 8K points, and (d) a MEM reconstruction using 2K points exponentially sampled from the first 16K points. All spectra have equal digital resolution.

(e.g., 1024). Conventional acquisition of this many points in t_1 is nevertheless often impractical (for sensitivity reasons) so exponential sampling should prove valuable.

ESTIMATION OF IDEAL SAMPLING REGIMES; THE EFFECT OF NOISE

A simulated FID was obtained using the SIMPLTN program (*11*). Varying amounts of Gaussian noise were then added to the FID which had 1024 points (with an acquisition time of $\sim 5 \times T_2$) and samples of 512, 256, 128, and 64 points were investigated. Figure 4 shows that for an almost theoretical FID, exponential sampling recovers excellent resolution from 128 points sampled out of 1024 points. When much noise is present it becomes advantageous to sample from as few as 256 points (e.g.,

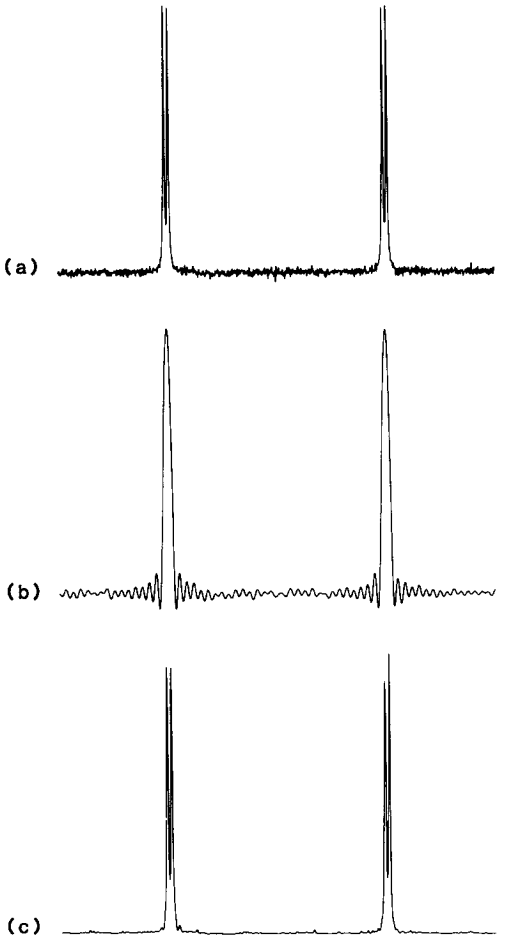

FIG. 4. Simulated spectra produced by (a) a conventional FT of a 1024 point FID after adding noise with a standard deviation equal to 1% of the first FID point, (b) a conventional FT of the first 128 points of the FID, and (c) a MEM reconstruction using 128 points exponentially sampled out of the entire 1024 point FID. All spectra have equal digital resolution.

Fig. 5). The technique does not break down even for very poor S/N, though intensities are affected.

In general, the more noisy the data are expected to be, the fewer points should be taken for the sampling range. Less resolution can reliably be obtained from noisy data but nevertheless much more than by using conventional sampling. In practice one would attempt to judge the sampling regime to be used in t_1 by first looking at one-dimensional spectra of the same molecule.

We have not attempted to compare S/N of MEM reconstructions and conventional spectra because the background noise level seen in MEM reconstructions can often be made arbitrarily low by use of a low default (*10*). Using MEM we can also improve S/N and resolution by the use of multiplication of the trial FID by a negative expo-

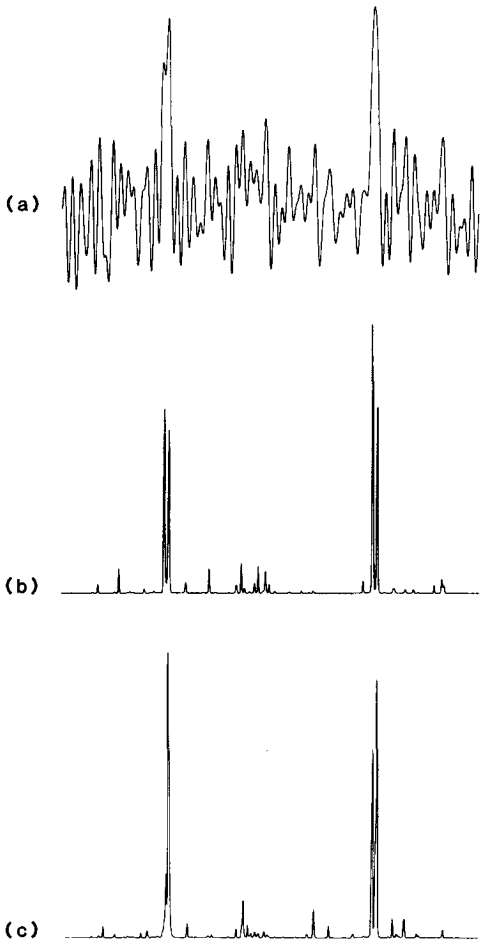

FIG. 5. Simulated spectra produced by (a) a conventional FT of the first 128 points of an FID as used in Fig. 4 but with 40% noise, (b) a MEM reconstruction using 128 points sampled exponentially from the first 256 points of the FID, and (c) a MEM reconstruction using 128 points exponentially sampled out of the first 512 points. All spectra have equal digital resolution.

nential (as part of the spectrum to mock data transform) (*2, 12*). No such multiplication has been used to produce any of the spectra shown in this paper. (The conclusions remain the same, however, when such multiplication is used.) Moreover, to produce a fair comparison of resolution, digital filtering was not used in any of the conventionally processed spectra.

SUMMARY

Exponential sampling of an FID has been shown to be an efficient way of producing high-resolution spectra for a given limit on the number of points to be sampled. It should prove to be capable of giving, after MEM reconstructions, greatly increased resolution in t_1 for a given recording time in homonuclear and heteronuclear 2D NMR spectra of biomolecules.

It should be noted that the method is not exactly suited to experiments where the signal does not simply decay exponentially in t_1 (e.g., double-quantum filtered COSY (*13*)). Modified sampling schemes should give improved results for such experiments, although exponential sampling should still be a useful first approximation.

ACKNOWLEDGMENTS

We thank the Science and Engineering Research Council for financial support, Dr. A. D. Bain for a copy of the SIMPLTN program, and Dr. R. N. Perham and Miss S. E. Radford for the ^{13}C spectra of the peptide. This contribution is from the Cambridge Biochemical NMR group. The NMR facilities were funded jointly by the Medical Research Council and the Science and Engineering Research Council.

REFERENCES

1. W. BRAUN, G. WAGNER, E. WÖRGÖTTER, M. VAŠÁK, J. H. R. KÄGI, AND K. WÜTHRICH, *J. Mol. Biol.* **187,** 125 (1986).
2. E. D. LAUE, M. R. MAYGER, J. SKILLING, AND J. STAUNTON, *J. Magn. Reson.* **68,** 14 (1986).
3. M. H. LEVITT, G. BODENHAUSEN, AND R. R. ERNST, *J. Magn. Reson.* **58,** 462 (1984).
4. P. J. HORE, *J. Magn. Reson.* **62,** 561 (1985).
5. J. C. HOCH, *J. Magn. Reson.* **64,** 436 (1985).
6. R. H. GRIFFEY, A. G. REDFIELD, R. E. LOOMIS, AND F. W. DAHLQUIST, *Biochemistry* **24,** 817 (1985).
7. D. NEUHAUS, G. WIDER, G. WAGNER, AND K. WÜTHRICH, *J. Magn. Reson.* **57,** 164 (1984).
8. J. C. J. BARNA, E. D. LAUE, M. R. MAYGER, J. SKILLING, AND S. J. P. WORRALL, *Biochem. Soc. Trans.* **14,** 1262 (1986).
9. M. H. LEVITT AND R. FREEMAN, *J. Magn. Reson.* **39,** 533 (1980).
10. J. SKILLING AND R. K. BRYAN, *Mon. Not. R. Astron. Soc.* **211,** 111 (1984).
11. T. ALLMAN AND A. D. BAIN, *J. Magn. Reson.* **68,** 533 (1986).
12. (a) S. SIBISI, J. SKILLING, R. G. BRERETON, E. D. LAUE, AND J. STAUNTON, *Nature (London)* **311,** 446 (1984). (b) E. D. LAUE, J. SKILLING, J. STAUNTON, S. SIBISI, AND R. G. BRERETON, *J. Magn. Reson.* **62,** 437 (1985).
13. U. PIANTINI, O. W. SØRENSEN, AND R. R. ERNST, *J. Am. Chem. Soc.* **104,** 6800 (1982).