Journal of Magnetic Resonance 214 (2012) 15–21

Contents lists available at SciVerse ScienceDirect

Journal of Magnetic Resonance

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

Phase-sensitive spectral estimation by the hybrid filter diagonalization method

Hasan Celik, Clark D. Ridge [1], A.J. Shaka *

Department of Chemistry, University of California, Irvine, CA 92697-2025, USA

ARTICLE INFO

Article history:
Received 27 July 2011
Revised 19 September 2011
Available online 29 September 2011

Keywords:
Filter diagonalization
FDM
Spectral estimation
Lorentzian-to-Gaussian
Phase-sensitive NMR spectra
Resolution enhancement

ABSTRACT

A more robust way to obtain a high-resolution multidimensional NMR spectrum from limited data sets is described. The Filter Diagonalization Method (FDM) is used to analyze phase-modulated data and cast the spectrum in terms of phase-sensitive Lorentzian "phase-twist" peaks. These spectra are then used to obtain absorption-mode phase-sensitive spectra. In contrast to earlier implementations of multidimensional FDM, the absolute phase of the data need not be known beforehand, and linear phase corrections in each frequency dimension are possible, if they are required. Regularization is employed to improve the conditioning of the linear algebra problems that must be solved to obtain the spectral estimate. While regularization smoothes away noise and small peaks, a hybrid method allows the true noise floor to be correctly represented in the final result. Line shape transformation to a Gaussian-like shape improves the clarity of the spectra, and is achieved by a conventional Lorentzian-to-Gaussian transformation in the time-domain, after inverse Fourier transformation of the FDM spectra. The results obtained highlight the danger of *not* using proper phase-sensitive line shapes in the spectral estimate. The advantages of the new method for the spectral estimate are the following: (i) the spectrum can be phased by conventional means after it is obtained; (ii) there is a true and accurate noise floor; and (iii) there is some indication of the quality of fit in each local region of the spectrum. The method is illustrated with 2D NMR data for the first time, but is applicable to *n*-dimensional data without any restriction on the number of time/frequency dimensions.

1. Introduction

The useful resolution of multidimensional (*n*D) NMR spectra depends on total acquisition time, line shape quality, spectral crowding, and noise level. First, the acquisition time AT_k in each of the n time dimensions $k = 1, \ldots, n$ imposes a minimum line width in the corresponding frequency dimension by the time–frequency uncertainty principle in the case of a Fourier transform (FT) spectrum. Secondly, discerning partially overlapping peaks in a contour plot, to analyze complex spectra and make assignments, is facilitated by the line shape in each dimension, a Gaussian line shape having the distinct advantage of clean elliptical contours that guide the eye to discern the true number of signals present. Finally, the noise level in the data is important to assess the significance of a possible peak, as features should exceed the noise by an acceptable margin to be included with confidence in a peak list to be used in subsequent distance constraints. "False positive" peaks can compromise structure calculations and can be time-consuming to identify and rectify.

The Filter Diagonalization Method (FDM) [1,2] has been used to analyze NMR data [3,4] and especially to provide an efficient way to handle severely truncated multidimensional signals [5–9] where the intrinsic resolving power of the high-dimensional data is high but the achievable digital resolution is usually limiting. The goal of FDM is to bypass the time–frequency uncertainty principle and achieve narrower peaks and much finer frequency resolution than the approximate limit $(AT_n)^{-1}$ of FT spectra. While FDM, in which the *n*D frequency data is expressed as *n*D Lorentzian peaks, is effective as a parametric multidimensional method for extremely large data sets, there are certainly details of the spectral construction itself that can be improved. The parameters are obtained by the solution of a generalized eigenvalue problem through eigenvalues and eigenvectors that describe Lorentzian peaks. The handling of the time-domain data this way has recently been analyzed in more detail [10] and, while capable of improvement, is still basically adequate. However, the follow up to obtain the spectrum itself can certainly be improved, especially for the case of somewhat noisy data. In such cases regularization [5] is used to improve the conditioning of the eigenvalue problem, and this regularization smoothes and broadens weak peaks and/or noise into the baseline, giving a false impression of the true noise level. A solution for better noise representation has recently been shown by a hybrid method, in which FDM was used to pick out and enhance the stronger peaks that the Lorentzian model fit well, and the FT was used to

* Corresponding author. Address: 1106 Natural Sciences 2, University of California, Irvine, CA 92697-2025, USA. Fax: +1 949 824 9920.
E-mail addresses: hcelik@uci.edu (H. Celik), clarkridge@gmail.com (C.D. Ridge), ajshaka@uci.edu (A.J. Shaka).
[1] Present address: Chemistry Department, University of Miami, Coral Gables, FL 33146-0431, USA.

1090-7807/$ - see front matter © 2011 Published by Elsevier Inc.
doi:10.1016/j.jmr.2011.09.044

obtain the residual, which was added to the idealized or *ersatz* FDM (EFDM) spectrum to create a so-called *hybrid* FDM (HFDM) spectrum [11]. We expand that treatment here to include multidimensional, true phase-sensitive FDM spectra that can be displayed, plotted, contoured and phased just like conventional FT spectra, but with higher resolution in dimensions that have too few time-domain data points. The approach here is numerically superior to our earlier treatment of HFDM, produces reliable spectra that have a more realistic noise floor than in our earlier work with multidimensional data, and can correctly handle peaks that have indeterminate phase, as in HMBC [12] spectra, for example.

2. Theory

As *n*D FDM spectra are always a linear combination of complex-valued *n*D peaks, the way to obtain a phase-sensitive absorption-mode spectrum and to make a Lorentzian-to-Gaussian transformation differs somewhat from the approach taken in *n*D FT spectra. We will clarify the key differences in the following sections. First, practical issues related to the FDM line shapes and the kind of time-domain data that is optimum will be covered. The new approach will be contrasted with more aggressive ways to construct the spectrum, ways that work in some circumstances but that can fail exactly when it is most critical to obtain better resolution, *i.e.* in cases of partial overlap. Next, the question of weighting the data to obtain more Gaussian line shapes will be covered, followed by a practical step-by-step procedure for obtaining the *n*D phase-sensitive hybrid spectrum. Finally, the question of how to handle the noise that is inevitably present in the spectra will be broached. Numerical examples show that the hybrid method can handle poor choices of the regularization parameter, and also spectra requiring strong linear phase corrections. Practical examples in the experimental section will illustrate the typical gains in resolution for 2D spectra. For 3D and higher spectra the gains are even more dramatic.

2.1. Multidimensional phase-sensitive FDM spectra

Two-dimensional data is the simplest case to consider. As a parametric method, FDM attempts to characterize a 2D spectrum as a combination of 2D peaks. The fewer the number of signal peaks in the data, the better the performance can be. Hence, FDM is best adapted to purely phase-modulated data sets where the signal can be written

$$C^{(P/N)}(n_1\tau_1, n_2\tau_2) \equiv c^{(P/N)}_{n_1 n_2} = \sum_{k_1=1}^{K_1}\sum_{k_2=1}^{K_2} d_{k_1k_2} e^{2\pi i f^{(P/N)}_{k_1} n_1\tau_1} e^{-n_1\tau_1/T_2^{(k_1)}} e^{2\pi i f_{k_2} n_2\tau_2} e^{-n_2\tau_2/T_2^{(k_2)}}, \quad (1)$$

where $d_{k_1k_2}$ is a complex-valued amplitude and $f^{(P)}_{k_1} \approx -f^{(N)}_{k_1}$. The superscripts *P*/*N* refer to "positive" or "negative" type data, as judged by the orientation of the frequency axis in F_1. The algebraic sign of the signal frequency in F_1 depends on whether *P*-type (+), or *N*-type (−) data is collected. As experimental data contains noise, and the *P*- and *N*-data are processed in separate FDM calculations, there need be no exact symmetry in F_1 between the two FDM spectra with respect to either peak positions or phases. Nevertheless, with low noise levels the spectra show approximate symmetry. Note that while *P*/*N* data results naturally in the case of gradient selection [13] of coherence transfer, in the alternative case of States-TPPI acquisition [14], the correct linear combination of cos ±*i* sin data sets would need to be made prior to FDM analysis. Analyzing either the cos or sin data sets themselves would require fitting twice as many peaks in each and, as we have as yet found no way to capitalize numerically on the known symmetry, would therefore require more experimental data to achieve comparable quality to that of pure phase-modulated data.

Simple 2DFT phase-sensitive spectra of *P*/*N* data are rare, as the 2DFT results in so-called "phase-twist" [15] line shapes, in which absorption- and dispersion-mode contributions are inextricably mixed together even when the data is correctly phased. It is difficult to discern peak positions and maxima in these complex spectra. However, it is the phase-twist spectrum that is the direct result of the FDM calculation, rather than the desired pure absorption spectrum. Thus, it must be decided how to follow up the FDM analysis to obtain a more useful spectral presentation.

Conventional processing software combines the phase-modulated *P*/*N*-data, if present, to obtain sin/cos amplitude-modulated data, and then proceeds as in Ref. [14] to avoid phase-twist line shapes. A parametric method like FDM, however, allows one to simply discard the dispersion-mode contribution to each individual peak entry, and thereby obtain a kind of pseudo-absorption spectrum from purely phase-modulated data, for example a single *N*-type data set [16]. Once the dispersion-mode has been deleted, however, it is no longer possible to phase the spectrum, either: the real and imaginary parts of the complex amplitude, $d_{k_1k_2}$, could refer either to the F_1 or F_2 dimension, or any linear combination of the two dimensions, for that matter. For this reason, previous work by our group has uniformly employed an "aggressive" method in two- and higher-dimensional spectra [5–7] to artificially "phase" the spectrum by *also* taking the real part of the complex amplitude $d_{k_1k_2} \rightarrow Re\{d_{k_1k_2}\}$ which in effect limits the "phase" of each absorption-mode Lorentzian peak to positive or negative. In addition, in the aggressive approach, we could also simply replace each absorption-mode Lorentzian peak, individually, with an absorption-mode Gaussian peak with the same integral and full width at half-maximum, thereby achieving a perfect Lorentzian-to-Gaussian line shape transformation independent of the actual widths of the peaks [5–7]. The drawbacks of this aggressive approach, however, have become apparent when overlapping peaks are characterized by badly out-of-phase entries in the FDM line list. The artificial phasing and line shape transformations in this case falsify the aggressive spectral estimate, even while the actual entire Lorentzian FDM phase-twist spectrum may still be an excellent fit to the phase-twist data. Furthermore, noise is similarly "phased-up" and can manifest itself as "new" peaks springing up from the baseline. Thus, although the aggressive approach produces pretty spectra, they are also unfaithful.

A better and more conservative approach, which we present here, is to utilize the pair of *P*/*N* data sets, process each one of these separately, and then combine the two phase-twist data sets by reflection in F_1 [17] to obtain an absorption-mode spectrum. An estimate of the infinite-time discrete FT spectrum of a finite two-dimensional time-domain *P*-type data set can be calculated from the FDM spectral parameters as:

$$S^{\infty}_{P}(f_1, f_2) = \tau_1\tau_2 \sum_{k_1,k_2} d_{k_1k_2}\left(\frac{1}{1-u^{(P)}_{k_1}/z_1} - \frac{1}{2}\right)\left(\frac{1}{1-u_{k_2}/z_2} - \frac{1}{2}\right). \quad (2)$$

Here the eigenvalues u_{k_1}, u_{k_2} give the position and width of the 2D peak in each dimension, and

$$z_1 = e^{2\pi i f_1\tau_1}; \quad z_2 = e^{2\pi i f_2\tau_2}. \quad (3)$$

For the *N*-type data, $u^{(P)}_{k_1} \approx \left(u^{(N)}_{k_1}\right)^*$. The inclusion of the $(-1/2)$ terms corrects for the first-point error in the two-dimensional spectrum [18], while explicitly including τ_i allows spectra with different spectral widths to be directly compared on an absolute scale. One should note that solving the FDM eigenvalue problem for both *P*- and *N*-type data will make the calculation a factor of two longer than the "aggressive" method, but the latter is only adequate when

peaks are sufficiently well resolved and need no phase correction, pre-conditions that are often unrealistic.

2.2. Lineshape transformation

The clean elliptical contours of Gaussian peaks make them preferable in congested spectral regions. As we have a "peak list" from the FDM eigenvalues for each spectral window, it is tempting to take the peaks, one-by-one, and convert them from the natural Lorentzian function of Eq. (2) to a corresponding Gaussian line shape. Such a transformation is impossible in FT deconvolution whenever peaks have different widths, and so it could be seen as a distinct advantage for the FDM approach. However, for a phase-sensitive presentation the dispersion-mode part of the Gaussian is also needed and, even in the case of an 1D integral FT, *i.e.*,

$$G(f)=\int_0^\infty d_k e^{-(t/T_G)^2}e^{-2i\pi ft}dt=d_k\left(\frac{\sqrt{\pi}T_G}{2}\right)e^{-\pi^2f^2T_G^2}\{1-i\,erfi(\pi fT_G)\} \tag{4}$$

we lack a useful numerical formula for the imaginary error function *erfi*(*z*), making the phase-sensitive Gaussian line shape time-consuming to evaluate. A proper discrete FT, which takes into account the aliasing of peaks, is even more involved. As the peak-by-peak substitution is not without problems, we have abandoned it in favor of conventional Lorentzian-to-Gaussian transformation by time-domain weighting. Rather than manipulating any of the peaks separately, the entire spectrum is treated as a whole. In this case the skirts of wide out-of-phase peaks, which may end up mostly canceling each other, also cancel correctly in the resolution-enhanced spectrum. The procedure is outlined in the next section.

2.3. Constructing the hybrid-FDM spectrum

The HFDM spectrum can recapture the weak features that do not appear in the EFDM spectrum obtained by the "aggressive" method using regularization, and additionally can restore the appropriate noise level to the smooth EFDM baseline [11]. Here we will discuss a two-dimensional recipe for the aforementioned phase-sensitive HFDM approach that can straightforwardly be generalized to the *n*-dimensional case.

The aim of HFDM is to use the superior resolution of FDM to capture all sharp peaks that are well above the noise, and use the discrete FT for its unbiased noise-estimate, and to ensure weak features are not suppressed entirely. The procedure employed adheres to the following sequence:

(i) Create the *P*/*N*-type EFDM spectra following Eq. (2). These phase-twist spectra have essentially no noise (although noise may have influenced the position or amplitude of the detected peaks) and can have a much finer frequency grid, enhancing resolution beyond the transform-limited line width.

(ii) Take the inverse-FT of *P*/*N*-type EFDM spectra to obtain corresponding calculated time signals:

$$C^{\infty}_{P/N}(n_1\tau_1,n_2\tau_2)=\delta f_1\delta f_2\sum_{n_1=0}^{N_{s_1}}\sum_{n_2=0}^{N_{s_2}}e^{2\pi if_1n_1\tau_1}e^{2\pi if_2n_2\tau_2}S^{\infty}_{P/N}(f_1,f_2). \tag{5}$$

Here, N_{s_i} are the number of the spectral grid points used in $S^{\infty}_{P/N}(f_1,f_2)$, while explicitly including δf_i allows spectra with different frequency steps to be correctly compared on an absolute scale. Also, the first point correction from Eq. (2) has to be undone to compare the FDM fit with the original data:

$$\widetilde{C}^{\infty}_{P/N}(n_1\tau_1,0)=2\times C^{\infty}_{P/N}(n_1\tau_1,0),$$
$$\widetilde{C}^{\infty}_{P/N}(0,n_2\tau_2)=2\times C^{\infty}_{P/N}(0,n_2\tau_2). \tag{6}$$

(iii) Over the time range for which there is recorded data, find the residuals between the true *P*/*N* time-domain data and the calculated counterparts from (ii):

$$\delta C_{P/N}(n_1\tau_1,n_2\tau_2)=\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1}C_{P/N}(n_1\tau_1,n_2\tau_2)-\widetilde{C}^{\infty}_{P/N}(n_1\tau_1,n_2\tau_2). \tag{7}$$

(iv) Apply standard Lorentzian-to-Gaussian transformation to the (longer) relatively noise-free *ersatz* FIDs, and an appropriate, but perhaps less aggressive, LG transformation to the residual, making sure to avoid truncation of the residual. This weighting can be interactively adjusted until a best compromise is obtained.

(v) Sum the *P*/*N* residuals into the *ersatz* *P*/*N* FIDs.

(vi) FT each hybrid time signal, taking the first point correction into account.

(vii) Flip the *N*-type spectrum in F_1 and add it to *P*-type counterpart to obtain absorption-mode line-shape [17] (Fig. 1).

(viii) Make the orthogonal linear combination to that in (vii) to allow the phase-sensitive HFDM 2D spectrum to be phased, if need be.

In step (vii) it is important to note that the Fourier frequency grid points are located at $-SW/2+k\,SW/N$, $k=0,\ldots,N-1$ and so wholesale reflection of the spectrum leads to a slight error in frequency registration. Instead, all points *except* the first point, at the edge of the spectral width, should be interchanged to obtain the correct flipped *N*-type spectrum. In [11] the residual spectrum for the "aggressive" method was obtained via an analytical discrete FT formula using the spectral parameters themselves (see Eq. (3) in [11]). This approach is computationally intensive because extended sinc-function line shapes must be evaluated over a fine fre-

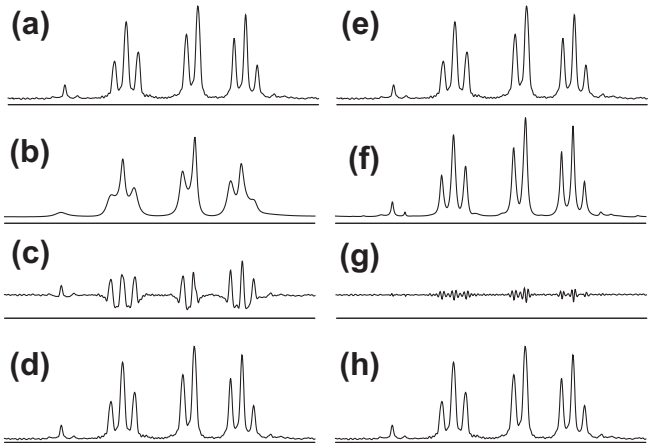

Fig. 1. The HFDM concept in action. (a) A portion of the FT spectrum of strychnine, showing several multiplets, some weaker peaks, and noise. The FID was sampled long enough that there are no issues with truncation. (b) An over-regularized *ersatz* FDM spectrum, showing the suppression of noise, and the differential broadening of less intense peaks [5]. (c) The FT of the residual, from the corresponding FIDs, (a) – (b) showing the inadequate fit, and the suppressed noise. (d) The *hybrid* FDM spectrum, (b) + (c). The excessive line broadening is corrected, and the proper noise level reintroduced. Panels (e) through (g) show the same sequence, with the only difference being that the regularization is appropriate to just remove the noise. The residual (g) is now far smaller, and the HFDM spectrum in (h) is comparable to the FT spectrum. This is the expected result, as the original data is not appreciably truncated. There is some very small resolution enhancement which gives rise to the small baseline undulations in (g) in the regions near the stronger peaks.

quency grid; and it can cause numerical instabilities in some cases, causing us to eventually abandon it. The take-home lesson from this work is that the spectrum, in its entirety, is the most reliable object to manipulate; reaching into the FDM line list to alter individual entries will sometimes produce quite misleading results.

As the FDM analysis proceeds via a number of frequency windows, each of which is usually far smaller than the entire spectral width, yet substantially larger than a typical peak width, we do not necessarily expect to be able to accurately capture wide lines of the order of the width of the frequency window. As such wide lines are not truncated in the time domain, there is no point in including them in the EFDM spectrum: they are safely deleted and relegated to the residual part of the spectrum, along with the noise. This culling of the FDM line list also improves the performance of the algorithm, by cutting down the number of peaks that need to be computed in each frequency window calculation. Likewise, extremely narrow peaks, such as may occur in a constant-time experiment, or by the influence of noise on short data records, must be broadened to at least the digital resolution of the frequency grid, or the peak may "disappear" between two adjacent grid points, neither of which is at its apex. By using an *in situ* analysis [19] the estimated standard deviation of the frequency accuracy for a typical peak can be evaluated [10], allowing a minimum line width threshold of this magnitude to be set. This minimum line width can be 10–100 times narrower than the transform-limited line width, however. The larger of this *in situ* estimate, or twice the digital resolution used in making the frequency spectrum, sets the minimum line width for the Lorentzian EFDM spectrum. Professionally-written software would automatically perform such an analysis in background, freeing the operator from any in-depth knowledge of the algorithm.

2.4. Improved regularization of the generalized eigenvalue problem

A two-dimensional spectrum necessitates the solution of two generalized eigenvalue problems, both of which may be ill-conditioned because the information may be overcomplete in a long dimension and yet also incomplete in a short one, especially when there is degeneracy in F_2. As a result, no useful 2D spectrum can be obtained without regularizing the generalized eigenvalue problems. The naive formulation is to solve

$$\begin{aligned}\mathbf{U}_1^{(\mathbf{1})}\mathbf{B}_{1_k} &= u_{1_k}\mathbf{U}^{(\mathbf{0})}\mathbf{B}_{1_k}\\ \mathbf{U}_2^{(\mathbf{1})}\mathbf{B}_{2_k} &= u_{2_k}\mathbf{U}^{(\mathbf{0})}\mathbf{B}_{2_k}\end{aligned} \tag{8}$$

where $\mathbf{U}_n^{(\mathbf{m})}$ is a matrix obtained by shifting the time-domain data by $\mathbf{m}$ points in the nth dimension. The matrices in (8) are all calculated in the Fourier basis for an n-dimensional frequency grid over a restricted frequency window comprising K_{win} total basis functions [2] and rarely exceed 500×500. The raw data matrices for a two-dimensional NMR spectrum could exceed $10^4 \times 10^4$, and are numerically intractable. (Raw matrices in 3D and 4D NMR surpass $10^6 \times 10^6$.) $\mathbf{U}^{(\mathbf{0})}$ is joint to both problems, and causes problems if it is ill-conditioned [5], which led to a reformation of (8) as

$$\begin{aligned}\mathbf{U}^{(\mathbf{0})\dagger}\mathbf{U}_1^{(\mathbf{1})}\mathbf{B}_{1_k} &= u_{1_k}\left\{\mathbf{U}^{(\mathbf{0})\dagger}\mathbf{U}^{(\mathbf{0})} + q_1^2\mathbf{I}\right\}\mathbf{B}_{1_k}\\ \mathbf{U}^{(\mathbf{0})\dagger}\mathbf{U}_2^{(\mathbf{1})}\mathbf{B}_{2_k} &= u_{2_k}\left\{\mathbf{U}^{(\mathbf{0})\dagger}\mathbf{U}^{(\mathbf{0})} + q_2^2\mathbf{I}\right\}\mathbf{B}_{2_k}\end{aligned} \tag{9}$$

where $\mathbf{I}$ is the diagonal identity matrix and the regularization parameters q_i^2 can be chosen to draw all those eigenvalues that refer to peaks with maxima within the frequency window under analysis to the interior of the unit circle, so that the fit refers only to decaying exponentials. In general, the shorter a time dimension is, the more regularization required to control the line width [10].

Here we adopt an alternative formulation that has also previously been described [20] in which a singular value decomposition (SVD) of the matrix $\mathbf{U}^{(\mathbf{0})}$ is performed

$$\mathbf{U}^{(\mathbf{0})} = \mathbf{V}\mathbf{\Sigma}\mathbf{W}^{\dagger} \tag{10}$$

where $\mathbf{\Sigma} = diag(\sigma_n)$ has real, non-negative entries and the matrices $\mathbf{V}$ and $\mathbf{W}$ are unitary. The SVD routine returns the singular values ordered so that $\sigma_1 > \sigma_2 > , \cdots , > \sigma_n$. The degree of ill-conditioning of $\mathbf{U}^{(\mathbf{0})}$ is measured by the ratio σ_1/σ_n and, in particular, if many singular values are very small or zero then the basis is too large; but in noisy NMR spectra the singular values rarely, if ever, show any sharp break between large and small entries, making it difficult to choose a smaller basis both correctly and automatically. Thus, rather than attempting this reduction, which would also perhaps discard some small but quite important peaks (*i.e.* NOEs) we typically adjust all the singular values in a conservative way, to improve the stability of the generalized eigenvalue problem and hence the spectrum obtained from it. In a recent analysis of ill-conditioned *linear* systems, the substitution

$$\sigma_n \rightarrow \sqrt{\sigma_n^2 + q^2} \tag{11}$$

has been shown to have advantageous properties [21] in controlling the pathology that such systems can manifest. We tried an analogous substitution in (8) but were generally dissatisfied with the results. Spectra were apparently distorted somewhat, while resolution was not as fine as with the older formulation in (9). However, a slight modification, namely

$$\sigma_n \rightarrow \sigma_n + q \tag{12}$$

seemed to give superior performance to any of the prior regularization methods, while still being economical, as the decomposition in (10) can be used for both F_1 and F_2, with different values of q as necessary. This substitution is similar in spirit to that in (9) but simpler to implement. It thus seems that the generalized eigenvalue problem, with its pair of matrices, has different issues than the stable inversion of a single matrix. We are still exploring the very best way to handle these ill-conditioned equations as they manifest in NMR spectra, and will report a more complete analysis in the future.

3. Numerical examples

Some straightforward simulated data sets serve to illustrate the performance of the new phase-sensitive method. Deliberately misadjusting the regularization parameter, by making it too large, will result in the loss of resolution, although such a loss would not be immediately obvious in the case of the EFDM spectrum unless comparison with a separate FT spectrum were conducted. Fig. 1 shows how the effect of inadvertent over-regularization is ameliorated. The FT spectrum shows a resolved multiplet with somewhat sparse digitization. The EFDM spectrum has been over-regularized by an incorrect choice of q, resulting in excessive line broadening. However, the residual spectrum shows an error between the true data and the broadened EFDM spectrum, and adding this error to the EFDM result largely restores the structure in the hybrid spectrum shown in the last trace. The ability to handle incorrectly phased data is shown in Fig. 2 in which the first data point in t_1 was omitted from the data set. For display purposes, a diagonal grid of peaks was synthesized. In the aggressive method the *P*-type data were analyzed and the real part of the amplitudes associated with an absorption-mode 2D peak. In the phase-sensitive method, both *N*- and *P*-type data sets analyzed, the phase-twist spectra constructed, and then the absorption-mode spectrum obtained. The expected linear phase correction in F_1 produces a satisfactory result, as it would in an FT spectrum with narrow-enough peaks. By contrast, the aggressive spectrum falsifies the peak integrals,

and shows negative peaks for the badly out-of-phase features at the edges of the spectral width.

4. Experimental

Two 2D ^{1}H–^{13}C HSQC spectra were used to test the phase-sensitive hybrid spectrum. The data were acquired at 500 MHz using a standard Varian triple resonance HCN proton observe probe with tri-axial pulsed field gradients. The pulse sequence used was previously described in [16], with the exception that the sensitivity enhancement period before acquisition was deleted in this case. The timing diagram is laid out in Fig. 3. The CLUB [22] gradient encoding step conveniently allowed us to zoom in on a chosen spectral region in ^{13}C by simply substituting frequency-modulated band-selective inversion pulses into the sequence, leaving all other parameters unchanged.

A 11.1 mg quinine sample in 800 μl $CDCl_3$ was used as the first test case. The data set comprised 32 × 2000 in the respective t_1 and t_2 dimensions over spectral widths of 20 and 8 kHz, respectively, resulting in acquisition times AT_1 = 1.6 ms and AT_2 = 250 ms. The more demanding HSQC spectrum of strychnine provided a better calibration of the method when there was additional spectral congestion, and additionally illustrates how convergence can be accelerated by zooming in on a congested region once the rest of the spectrum has stabilized. The very large ^{13}C spectral range was decoupled using an adiabatic decoupling scheme described in literature [27] (STUD) or in the case of more restricted bandwidth experiments either WALTZ-16 [28] or GARP [29] as appropriate. An effective rf, in terms of heating, of around 2 kHz on ^{13}C, achieved a 20 kHz adiabatic decoupling bandwidth with a 1 ms inversion pulse, but resulted in appreciable cycling sidebands that appeared along F_2 traces in the 2D spectrum. No particular suppression of these artifacts was attempted aside from a high-power

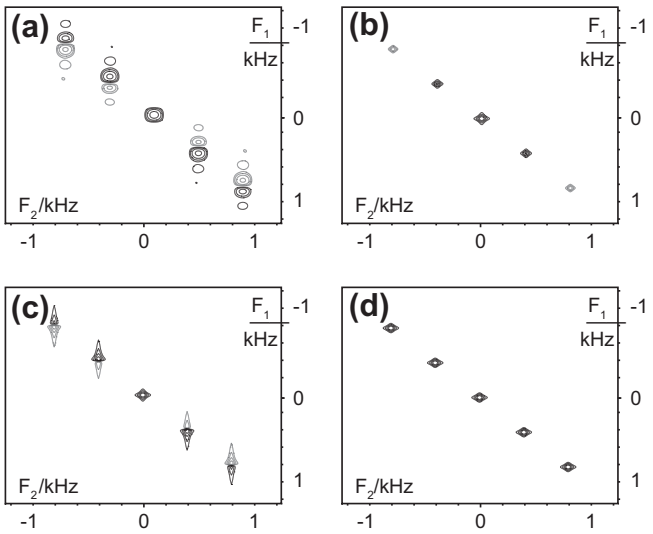

Fig. 2. Performance of phase-sensitive FDM compared with the aggressive approach of artificially phasing each peak. A synthetic data set, with no noise, comprises a diagonal set of peaks. The first point in t_1 has been omitted, resulting in a large linear phase error in F_1. (a) The conventional FT spectrum, showing the lack of resolution in F_1 due to a short data record, and the strong linear phase error. (b) The EFDM spectrum using the aggressive approach, in which every "peak" is artificially phased and an absorption-mode 2D peak placed in position. The phase error gets converted into an amplitude error, and an incorrect algebraic sign at the edges of the spectral width. (c) The phase sensitive HFDM spectrum, in which both *N*- and *P*-type spectra are analyzed, and phase-sensitive phase-twist peaks are used. (d) The result of (c) when a conventional linear phase correction is applied. As in the FT case, the correct spectrum is obtained. However, the resolution is far superior compared to that in (a).

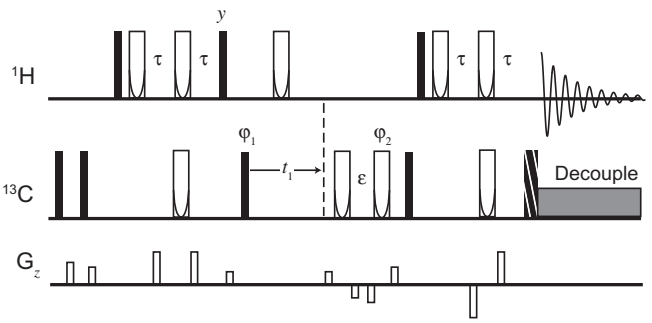

Fig. 3. The pulse sequence timing diagram for the 2D HSQC experiments. Native carbon-13 magnetization is scrubbed by two 90° pulses (narrow filled icons) and gradients (on the G_z stave). Open icons, with a parabolic phase profile inscribed at bottom, are BIP [23] inversion pulses. At the end of the t_1 encoding of the carbon-13 transverse magnetization is accomplished with a CLUB [22] and decoded with a bipolar gradient. The isotope-selective CLUB unit also proved useful in aqueous solutions [data not shown] where spurious peaks resulting from intermolecular multiple-quantum coherence [24] are avoided. The final "barber pole" icon is a broadband compensated 90° that purges anti-phase magnetization and seems to reduce the adiabatic decoupling sidebands in F_2. The phases ϕ_1 and ϕ_2 are used for TPPI [25,26] and to subtract out the residual proton signal from the carbon-12 isotopomers, respectively. The delay ϵ is used to ensure that the first increment has effectively zero time evolution in t_1 [22], avoiding any frequency-dependent phase correction in F_1.

frequency-modulated 90° ^{13}C pulse just prior to decoupling, to purge anti-phase magnetization from the first point of the ^{1}H FID. The sidebands, remaining visible above the thermal noise, were used as a convenient test of the hybrid method's ability to recover weak signals along with noise.

The solution of Eq. (8) was conducted over a frequency window that included the entire F_1 spectral width, and a narrow frequency window in F_2 such that the number of 2D basis functions was kept below 200 × 200 regardless of the actual number of data points. Window calculations were then quilted together into the whole spectrum as described previously. This ability of FDM to handle huge data sizes that would be prohibitive to tackle in one matrix diagonalization [1,2] is a major strength of the approach.

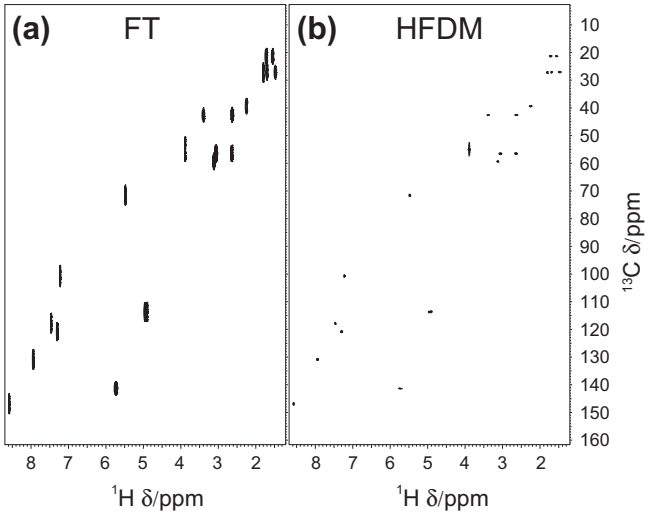

Fig. 4. A comparison of (a) FT and (b) HFDM processing of a 2D HSQC spectrum of quinine, a molecule with peaks across a large fraction of both the ^{1}H and ^{13}C chemical shift range. Only 32 increments in t_1 were taken, so that peak widths in F_1 in the FT spectrum are of the order of 5 ppm (625 Hz). Although the natural width is many times less than the transform-limited width, the contour plot in (a) is relatively well-resolved in the 2D spectrum. Thus, while the resolution in (b) is far better in F_1, the qualitative information content of the two renditions is comparable. We consider (b) to be essentially a "cosmetic" improvement of (a).

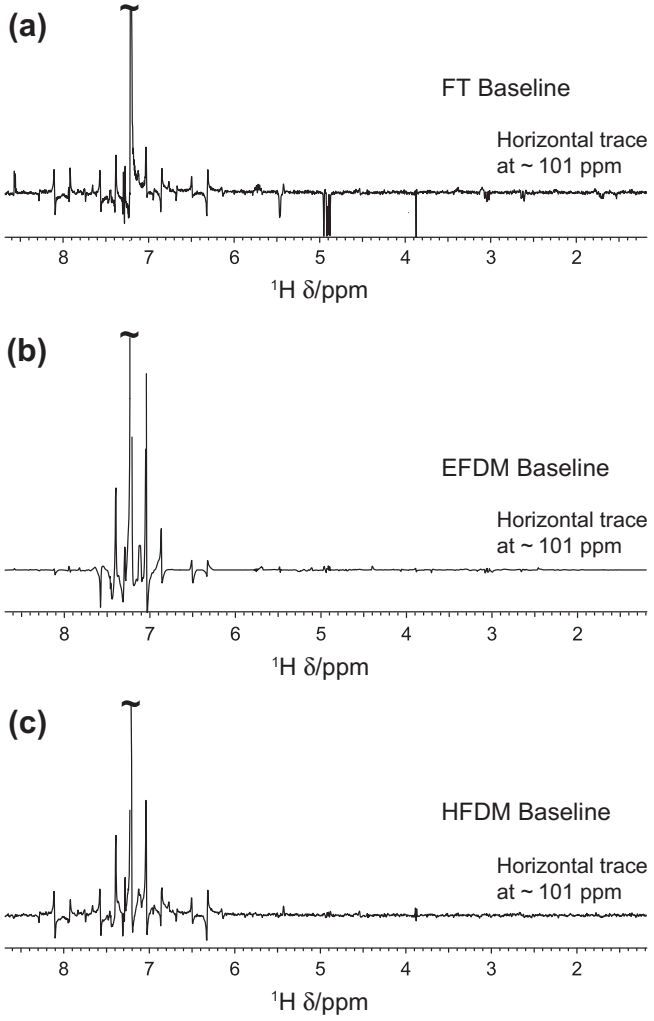

Fig. 5. Proton horizontal traces from the 2D HSQC spectra near 101 ppm in the ^{13}C dimension. (a) The FT spectrum, showing the major peak, cycling sidebands apparent as a host of smaller out-of-phase peaks, noise, and some breakthrough from other peaks at different ^{13}C chemical shifts, as the 2D peaks are wide in F_1. (b) The regularized, but phase-sensitive EFDM spectrum. Much of the true noise has been suppressed, and the dispersion-mode peak near 6.35 ppm is broadened by the regularization. (c) The hybrid FDM spectrum restores the noise and weaker cycling sidebands. The main resonance is more intense than in (a) because the 2D peak in (c) is much narrower in F_1 compared with the transform-limited line width in (a). The disappearance of the doublet near 8.6 ppm is due to the narrower line width in F_1 as well, as these peaks arise from a different 2D resonance.

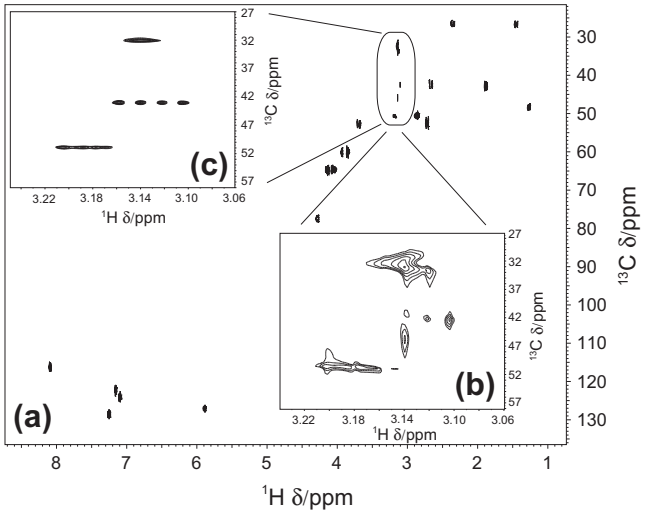

Fig. 6. The HFDM 2D HSQC spectrum of strychnine. (a) The survey spectrum obtained with only four t_1 increments. High-resolution proton multiplets appear at the chemical shift positions of the carbon-13 spins, and a good fit is obtained over 90% of the spectrum. But a region with several partially overlapping proton multiplets cannot be resolved, as shown by the poor result in (b), an expanded view of that in (a). Focusing on this region alone, by using band-selective frequency-modulated inversion pulses in the CLUB gradient encoding step, allowed a narrowing of the spectral width in F_1 by a factor of five, and repeating the acquisition using six increments over this region then gave the stable and reliable result shown in (c).

5. Results and discussion

Fig. 4 compares the 2D FT and 2D HFDM spectra of quinine using the pulse sequence of Fig. 3. The peaks are quite well resolved even in the transform-limited FT spectrum. Thus, although the HFDM spectrum has higher resolution, the information that an experienced spectroscopist is likely to be able to extract is roughly comparable. In this case the improvement is mostly cosmetic rather than material. (The narrower peaks in F_1 do, however, reduce the relative adiabatic decoupling sideband intensity, because the weak sidebands remain wide in F_1 as they are too weak to contribute to the regularized FDM line list. The *apparent* signal-to-noise ratio is also improved, although it is well known that this improvement is entirely illusory because the sensitivity remains the same and only strong-enough peaks, which are easily recognized above the noise level, can be captured and narrowed.)

Fig. 5 shows a trace near 101 ppm in the ^{13}C dimension, with the vertical gain increased so that the noise and cycling sidebands are apparent. The FT spectrum (a) shows the intrinsic sensitivity and the out-of-phase sidebands, as well as some breakthrough from large, broad peaks at different ^{13}C chemical shifts.

However, the high-resolution spectrum is advantageous when certain regions are well resolved while others are more crowded, as it is far easier to see which areas could benefit from more complete data acquisition, and which regions have essentially converged onto a final result. Fig. 6a shows the 2D HFDM spectra of strychnine using the pulse sequence of Fig. 3 and only *four* time increments in t_1, for an acquisition time of just 200 μs. Most of the spectrum is free and clear, with a small congested region in which there are noticeable distortions, as clearly shown in the inset region blown up in (b). In this local region the number of strychnine 2D peaks exceeds the number of peaks that FDM can use to fit the data, based on the information content of the 2D signal. This can also be confirmed by the higher residual in this region compared to similar 2D areas elsewhere, where the fit is accurate. However, as the rest of the spectrum is well-resolved, it is a simple matter to "zoom in" on the region in question by using a pair of identical band-selective FM refocusing pulses in the CLUB gradient encoding step. This allows the 20 kHz spectral width to be narrowed to 4 kHz and, using just six increments over this restricted region, the full and correct proton multiplet structure is revealed, in panel (c). A 2D FT survey spectrum of strychnine with only 10 increments in total, by contrast, shows such poor resolution in F_1 that it is not apparent which, if any, of the regions actually *have* peaks that are unresolved. Thus, the ability to rapidly and accurately identify high-resolution regions with sparse resonances using HFDM can aid high-throughput applications by allowing one to focus on a few congested regions without any ambiguity in the rest of the 2D plane. These areas may then be substituted, if desired, for the corresponding unresolved frequency region in the survey spectrum, to arrive at a single display containing all the information. The incorporation of the CLUB element allows the same pulse sequence to be used for survey or band-selective experiments, and has proven convenient and flexible.

6. Conclusion

Short cuts, as the aphorism reads, lead to long delays. This has certainly been true in the construction of a proper frequency spectrum from the output of the FDM algorithm. The temptation to take "peaks" at face value and manipulate each of them individually, changing phase and line shape to achieve a clean-looking spectrum, was overwhelming. However, such spectra are potentially quite misleading when it comes to assignment and the extraction of information that is ultimately used to arrive at chemical structure. Cosmetic enhancement of the spectrum itself should not be the goal, which should rather be to facilitate the more rapid, more reliable assignment and structure determination of unknowns. In cases where sensitivity is adequate, the HFDM spectrum can enhance the resolution of sharp, stronger peaks, while still producing a spectrum in which all the original imperfections are present, and a true noise floor prevents inadvertent contouring of features that look like resonances but that would in fact be buried in noise in any decent spectral estimate. Using *in situ* analysis, error bounds can be given, and areas of poor local fit can be selected and rapidly improved by a trivial modification of the pulse sequence. Regularization via SVD has some advantages for multidimensional spectra and, as it seems to have no obvious disadvantage in other regards, can be recommended for all FDM approaches. Taken together, these features make HFDM an attractive all-purpose method to estimate many different kinds of multidimensional NMR spectra, lower the burden on NMR spectrometer time, and make best use of the new generation of high-sensitivity NMR probes. The CLUB–HSQC sequence allows one to zoom in on restricted spectral region, and HFDM makes it possible to identify trouble spots rapidly and obtain extra data to ensure sufficient resolution while ignoring larger swaths where spectrum is essentially already fully converged.

Acknowledgments

Hasan Celik was supported by a Fellowship from the Amgen Foundation and a UCI Graduate Dean's Dissertation Fellowship. We thank Prof. V.A. Mandelshtam for his insight into various problems with FDM, and his willingness to share his considerable expertise with us.

References

[1] M.R. Wall, D. Neuhauser, Extraction, through filter-diagonalization, of general quantum eigenvalues or classical normal-mode frequencies from a small number of residues or a short-time segment of a signal. 1. Theory and application to a quantum-dynamics model, J. Chem. Phys. 102 (20) (1995) 8011–8022.
[2] V.A. Mandelshtam, H.S. Taylor, Harmonic inversion of time signals and its applications, J. Chem. Phys. 107 (17) (1997) 6756–6769.
[3] 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 (1) (1998) 76–87.
[4] J.H. Chen, V.A. Mandelshtam, Multiscale filter diagonalization method for spectral analysis of noisy data with nonlocalized features, J. Chem. Phys. 112 (2000) 4429–4437.
[5] J.H. Chen, V.A. Mandelshtam, A.J. Shaka, Regularization of the two-dimensional filter diagonalization method: FDM2K, J. Magn. Reson. 146 (2000) 363–368.
[6] J.H. 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.
[7] J.H. Chen, D. Nietlispach, A.J. Shaka, V.A. Mandelshtam, Ultra-high resolution 3D NMR spectra from limited-size data sets, J. Magn. Reson. 169 (2004) 215–224.
[8] G.S. Armstrong, K.E. Cano, V.A. Mandelshtam, A.J. Shaka, B. Bendiak, Rapid 3D NMR using the filter diagonalization method: application to oligosaccharides derivatized with C-13-labeled acetyl groups, J. Magn. Reson. 170 (1) (2004) 156–163.
[9] 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 (1) (2005) 160–168.
[10] H. Celik, A.J. Shaka, Filter diagonalization using a "sensitivity-enhanced basis": improved performance for noisy NMR spectra, J. Magn. Reson. 207 (2010) 17–23.
[11] C.D. Ridge, A.J. Shaka, "Ersatz" and "hybrid" NMR spectral estimates using the filter diagonalization method, J. Phys. Chem. A 113 (2009) 2036–2052.
[12] A. Bax, M.F. Summers, H-1 and C-13 assignments from sensitivity-enhanced detection of heteronuclear multiple-bond connectivity by 2D multiple quantum NMR, J. Am. Chem. Soc. 108 (1986) 2093–2094.
[13] R.E. Hurd, Gradient-enhanced spectroscopy, J. Magn. Reson. 87 (1990) 422–428.
[14] D.J. States, R.A. Haberkorn, D.J. Ruben, A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants, J. Magn. Reson. 48 (1982) 286–292.
[15] G. Bodenhausen, R. Freeman, R. Niedermeyer, D.L. Turner, Double Fourier transformation in high-resolution NMR, J. Magn. Reson. 26 (1977) 133–164.
[16] S. Keppetipola, W. Kudlicki, B.D. Nguyen, X. Meng, K.J. Donovan, A.J. Shaka, From gene to HSQC in under five hours: high-throughput NMR proteomics, J. Am. Chem. Soc. 128 (2006) 4508–4509.
[17] P. Bachmann, W.P. Aue, L. Muller, R.R. Ernst, Phase separation in two-dimensional spectroscopy, J. Magn. Reson. 28 (1977) 29–39.
[18] E.O. Brigham, The Fast Fourier Transform, Prentice-Hall, 1974.
[19] J.C. Hoch, A.S. Stern, NMR Data Processing, Wiley-Liss, 2005.
[20] G.S. Armstrong, J.H. Chen, K.E. Cano, A.J. Shaka, V.A. Mandelshtam, Regularized resolvent transform for direct calculation of 45° projections of 2D J spectra, J. Magn. Reson. 164 (1) (2003) 136–144.
[21] J.R. Jensen, D.E. Ramirez, Anomalies in the foundations of ridge regression, Int. Stat. Rev. 76 (2008) 89–105.
[22] H.T. Hu, A.J. Shaka, Composite pulsed field gradients with refocused chemical shifts and short recovery time, J. Magn. Reson. 136 (1) (1999) 54–62.
[23] M.A. Smith, H. Hu, A.J. Shaka, Improved broadband inversion performance for NMR in liquids, J. Magn. Reson. 151 (2001) 269–283.
[24] Q.H. He, W. Richter, S. Vathyam, W.S. Warren, Intermolecular multiple-quantum coherences and cross correlations in solution nuclear magnetic resonance, J. Chem. Phys. 98 (1993) 6779–6800.
[25] G. Drobny, A. Pines, S. Sinton, D.P. Weitekamp, D. Wemmer, Fourier transform multiple quantum nuclear magnetic resonance, Faraday Symp. Chem., S. 13 (1978) 49–55.
[26] G. Bodenhausen, R.L. Vold, R.R. Vold, Multiple-quantum spin-echo spectroscopy, J. Magn. Reson. 37 (1980) 93–106.
[27] M.R. Bendall, Broad-band and narrow-band spin decoupling using adiabatic spin flips, J. Magn. Reson. Ser. A 112 (1995) 126–129.
[28] A.J. Shaka, J. Keeler, R. Freeman, Evaluation of a new broad-band decoupling sequence – WALTZ-16, J. Magn. Reson. 53 (1983) 313–340.
[29] A.J. Shaka, P.B. Barker, R. Freeman, Computer-optimized decoupling scheme for wideband applications and low-level operation, J. Magn. Reson. 64 (1985) 547–552.