NMR Data Processing*

**Gareth A. Morris**

School of Chemistry
University of Manchester
Oxford Road
Manchester
M13 9PL
UK

email g.a.morris@manchester.ac.uk

tel +44 161 275 4665

fax +44 161 275 4598

Keywords

Baseline; 2D; DOSY; Fourier transform; Free induction decay; Linear prediction; Lorentz-Gauss transformation; Maximum entropy; Phase; Resolution enhancement; STOCSY; Zero-filling

Abstract

The use of data processing methods in high resolution NMR spectroscopy is explained. These approaches include manipulation of the NMR FID to improve the signal-to-noise ratio or spectral resolution, manipulations of the frequency spectrum to correct phases and base line errors, and the determination of peak positions and intensities. Various approaches for improving incomplete data sets are also outlined. Finally, some more recent approaches for fast 2-dimensional NMR and for statistical analysis of spectra are covered.

$F$, $v$
frequency

i
$\sqrt{-1}$

$m$
number of complex coefficients

$M$, $N$
number of complex points

$p$
probability distribution

$s_k$
$k$th point in free induction decay

$S_n$
$n$th point in spectrum

$S\,(p)$
Shannon informational entropy

$S_\mathrm{p}(v)$
phased spectrum

$s\,(t)$
free induction decay

$t_1$
evolution time

$t_2$
real time

$t_e$
exponential time constant

$t_g$
Gaussian time constant

$t_s$
time shift of Gaussian

$T$
decay constants

$W$
line width

$W(t)$
weighting function

$\alpha_j$
$j$th complex amplitude

$\beta_j$
$j$th complex exponential

$\delta$
delay time

$\phi$
phase shift

# Introduction

Computers play a central part in modern NMR spectroscopy. Their use for the real-time control of pulsed NMR experiments has enabled the development of multiple pulse techniques such as two-dimensional NMR; this article deals with the part played by computers in the acquisition, processing and presentation of experimental NMR data.

All NMR experiments rely on the excitation of a nuclear spin response by a radiofrequency magnetic field, usually in the form of a short pulse or pulse sequence. This generates a rotating nuclear magnetic moment, which induces a small oscillating voltage in the probe coil: a 'free induction decay' (FID). The spectrometer receiver amplifies this voltage, and it is shifted down into the audio frequency range, with high-frequency components being filtered out, before the digitized signal is recorded for processing. Two signals with phases 90° apart are recorded ('quadrature detection'), to allow the relative signs of frequencies to be distinguished. In current spectrometers the frequency shifting, filtration, and quadrature detection are all performed digitally ("digital signal processing", DSP), the high-frequency signal being digitized by a fast analogue-to-digital converter (ADC). Previous spectrometers used two independent 0° and 90° analogue electronic channels for the frequency shifting ("mixing") and filtration, digitizing only at the last stage when the range of signal frequencies is low and hence a relatively slow ADC suffices.

From this point onwards the signal is handled digitally until it is presented to the experimenter as a printed spectrum or an interactive spectral display. The three main stages of processing are first, the acquisition, using analogue or digital signal processing, of a filtered, averaged time-domain recording of the nuclear spin response; second, the generation of a frequency-domain spectrum, usually but not always by Fourier transformation; and third, postprocessing of the spectrum to aid its interpretation. The three stages are summarized briefly below, followed by more detailed discussions of the techniques used and some practical illustrations.

# Data Acquisition

NMR experiments generally require the co-addition of a number of recordings of the nuclear spin response ('transients'), often using different permutations of radiofrequency pulse and receiver phases ('phase cycling'). The data are recorded as a set of complex numbers which sample the in-phase and quadrature NMR signals as a function of time. Early Fourier transform spectrometers had limited word length and required some scaling of the data before coaddition, but this is no longer necessary and successive transients are simply added to memory. The use of DSP gives better filtration of unwanted noise and signals, better effective ADC resolution, and less baseline distortion.

In two-dimensional (2D) NMR a series of FIDs is acquired using a pulse sequence containing a variable evolution period, which is incremented regularly to map out the behaviour of the nuclear signal as a function both of real time $t_2$ during the FID, and of the evolution time $t_1$. A typical 2D NMR experiment might acquire 512 FIDs of 1024 complex points, which would then be doubly Fourier transformed as a function of the two time variables. 3D and 4D NMR extend the

principle to two and three evolution periods respectively, but time constraints limit the number of samples in each dimension more severely, and increase the pressure to use non-uniform sampling of the signal (see below). A very effective alternative where high resolution is not required is to use methods borrowed from echo-planar imaging to encode indirect spectral dimensions spatially, allowing a complete 2D dataset to be recorded in less than a second.

# Spectrum Generation

To make the raw experimental data interpretable they must be converted into a frequency-domain spectrum. The classical, and commonest, method is discrete Fourier transformation. This is a linear operation: the information content of the data is unchanged. Alternative methods such as maximum entropy reconstruction and linear prediction are nonlinear, changing the information content. In Fourier processing a weighting function is usually applied to the time-domain data before transformation; a DC correction based on the last portion of the data may also be performed. After weighting and any zero-filling (see below), the fast Fourier transform (FFT) algorithm is used to produce the frequency spectrum. The resultant spectrum consists of a set of complex numbers sampling a defined frequency range at equal intervals. In general the real and imaginary parts of the spectrum will both be mixtures of absorption mode and dispersion mode signals; a spectrum suitable for display and analysis is obtained by taking linear combinations of the real and imaginary data ('phasing'), or, if that is not possible, by taking the modulus ('absolute value mode') or square modulus ('power mode').

In 2D NMR, weighting and zero-filling are carried out on the FIDs as normal, but after Fourier transformation with respect to $t_2$ to give a series of spectra $S(t_1, F_2)$ the data matrix is transposed to give a set of 'interferograms' $S(F_2, t_1)$. A second Fourier transformation, with respect to $t_1$, yields the 2D spectrum $S(F_1, F_2)$. Because most coherence transfer processes cause amplitude rather than phase modulation as a function of $t_1$, information on the signs of $F_1$ frequencies is usually missing; it can be recovered by making two measurements using different phase cycles. If the two sets of measurements are combined before Fourier transformation, converting amplitude modulation into phase modulation, absolute value mode presentation is generally required because individual signals will show a 'phase-twist' line shape; this is also the case with simple experiments which use pulsed field gradients for coherence transfer pathway selection. If the two data sets are recombined appropriately after the first Fourier transformation, both phase cycled and pulsed field gradient 2D experiments can be made to yield pure double absorption mode line shapes. Two common recombination schemes are the 'hypercomplex' method of States, Haberkorn and Ruben, and the time-proportional phase incrementation (TPPI) method of Marion and Wüthrich.

# Postprocessing

Many chemical questions can be answered with a simple spectrum, but to extract all the useful information from an NMR experiment it can require a wide range of data processing methods. Integration (usually after baseline correction) and peak picking allow the relative numbers of spins responsible for different multiplets to be found, and chemical shifts and scalar couplings to be measured. Better measures of signal intensity and line shape can be found by least-squares

fitting, if necessary in conjunction with line shape correction by reference deconvolution. Computers can also aid in the analysis of strongly coupled spectra and the spectra of dynamic systems.

# Zero-Filling

Data acquisition produces a set of $N$ complex points, sampled at equal time intervals $\Delta t$, which describe the nuclear FID as $2N$ independent pieces of information. Discrete Fourier transformation of these data will produce a spectrum of $N$ complex points, at frequency intervals $1/(N\Delta t)$. A final absorption mode spectrum will contain just $N$ independent pieces of information, only half the amount originally acquired; the remaining $N$ points form the dispersion mode spectrum. Full use of the experimental data can be achieved if $N$ complex zeroes are appended to it before Fourier transformation ('zero-filling'). Transforming $N$ data points plus $N$ zeroes generates a spectrum of $2N$ complex points, at frequency intervals $1/(2N\Delta t)$ Hz. The $2N$ real points are independent, and contain the same information as the $2N$ imaginary points: the real and imaginary data are correlated, so the total information content of the spectrum is unchanged. Zero-filling thus improves the digital resolution of the frequency spectrum twofold, as can be seen from Figure 1. Appending more than $N$ zeroes before transformation cannot increase the information content of the spectrum: the extra data points obtained simply interpolate between those produced by a single zero-filling.

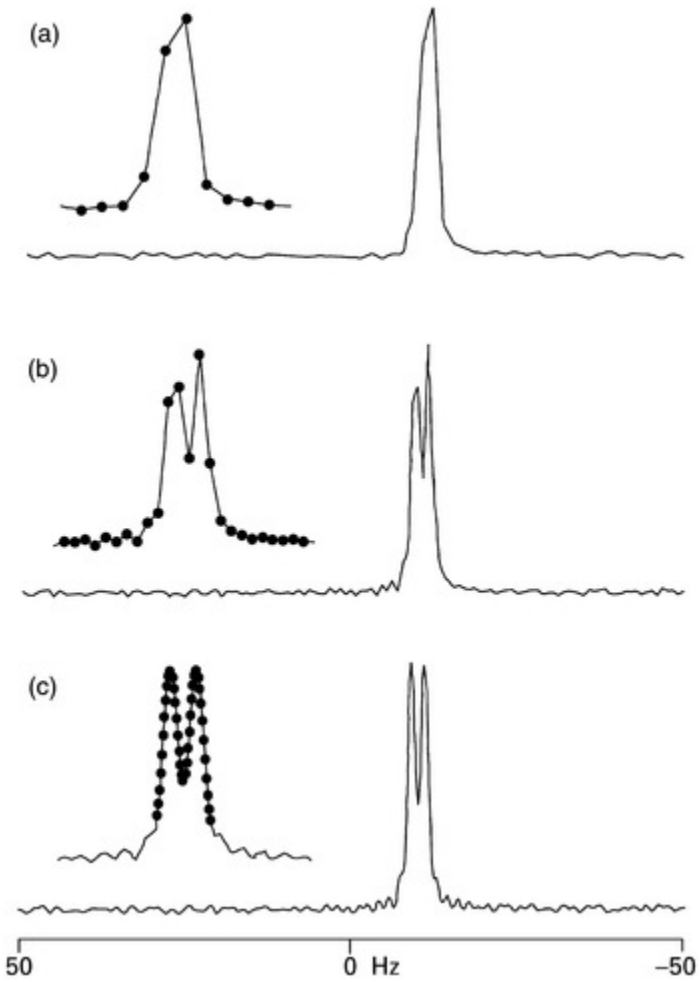


**Figure 1**

Spectra of a doublet with splitting 2 Hz, centred at −10 Hz, calculated for a 64 complex point FID with (a) no zero-filling; (b) one zero-filling, to 128 complex points; and (c) four zero-fillings, to 1024 complex points. The splitting becomes visible after one zero-filling; further zero-filling is equivalent to interpolation between the data points with a $\sin(x)/x$ function.

# Time-Domain Weighting

All experimental NMR signals decay, sooner or later: most well-designed experiments sample the signal until it has decayed close to zero. The later stages of the recorded data contain less signal, but the noise remains more or less constant. Recording for too short a time will lose valuable data; recording for too long will emphasize the noise relative to the signal. The widths of spectral lines depend on the rate of decay of the NMR signal: resolution can be improved if the natural decay of the signal is counteracted. Both issues can be addressed by weighting the time-domain signal before Fourier transformation. The choice of weighting function determines the compromise between resolution and signal-to-noise ratio in the resultant spectrum.

The signal-to-noise ratio of a spectrum can be optimized by multiplying the experimental signal by a weighting function which matches the experimental decay envelope: 'matched filtration'. The Fourier transform of an exponential decay with time constant $T$ is a Lorentzian line shape with a full width at half height of $1/(\pi T)$ Hz. Thus to obtain the best signal-to-noise ratio for a Lorentzian line of width $W$ Hz the experimental data should be multiplied by a decaying exponential of time constant $1/(\pi W)$ s. This gives a spectrum with optimum signal-to-noise ratio, at the expense of a doubling of the line width to $2W$ Hz. Where a spectrum with poor signal-to-noise ratio contains lines with a range of widths it can be helpful to try exponential weighting with several different time constants.

Time-domain weighting is equivalent to frequency-domain convolution. The convolution theorem states that the Fourier transform of the product of two functions $a(t)$ and $b(t)$ is the convolution of the two individual transforms $A(v)$ and $B(v)$:
[1]

$$\begin{aligned}\mathbf{FT}^{-}[a(t)\times b(t)] &\equiv \int_{-\infty}^{\infty} a(t)\times b(t)\ \exp(-2\pi i v t)\ \mathrm{d}t \\ &\equiv \int_{-\infty}^{\infty} A(v')\times B(v-v')\ \mathrm{d}v' \\ &\equiv \mathbf{FT}^{-}[a(t)]\otimes\mathbf{FT}^{-}[b(t)] \\ &\equiv A(v)\otimes B(v)\end{aligned}$$

where FT$^{-}$[ ] indicates Fourier transformation and ⊗, denotes convolution. Thus time-domain exponential weighting is equivalent to convolution, or smoothing, with a Lorentzian lineshape in the frequency domain. Matched filtration corresponds to smoothing the raw experimental spectrum with a function which matches the experimental line shape, as Figure 2 illustrates.

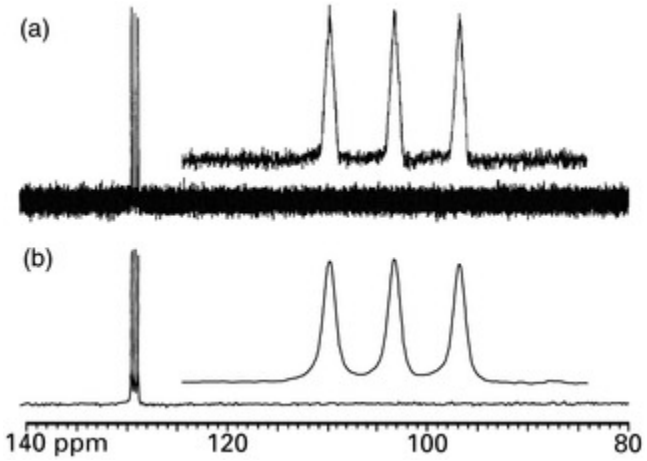


**Figure 2**
75.4 MHz spectra of the $^{13}$C triplet of deuteriobenzene in the ASTM (American Society for Testing and Materials) sensitivity test sample (60% deuteriobenzene/40% dioxane), with and without matched filtration. The unweighted spectrum (a) shows a signal-to-noise ratio of 18:1; the same data given an exponential multiplication with a time constant $1/(3\pi)$ s before Fourier

transformation, corresponding to a 3 Hz Lorentzian line broadening, show (b) a signal-to-noise ratio of 92:1. An acquisition time of 5.462 s was used, with a spectral width of 12000 Hz and one zero-filling. The insets show expansions of the triplet signal, illustrating the broadening of the lines and the smoothing of the noise caused by the exponential multiplication.

Time-domain weighting is also extensively used for resolution enhancement. Since this emphasizes the later part of the experimental signal, the noise energy is increased with respect to the signal energy, and resolution enhancement reduces the signal-to-noise ratio of the resultant spectrum. The aim of resolution enhancement is to reduce line widths without degrading the signal-to-noise ratio unacceptably. The natural decay of individual NMR signals is normally exponential, but countering this decay by multiplication with a rising exponential would lead to steeply rising noise. To stop the exponential rise in noise a further weighting using a function with a steeper decline is required. A weighting function $W(t)$ composed of a rising exponential with time constant $t_e$ and a falling Gaussian with time constant $t_g$

[2]

$$W(t) = \exp(+t/t_e)\ \exp[-(t/t_g)^2]$$

is generally the method of choice for resolution enhancement. $W(t)$ can also be written as a time-shifted Gaussian

[3]

$$W(t) = \exp(+t_s^2/t_g^2)\ \exp[-(t - t_s)^2/t_g^2]$$

where $t_s=t^2_g/(2t_e)$. If $t_e$ is equal to the decay constant of the experimental NMR signal, then multiplication by $W(t)$ before Fourier transformation converts a Lorentzian lineshape of width $1/(\pi t_e)$ Hz to a Gaussian of width $(2/\pi t_g)\sqrt{\log_e 2}$ Hz. Since spectra normally contain a range of line widths, it is usually necessary to experiment with $t_e$ and $t_g$ to find the best values for a given region of a spectrum. Because instrumental effects such as field inhomogeneity make experimental line shapes non-Lorentzian, resolution enhancement is best combined with reference deconvolution (see below). Figure 3 shows the application of Lorentz–Gauss resolution enhancement to a proton multiplet.

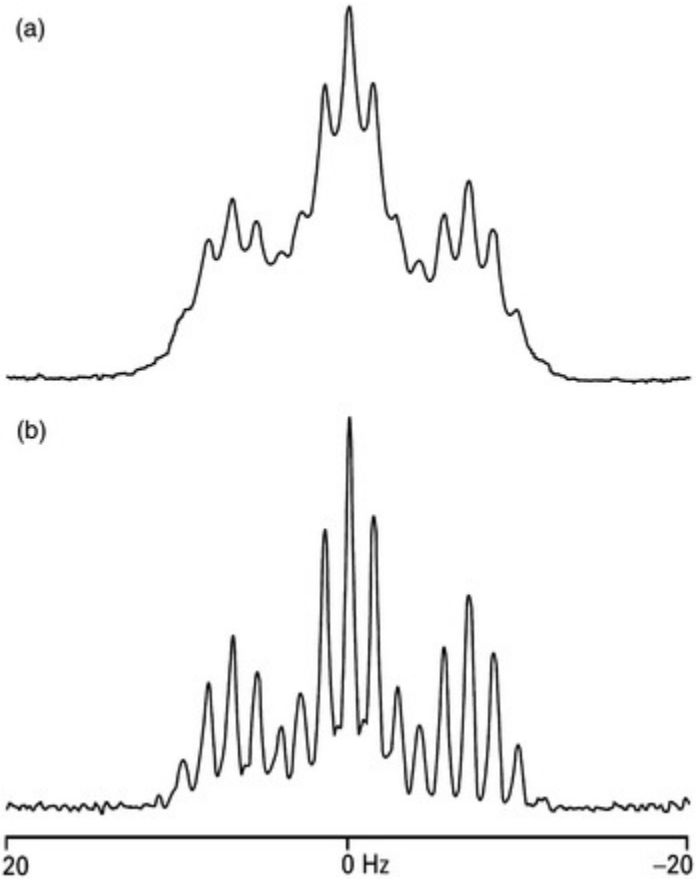


**Figure 3**

Expansions of the multiplet at 5.1 ppm in the 400 MHz proton spectrum of geraniol in deuteriomethanol: (a) raw spectrum; and (b) spectrum after Lorentz–Gauss conversion using rising exponential weighting with a time constant of $1/\pi$ s and Gaussian weighting with a time constant of 1 s.

Even where neither sensitivity nor resolution enhancement is sought, time-domain weighting is desirable where some NMR signal survives at the end of the sampled data. Such a truncated dataset is equivalent to the full, untruncated signal multiplied by a window function; the convolution theorem shows that the resultant spectrum will contain the true line shapes convoluted by a 'sinc' [$\sin(x)/x$] function, giving rise to 'wiggles' on either side of lines. Applying a weighting function $W(t)$, which brings the time-domain data smoothly to zero ('apodization'), can reduce or suppress such undesirable artefacts.

Most NMR spectra are presented in phase-sensitive mode, but this is not appropriate where the phases of signals vary rapidly or unpredictably with position, as in some magnetic resonance imaging and multidimensional NMR experiments. The modulus of a complex Lorentzian line shape shows a very broad base because of the contribution from the (imaginary) dispersion mode component. This can be suppressed if time-domain weighting is used to force the experimental signal into a form that is time-symmetric, for example using the function $W(t)$ above with a small $t_e$ ('pseudo-echo' weighting) or using a half sine-wave ('sine-bell' weighting). Although it is common to arrange for such weighting to leave the maximum of the weighted signal at the midpoint of the experimental data, this is neither necessary nor always desirable. Absolute value

mode presentation is the norm where phase cycling or pulsed field gradients are used to produce signal phase modulated as a function of $t_1$ in 2D NMR; it is to be avoided where possible because overlapping peaks are distorted by interference between their dispersion mode parts.

# Fourier Transformation

The classical frequency domain spectrum $S(\nu)$ is the Fourier transform of the FID $s(t)$:
[4]

$$S(\nu) \int_0^{t_a} s(t)\ \exp(-2\pi i \nu t)\ \mathrm{d}t$$

where the integration limits reflect the fact that the FID starts at time zero and is recorded for a time $t_a$. Practical spectrometers use digital technology, so the FID $s(t)$ is digitized at regular intervals $\Delta t$ to give a time series of $M$ points $s_k=s[(k-1)\Delta t]$, where $(M-1)\ \Delta t=t_a$, and a discrete Fourier transform (DFT) is carried out using the Cooley–Tukey FFT algorithm. The DFT of a time series of $N$ complex points with spacing $\Delta t$ generates a frequency spectrum which is a series of $N$ complex points with spacing $1/(N\Delta t)$ Hz:
[5]

$$S_n = \frac{1}{\sqrt{N}} \sum_0^{N-1} s_k\ \exp(-2\pi i k n/N)$$

where the frequency of the $n$th point is $(n-1)/(N\Delta t)$ Hz. Both the continuous and the discrete Fourier transform can use several different sign and normalization conventions; those given here are widely used, but others are equally valid. Spectrometers are almost invariably restricted by the FFT to Fourier transforming numbers of points $N$ which are powers of 2.
Discrete sampling in the time domain introduces an ambiguity into the frequency domain: signals at frequencies separated by multiples of $1/(\Delta t)$ Hz are indistinguishable, since their relative phases only change by multiples of $2\pi$ between sampling points. NMR signals which lie outside the range 0 to $(N-1)/(N\Delta t)$ will be 'aliased' by adding or subtracting the spectral width $1/\Delta t$ until they lie within this window, as seen in Figure 4. Since the signals that emerge from the spectrometer receiver may have positive or negative frequencies, the output of the DFT needs to be rotated by $N/2$ points [$1/(2\Delta t)$ Hz] so that the digitized spectrum covers the range $-1/(2\Delta t)$ to $(N/2-1)/(N\Delta t)$ Hz. By convention, the resultant spectrum is plotted with frequency increasing from right to left, so that the most shielded nuclei (those with lowest chemical shift) lie at the right. Some older spectrometers sample the real and imaginary receiver channels alternately rather than simultaneously, using a real rather than a complex Fourier transform; signals outside the spectral width then 'fold' back into the spectrum by reflection about the frequency limits $\pm 1/(2\Delta t)$.

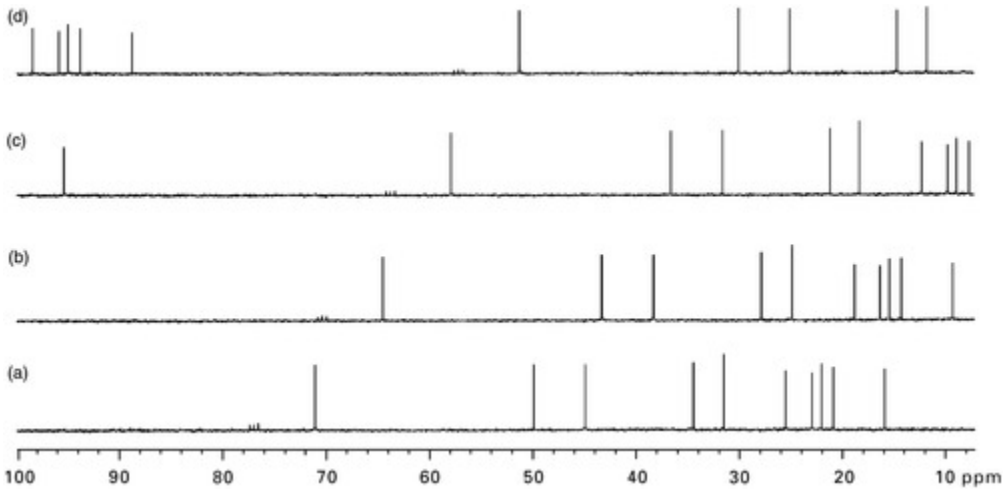


**Figure 4**

75.4 MHz proton-decoupled $^{13}C$ spectra of 30% menthol in deuteriochloroform, (a) recorded with all signals within the spectral window, and (b)–(d) with the transmitter displaced to high field in 500 Hz steps. Spectra (c) and (d) show the aliasing of the high-field signals to reappear at the low-field end of the spectrum.

# Phasing

The measurement of NMR data must wait until the radiofrequency pulse and its after-effects have died away, which can take several tens of microseconds; analogue or digital filtration also delays the arrival of the NMR signal at the receiver output. The effect on an NMR signal of a delay time $\delta$ is to add a phase shift $\exp(2\pi i\nu\delta)$ to a signal of frequency $\nu$, causing the signal phase to vary linearly across the spectrum. In addition, the relative phases of the receiver reference signals and the transmitter pulse are arbitrary, so both a zeroth- and a first-order phase correction are needed to bring all signals into absorption mode. The phased spectrum $S_p(\nu)$ can be written as:

[6]

$$S_p(\nu) = S(\nu)\ \exp[-i(\phi_0 + \phi_1\ \nu\Delta t)]$$

where the zeroth-order and first-order phase shifts $\phi_0$ and $\phi_1$ are normally determined either automatically, or by the spectrometer operator using an interactive display. Figure 5 shows a typical spectrum before and after phasing.

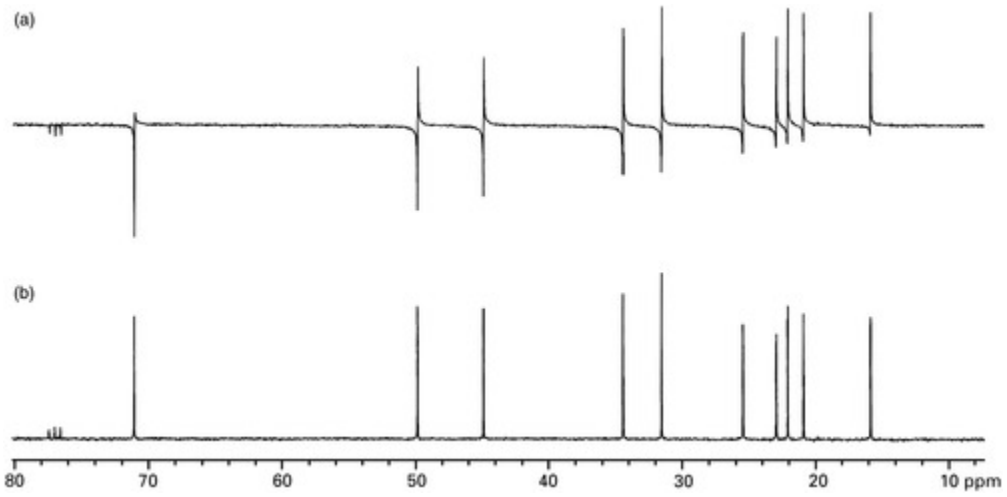


**Figure 5**

75.4MHz proton-decoupled $^{13}C$ Spectra of 30% menthol in deuteriochloroform, (a) before and (b) after zeroth- and first-order phase correction.

First-order phase correction has one insidious effect, baseline distortion. A frequency-dependent phase shift cannot make up for the data that were lost during $\delta$; the baseline error is just the DFT of the missing data. However, provided $\delta$ is small compared to $\Delta t$, this baseline curvature can easily be corrected during postprocessing of the spectrum. In 2D NMR, there will be different time delays $\delta_1$ and $\delta_2$ in the two time dimensions; phasing again is normally carried out using an interactive display.

# Linear Prediction

A FID and its Fourier transform contain exactly the same information, but sometimes this is insufficient to give a readily interpretable spectrum. Where there is adequate internal evidence within the FID, it may be possible to extrapolate the NMR signals forwards and/or backwards in time to synthesize missing data and hence create a time-domain signal that transforms to a clearer spectrum. The two commonest uses of such an extrapolation are backwards in time to replace data lost during the time $\delta$, and forwards in time to improve resolution. A typical digitized experimental FID contains a series of $n$ exponentially damped, complex signals, plus a background of random noise. The NMR signal can be written as:

[7]

$$s_k = \sum_{j=1}^{n} \alpha_j \beta_j^k$$

where the complex number $\alpha_j = A_j \exp(i\phi_j)$ defines the phase $\phi_j$ and amplitude $A_j$ of the $j$th signal, and $\beta_j = \exp(2\pi i \Delta t \nu_j) \exp(-\Delta t / T_j)$ is determined by the frequency $\nu_j$ and decay constant $T_j$. The contribution made by component $j$ to point $s_k$ is just the contribution to $s_{k-1}$ multiplied by $\beta_j$. Linear prediction (LP) algorithms[1,2] take a FID of $M$ points and fit this time series with a set of $m$ complex coefficients $a_j$

[8]

$$s_k = \sum_{j=1}^{m} a_j s_{k-j}$$

so that point $k$ is expressed as a linear combination of the previous $m$ points (forward prediction), or of the subsequent $m$ points (backward prediction)

[9]

$$s_k = \sum_{j=1}^{m} a_j s_{k+j}$$

A variety of algorithms exist for finding the coefficients $a_j$ and multipliers $\beta_j$, with which the experimental data can be extrapolated forwards or backwards; all share some common problems. Linear prediction has difficulty distinguishing between positive and negative decay constants $T_j$, and so is best suited to time series in which all the decay constants are either positive or negative, allowing spurious $\beta$ values to be rejected. The number $m$ of coefficients $a_j$ to be used has to be decided by the experimenter: too few, and peaks will be missed: too many and noise will be treated as signal. NonLorentzian line shapes make exponential damping a poor approximation, increasing the number of coefficients needed.

Linear prediction is (despite its name) a nonlinear method and can produce very misleading results, but with care it can greatly ease interpretation of poorly digitized spectra. It is particularly useful in 2D NMR, where signals are routinely truncated in the $t_1$ dimension. Several other processing methods designed to supplement measured data with inter- and extrapolated data points are described below.

# Maximum Entropy Reconstruction and the Filter Diagonalization Method

Although linear prediction can be used to extract spectral data directly from a FID ('parametric LP'), it is commonly used to extrapolate the experimental time-domain data, which are then weighted and transformed as normal. Maximum entropy reconstruction[3,4], in contrast, seeks to fit the experimental FID with a model function that contains the minimum amount of information consistent with fitting experiment to within the estimated noise level. The criterion of minimum information corresponds to the maximum Shannon informational entropy $S(p)$, which for a probability distribution $p$ is defined as

[10]

$$S(p) = -\sum_{n} p_n \log_e p_n$$

Maximum entropy methods have generated considerable controversy; they have been described (a little unkindly) as generating more heat than light. Their results can show spectacular improvements in signal-to-noise ratio, but this should not be confused with sensitivity of detection. Maximum entropy methods successfully pick out those signals that are above a defined threshold, but miss those below it; the signal amplitude estimates produced are comparable to those obtainable by simply fitting a model line shape to the Fourier transform spectrum. Thus for well-sampled experimental data the advantages of maximum entropy

methods are largely cosmetic, and come at a high computational cost. Where such methods can be very valuable is with data sets that are damaged, incomplete, not sampled at uniform time intervals, or require deconvolution.

One alternative method for processing truncated datasets is the filter diagonalisation method (FDM)[5,6], which uses linear algebra methods to estimate the frequencies, linewidths and intensities of Lorentzian lines in a spectrum. For experimental data with good signal-to-noise ratio the FDM can give very large improvements in resolution.

# Non-Uniform Sampling

So far, the nonlinear processing methods described have been applied to conventional experimental data, with data points sampled at a constant rate. While such data still form the backbone of NMR applications, and almost all FIDs are still sampled uniformly, there can be significant advantages to other sampling strategies, for example in the indirect dimensions of multidimensional experiments. One of the primary reasons for multidimensional NMR is to gain resolution, separating signals according to several different parameters, e.g. the chemical shifts of different isotopes. Multidimensional spectra are therefore significantly sparser than 1D spectra, with the signals occupying a much smaller proportion of the spectral range. One consequence is that as the number of dimensions increases the signal parameters become more and more over-determined by the (fully sampled) experimental data: far more data points are acquired than are needed to characterize the information content of the multidimensional spectrum. It is thus possible, with suitable data processing, to obtain the desired information with far fewer individual measurements of FIDs, by using non-uniform sampling in the indirect dimensions.

A wide variety of methods have been proposed that allow complete spectra to be obtained from partial multidimensional datasets. The choice of method depends among other things on the nature of the 1D spectrum, the resolution needed, the degree of overlap anticipated, the data processing algorithms available, and the extent to which spectral artefacts can be tolerated.

Where the 1D spectrum is relatively sparse, with little overlap between multiplets, Hadamard encoding using prior knowledge of the 1D spectrum allows a multidimensional spectrum to be generated using a very small number of distinct acquisitions, tailored to the frequencies required[7,8]. In Hadamard spectroscopy, the evolution time in the indirect dimension is replaced by phase-modulated multi-site frequency selective excitation. The minimum number of acquisitions needed is equal to the number of different chemical shifts for which information is required, rather than the number needed to digitize the full range of chemical shifts studied. Recent applications of Hadamard-encoded NMR spectroscopy have shown increased time resolution resulting from very fast spectral acquisition, while maintaining the spectroscopic resolution associated with 2D spectra. Specific examples include measurement of proton-deuteron exchange kinetics, and frequency-resolved diffusion and relaxation measurements of chemically similar species in mixtures. The use of Hadamard-encoded filters to measure methyl–methyl cross-relaxation has also been demonstrated.

Where spectra of dimension greater than 2 are required, conventional acquisition is extremely time-consuming because of the need to record a large number of increments of each evolution time. A particularly simple scheme for non-uniform sampling of the indirect dimensions is to record FIDs only for certain fixed ratios of evolution times, incrementing the two times in sympathy[9,10]. Each set of radially-sampled data, along a fixed angle in time space, Fourier transforms to give an integral projection of the signal intensity along the same angle in frequency space. The advantage to this type of choice of which data points to sample is that the full spectrum can then be deduced using projection-reconstruction algorithms, common in imaging methods such as computed tomography and MRI, in a process analogous to deducing the three dimensional shape of a person's head by examining a series of silhouettes drawn at different angles. Practical problems include the need to deal with ambiguities introduced by signals overlapping in projections, the effects of noise, and artefacts introduced by the reconstruction. Projection-reconstruction methods can allow the analysis of very high dimensional spectral data ("hyperdimensional spectroscopy").

The most general case is where FIDs are sampled for only a small fraction of the full range of values of evolution periods. Provided that, as noted above, sufficient measurements are made to define unambiguously the frequencies of all the signals in a multidimensional spectrum, very large time savings become possible. As the number of dimensions increases, so the proportion of the potential data space that needs to be acquired reduces, keeping the overall experiment duration required manageable. The first problem is to decide on a rational basis for choosing which data to acquire, i.e. which values to use for sampling the different evolution periods[4]. Although in principle it is possible to use any values for these, in practice values are usually constrained to lie on a Nyquist sampling grid – in other words measurements are made for a subset of the evolution time values that would be used in a complete conventional acquisition. Common choices include pseudorandom sampling and Poisson Gap scheduling. Once data have been acquired, a variety of different algorithms can be used to construct spectra. Choice of method depends amongst other things on the sparsity or otherwise of the spectrum, the signal-to-noise ratio, and prior knowledge. Compressed sensing methods such as iterative soft thresholding (IST) and iterative re-weighted least squares (IRLS) have proved popular and effective[11,12]; maximum entropy is also effective[4] but has heavy computational demands; and multiway decomposition methods, which in nondegenerate systems can provide unique decompositions for datasets of dimension three or greater are also effective and have proved popular in protein NMR[13,14].

# Postprocessing Techniques

The simplest and most widely used form of postprocessing is integration of the signal intensity. The integral of a resonance is proportional to the number of spins contributing to it; thus the relative numbers of nuclei in different chemical groupings can be found by comparing the integrals of their signals under appropriate experimental conditions. Accurate integration almost always requires operator intervention to carry out baseline correction, varying according to need from simple offset and slope correction through to the subtraction of a baseline calculated by spline or polynomial fit to operator-defined regions of empty baseline.

A second common example of postprocessing is the listing of signal heights and positions ('peak picking'), for the measurement of chemical shifts and coupling constants, or as the first step in the extraction of parameters such as relaxation times, rate constants or diffusion coefficients. In principle, the integral of a signal should give a better estimate of signal amplitude than peak height, being independent of line shape; in practice, baseline errors and signal overlap mean that peak height measurements are usually preferred for intensity comparisons between corresponding signals in different spectra.

Where signals overlap, neither integration nor peak picking gives accurate signal intensities. Here iterative fitting can be used to decompose the experimental spectrum into contributions from individual lines, typically assumed to have Lorentzian or Gaussian shapes; the positions, amplitudes and widths of the theoretical line shapes are varied to minimize the sum of the squares of the differences between the experimental and calculated spectra. Such least-squares fitting is easily perturbed by instrumental distortion of the line shapes, for example as a result of static field inhomogeneity, so the best results require either great care with shimming or some form of compensation for instrumental line shape contributions.

One effective way to compensate for many instrumental sources of error is reference deconvolution[15]. Since most instrumental errors (e.g. static field inhomogeneity and magnetic field instability) affect all signals equally, multiplying the experimental FID by the complex ratio of the theoretical and experimental signals for a reference resonance leads to a spectrum in which such errors have been corrected. This technique can be used to ensure that all lines are basically Lorentzian, and also to enforce strict comparability between different spectra in a series, for example to correct $t_1$ noise in multidimensional NMR.

Many other data processing techniques are used to extract useful information from experimental NMR spectra. Signal intensities compiled by peak picking may be fitted to an exponential or Gaussian function, as in the determination of relaxation times and in pulsed field gradient spin-echo measurements of diffusion coefficients. The latter experiment can be extended to the construction of a pseudo-2D spectrum in which signals are dispersed according to chemical shift in one dimension and diffusion coefficient in the other (see "Diffusion-ordered spectroscopy" below). In many spectra the extraction of chemical shift and coupling constant values is hindered by second-order effects ('strong coupling'); the analysis of strongly coupled spectra is most effectively carried out using quantum-mechanical simulation. This can be partially automated in favourable cases, the experimental spectrum being used as a target for least-squares fitting in which the variable parameters are the chemical shifts and coupling constants rather than simply the signal positions, amplitudes and widths. The extraction of kinetic parameters from exchange-broadened band shapes ('lineshape analysis') can be similarly automated. An important application of automated post-processing of signal intensities, sometimes after the agglomeration of intensity from individual spectral regions ("binning"), is in metabolomics/metabonomics[16].

# Ultrafast multi-dimensional methods

Whether full or partial sampling of the indirect dimensions of a multidimensional NMR experiment is used, the experimental time demands are considerable. One alternative is to use

multiplexing to encode information on evolution time into a spatial dimension, allowing all the data needed for a complete two-dimensional spectrum to be recorded in a single scan[17,18]. Some changes in data processing are required, but these are normally based on conventional Fourier methods. Derived from echo planar methods in magnetic resonance imaging, this approach makes some sacrifices in sensitivity and NMR resolution but allows a complete spectrum to be recorded in as little as a second, enabling 2D methods to be used to study phenomena such as refolding processes and hydrogen-deuterium exchange in proteins to be studied. In combination with hyperpolarization, ultrafast methods allow 2D correlation spectra of proteins to be recorded at submicromolar concentrations[19].

# Diffusion-ordered spectroscopy

Diffusion-ordered spectroscopy (DOSY) separates the NMR signals of different species according to their diffusion coefficient. In a pulsed field gradient spin echo experiment, the signals from a given substance in a mixture normally decay as a function of gradient amplitude at the same rate, determined by its diffusion coefficient, allowing a bilinear NMR data set recording signal variation as a function of chemical shift and of gradient amplitude to be assembled. In the simplest DOSY analysis, fitting the signal decay for each peak to the appropriate form of the Stejskal-Tanner equation yields a diffusion coefficient for each signal. These data can then be used to synthesise a 2D spectrum in which the normal 1D spectrum is extended into a second, diffusion dimension, giving a contour plot of signal strength against chemical shift diffusion coefficient. In 2D DOSY the initial diffusion-weighted spectra are 1D; adding diffusion weighting to 2D experiments such as COSY, NOESY or HMQC gives the corresponding 3D DOSY spectra. The difficulty of extracting accurate and meaningful diffusion parameters from overlapping signal decays sets strict limits on the diffusion resolution. To identify small differences reliably, it is necessary that systematic errors be reduced to an absolute minimum. The challenge of designing successful high-resolution DOSY techniques is to combine experimental methods that minimize such errors with data analysis procedures which compensate as accurately as possible for the errors that remain. Applications of DOSY NMR include identification of the components and impurities in complex mixtures, such as body fluids, or reaction mixtures, and technical or commercial products, for example, comprising polymers or surfactants.

Data processing is a key component of DOSY NMR. Both univariate and multivariate methods have been used, but all have difficulties when applied to real-world cases. The big challenge is that signals overlapping in the NMR dimension(s) of the experiment can be very difficult to disentangle in the diffusion dimension, because of the intractable problem of separating different superimposed exponential signal decays. A great deal of ingenuity has gone into designing experiments to minimize signal overlap, and data processing methods that allow the maximum interpretable information to be extracted from the experimental measurements. The former include pure shift methods, in which the effects of homonculear scalar couplings are suppressed in order to yield spectra with just a single signal for each distinct chemical shift[20]. Examples of the latter include single-channel methods such as multiexponential fitting (SPLMOD) and continuous diffusion spectrum fitting (CONTIN), and two multivariate such as the direct

exponential curve resolution algorithm (DECRA), multivariate curve resolution (MCR), and the CORE and SCORE methods.

# Statistical methods

Conventional multidimensional NMR methods generate spectra that display relationships between signals that have a direct physical basis, for example a scalar coupling (e.g. in COSY or HSQC) or a through-space coupling (e.g. in NOESY or ROESY). A more general approach to defining relationships is to plot the covariance between signals in different dimensions[21,22]. This can be used as an alternative to conventional Fourier processing, e.g. in COSY or TOCSY, to generate alternative representations of the data ("covariance processing"), and can be extended to allow construction of spectra that cannot be obtained directly by conventional means, for example correlating nuclei that have no direct coupling ("indirect covariance processing"). In combination with pure shift methods, covariance processing can yield 2D correlation spectra in which all homonuclear coupling structure has been suppressed[23].

The implementation of an approach known as statistical total correlation spectroscopy (STOCSY) for aiding the identification of the various NMR peaks from a given molecule of potential biomarker interest in complex biochemical mixtures, has been a recent development[24]. STOCSY takes advantage of the multi-collinearity of the intensity variables in a set of 1D spectra (e.g. $^{1}$H NMR spectra) to generate a pseudo-2D NMR spectrum that displays the correlations among the intensities of the various peaks across the whole dataset. This method is not limited to the usual connectivities that are deducible from more standard 2D NMR spectroscopic methods such as COSY or TOCSY, and hence is applicable to singlets and to peaks from functional groups that are far apart in a molecule.

# See also

Fourier Transformation and Sampling Theory, Laboratory Information Management Systems (LIMS), NMR Principles, NMR Pulse Sequences, NMR Spectrometers, Two-Dimensional NMR.

# Further Reading


Freeman, R., *A Handbook of Nuclear Magnetic Resonance (2/e)*. (1987) Longman, Harlow, UK .
Freeman, R., *Spin Choreography*. (1997) Spectrum, Oxford, UK .
Hoch, J.C.; Stern, A.S., *NMR Data Processing*. (1996) Wiley-Liss, New York .
Kupce, E.; Nishida, T.; Freeman, R., Hadamard NMR spectroscopy, *Progress in Nuclear Magnetic Resonance Spectroscopy* **42** (2003) 95–122.
Lindon, J.C.; Ferrige, A.G., Digitisation and data processing in Fourier transform NMR, *Progress in NMR Spectroscopy* **14** (1980) 27–66.
Rutledge DN (ed.) (1996) Signal treatment and signal analysis in NMR. In: vol. 18 of *Data Handling in Science and Technology*. Amsterdam: Elsevier.

*This article is reproduced from the previous edition, volume 2, pp 1514–1521, © 1999, Elsevier Ltd., with revisions made by the Editor and by the author.

Change History: G A Morris corrected copy-editing errors throughout the Word file supplied and added new material to almost every section. Only the section Phasing and the Figures remain unchanged; captions to the latter have been corrected.

# References

[1] Abidgaard, F.; Gesmar, H.; Led, J.J., *J. Magn. Reson.*, **1988**, *79*, 78-89.
[2] Led, J.J.; Gesmar, H., Fourier Transform and Linear Prediction Methods. In *Multidimensional NMR Methods for the Solution State; Morris, G.A. and Emsley, J.W.* Eds.; John Wiley & Sons Ltd: Chichester, UK, 2010; Chapter 10, pp 131-141.
[3] Mobli, M.; Maciejewski, M.W.; Gryk, M.R.; Hoch, J. C., *J. Biomol. NMR*, **2007**, *39*, 133-139.
[4] Hoch, J.C.; Mobli, M., Maximum Entropy Methods in Multidimensional NMR. In *Multidimensional NMR Methods for the Solution State; Morris, G.A. and Emsley, J.W.* Eds.; John Wiley & Sons Ltd: Chichester, UK, 2010; Chapter 8, pp 107-117.
[5] Mandelshtam, V., *Prog. NMR Spectrosc.*, **2001**, *38*, 159-196.
[6] Shaka, A.J.; Mandelshtam, V., Filter Diagonalization Methods for Time-Domain Signals. In *Multidimensional NMR Methods for the Solution State; Morris, G.A. and Emsley, J.W.* Eds.; John Wiley & Sons Ltd: Chichester, UK, 2010; Chapter 9, pp 119-130.
[7] Freeman, R.; Kupče, Ē., *J. Biol. NMR* **2003**, *27*, 101-113.
[8] Freeman, R.; Kupče, Ē., Fast Multidimensional NMR by Hadamard Spectroscopy. In *Multidimensional NMR Methods for the Solution State; Morris, G.A. and Emsley, J.W.* Eds.; John Wiley & Sons Ltd: Chichester, UK, 2010; Chapter 4, pp 61-71.
[9] Kupče, Ē.; Freeman, R., *Prog. NMR Spectrosc.*, **2008**, *52*, 22–30.
[10] Freeman, R.; Kupče, Ē., Multidimensional NMR by Projection-Reconstruction. In *Multidimensional NMR Methods for the Solution State; Morris, G.A. and Emsley, J.W.* Eds.; John Wiley & Sons Ltd: Chichester, UK, 2010; Chapter 5, pp 73-83.
[11] Hyberts, S.G.; Milbradt, A.G.; Wagner, A.B.; Arthanari, H.; Wagner, G., *J. Biol. NMR* **2012**, *52*, 315-327.
[12] Misiak, M.; Koźmiński, W.; Chmurskia, K.; Kazimierczuk, K., *Magn. Reson. Chem.* **2013**, *51*, 110-115.
[13] Orekhov, V.Yu.; Ibraghimov, I.; Billeter, M., *J. Biomol. NMR*, **2003**, *27*, 165-173.
[14] Billeter, M.; Staykova, D.K., Rapid Multidimensional NMR: Decomposition Methods and their Applications. In *Multidimensional NMR Methods for the Solution State; Morris, G.A. and Emsley, J.W.* Eds.; John Wiley & Sons Ltd: Chichester, UK, 2010; Chapter 6, pp 85-95.
[15] Morris, G.A.; Barjat, H.; Horne, T.J., *Prog. NMR Spectrosc.*, **1997**, *31*, 197–257.
[16] Ebbels, T.M.D.; Cavill, R., *Prog. NMR Spectrosc.*, **2009**, *55*, 361-374.
[17] Gal, M.; Frydman, L., Ultrafast Multidimensional NMR: Principles and Practice of Single-Scan Methods. In *Multidimensional NMR Methods for the Solution State; Morris, G.A. and Emsley, J.W.* Eds.; John Wiley & Sons Ltd: Chichester, UK, 2010; Chapter 3, pp 43-60.
[18] Tal, A.; Frydman, L., *Prog. NMR Spectrosc.*, **2010**, *57*, 241-292.
[19] Frydman, L.; Blazina, D., *Nature Physics*, **2007**, *3*, 415–419.
[20] Nilsson, M.; Morris, G.A., *Chem. Commun.*, **2007**, 933-935
[21] Trbovic, N.; Smirnov, S.; Zhang, F.; Brüschweiler, R., *J. Magn. Reson.*, **2004**, *171,* 277–283.

[22] Snyder, D.A.; Brüschweiler, R., Multidimensional Correlation Spectroscopy by Covariance NMR. In *Multidimensional NMR Methods for the Solution State; Morris, G.A. and Emsley, J.W.* Eds.; John Wiley & Sons Ltd: Chichester, UK, 2010; Chapter 7, pp 97-105.
[23] Aguilar, J.A.; Colbourne, A.A.; Cassani, J.; Nilsson, M.; Morris, G.A., *Angew. Chem. Int. Ed.*, **2012**, *51*, 6460-6463.
[24] Cloarec, O.; Dumas, M.E.; Craig, A.; Barton, R.H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J.C.; Holmes, E.; Nicholson, J. , *Anal. Chem.*, **2005**, *77*, 1282–1289.