"Denoising 1D Solution NMR (DSNMR)"
Fewer Transients - More Sensitivity
Magnet Resonance Analysis by
Singular Value Decomposition Harmonic Inversion

Introduction

Random noise arising from the measurement process and equipment is the reason that large numbers of transients are needed in NMR measurements. To observe the signal features in the Fourier time to frequency method (FT) sufficient transients (N_{tr}) must be collected to raise the signal to noise (S/N) ratio to about three. For the method presented here, which is valid only for gas and liquid phase samples as it assumes that isolated signal lines are Lorentzian, it is empirically found that an acceptable spectrum can be achieved at about an S/N ratio of about 1.10. As such, a transient reduction by about a factor of nine is expected and can be obtained for challenging problems such as those involving ^{15}N, ^{17}O and some ^{13}C (quaternary type). For problems where signal to noise is not an issue, this method offers few advantages over the standard FT method. The present method requires, for reasons of internal statistical assumptions, at least 50 transients. For reasons to be made clear below the method will be referred to as the Singular Value Decomposition Harmonic Inversion Method and denoted as SVDHI.

The SVDHI method has two additional smaller advantages. First is that resolution is enhanced relative to the FT method by a factor of $\Delta_{average}/\Delta_{minimum}$, Δ being an adjacent interline distance; this is why the Harmonic Inversion (HI) time to frequency processor is called a high resolution processor. Second the HI method actually tabulates the three parameters, ω; the frequency, d; the area and Γ; the width of each Lorentzian in the spectrum prior to producing a spectrum and draws the Lorentzian curves to produce a spectrum. Two disadvantages, or better called annoyances, exist. First is that often, but not always, the line heights are misleading (but areas aren't). The heights are given as d/Γ and Γ is the most inaccurate of the parameters while ω is the most accurate. This arises from the fundamentals of the theory. Second is the fact that the method works in windows so as to limit the number of spectra features. Numerical problems in the internal computations would occur if this wasn't done. In many situations (e.g. looking for a multiplet) a single window is all that is needed. To obtain a full spectrum, half overlapping windows need be patched together as inaccuracies occur at the

window edges. For spectra over the total bandwidth the number of transients needed will be determined by the window with the weakest feature. So if you know where the weakest features will be, work there first. The reason for this will become clearer below. The assumption in the theory of the SVDHI method that isolated features can be represented by a Lorentzian agrees with NMR theory only for the gas and liquid phase and is the reason for the methods limitation. On the other hand it is this assumption that is "extra information" that gives the SVDHI method its advantages over the FT method.

Background

Before describing the program that implements SVDHI, some words about how the method achieves it's aims are in order. In the FT method the experimentally observed FID is represented in a basis supplied by the discretization of the time representation of the function exp(iωt). When this representation is Fourier transformed, the result is a stick spectrum which at the Fourier grid points represents the weight or intensity of the spectrum at each grid point. Ignoring the resolution problem, this result can be used to represent almost any functional form of the FID and of the final spectrum. For a given number of transients, the transient averaged FID represents signal and noise contributions to the spectrum equally well. In particular each "stick" can have both signal and noise contributions with no obvious way to separate them. The only generic procedure to reduce noise and to obtain the signal contribution is to collect more transients, i.e. increase N_{tr} so that the noise decreases proportional to $\sqrt{N_{tr}}$. The SVDHI method uses this same FID at a given N_{tr}. This is windowed and then sent to the SVD part of the program.

The SVD function will use the signal averaged FID to produce a "cleaned" or "noise reduced" windowed FID. The HI method takes this latter FID and produces a spectrum made up of separate and/or a small number of interfering Lorentzian functions, each indexed by it's parameters, ω, d and Γ. The catch is that the SVD method of success (the word windowed is now implicit) in "cleaning" the FID depends on a decision made by the operator viewing the SVD graph. This graph shows a plot of a set of usually about 40 positive real numbers called singular values (SV) whose value on the ordinate goes down as their index goes up on the abscissa. A log plot is also available and can be useful in observing the spacing between adjacent SV.

A singular value can contain any mix of signal and noise components. For a FID made only of noise, each SV (from theory) has a value equal to the standard

deviation. The SV graph would then exhibit singular values that appear as a horizontal "string of pearls." As N_{tr} increases, the line's distance from the abscissa would decrease as $\sqrt{N_{tr}}$. On the other hand, if there had been signal information in the noisy FID and if this would have eventually lead to say K Lorentzian spectral features, then as N_{tr} increased and the noise decreased, the first K singular values would decrease in value in a process that wrings the noise out of them leaving only the signal component. The latter would no longer decrease with more transients and hence, would roughly stabilize in value on their respective vertical index lines. The biggest lines with the most signal would stabilize first and so on. Empirically, this occurs at a S/N ratio which we find to be locally about a line at about 1.10. Therefore as N_{tr} increases stabilization of K singular values resembles a situation where one, two, or several of the left most pearls no longer appear to continuously decrease in value as rapidly as the remaining string of pearls. This causes the SVs to lie on a curve that for example resembles that shown in Figures 1a and 1b. If K was not known, sufficient transients would be needed to get to a graph as in Figure 1c where the distance between the last stabilized SV and the left most SV on the string of pearls is well defined. This distance is called the "gap". This in fact determines K in the usual situation where it is not known. At an N_{tr} corresponding to fig. 1b, it is not clear if the 4th singular value will detach and stabilize. K cannot be specified; it could be 3; it could be 4. This region of the curve where this uncertainty exists always has the shape of an "elbow" and is referred to as such. As seen in fig. 1c, the 4th SV decides to join with the string of pearls determining in our example K as 3.
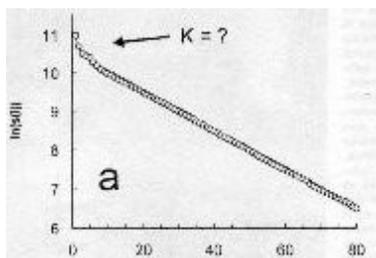
Figure 1a
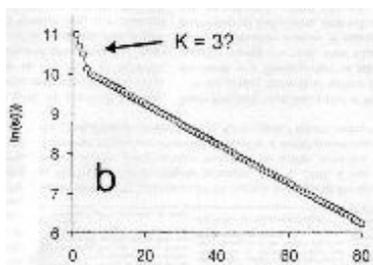
Figure 1b
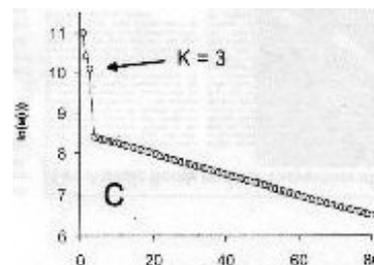
Figure 1c

Figure 1c is the ideal situation; the gap is big. However this could take so many transients that the savings over the FT method might not be as impressive. To obtain larger savings it would be advantageous to be able to spot the gap earlier in the evolution of the SVD graph as a function of $N_{tr.}$. This is possible if we watch this evolution as opposed to looking at a single graph. This gap appears when in the elbow region the evolution over a reasonable number of transients convincingly demonstrates that every SV is either stable and seems to fail to keep up with the dropping string of pearls or in contrast is moving to join the string of pearls. In the

special case where it is known (foreknowledge) that there will be only one spectral feature in the window, a damping of the original FID can decrease noise and make the gap appear earlier. For more than one feature, if they are smaller than the most prominent feature, they will get suppressed with the noise and make the fact that K should be greater than one be harder to observe i.e. more transients rather than less would be needed to observe the gap. What does help in this situation is to remove the noisiest part of the FID by dropping the last half or even three quarters of its samples. This decrease in resolution is usually more than made up for by the mentioned high resolution ability of the Harmonic Inversion method.

When in the eventually obtained spectrum the features are well separated and not standing on a background or on the base of a much larger overlapping line, the largest SV will be associated with the most prominent spectra feature, the 2nd with the 2nd etc. When small features lie near big features (dynamic range problems) many more transients will be needed to see them. The big features may be solvent lines that lay in the window or could arise because one could be trying to see a splitting that lies about a big chemical shifted feature. In such cases a first "gap" will appear below the number of singular values needed for the big lines and the problem will reduce to finding, now collecting many more transients, a second gap after the singular values for the small features appear. This is all analogous to further transient collection in the FT method. If the small features lie on the base of big lines, they can no longer be expected to be well represented by a single Lorentzian and the 2nd gap may appear only after several SV's detach and stabilize from the left of the first gap's string of pearls. Another example for the need to not stop at the first gap is when a single peak is known to really be the low solution of say a triplet. The first gap will be at K=1 where the single peaks appears. As in the FT method, many more transients will be needed to see the stabilizing SV's that give the triplet and even then more than three SV's may be needed to describe this overlapping case. Clearly any help in specifying K would be welcome. A "helping" technique is available but it requires we first explain how the spectrum is obtained.

At a given N_{tr} the SVD program takes the available windowed transient averaged FID and creates data vectors as follows. As decided by computer memory size, the program might take the hundred data bits starting with the first data point as the first vector; the hundred starting with the second data point would be taken as the second vector and so on. These vectors then become the columns of a so called data matrix which when multiplied from the right by its conjugate transpose gives by construction a Hermitian matrix whose eigenvalues must be positive. The eigenvalues, which are our SVs, and eigenvectors of this matrix are then obtained. The data vectors can then be expanded on these eigenvectors. It can be shown that

the expansion coefficient for each eigenvector is the square root of it's own eigenvalue (singular value). The program then sets in the expansion the eigenvalues numbered K+1, K+2, etc to zero, thereby eliminating most noise by eliminating all components of the data vector that are deemed to be noise. What remains is a "cleaned" expansion of the data vectors, which components are the cleaned data bits needed to construct cleaned FID elements and equivalently a cleaned FID. This is then fed into the H.I. program which "fits" the data to K Lorentzian functions by solving for their individual parameters ω, d, and Γ. The Lorentzians are then drawn along a frequency axis to give the spectrum.

A method for building confidence in the choice of K at an early stage of a possible separation of the SVD curve is clearly needed. An error in choosing K comes at a high price. Too small a K value omits signal features, too large adds false ones. The method used here depends on the fact that as the noise is random, it is not represented well by Lorentzian functions. Changing N_{tr} then shows changes in their parameters and their associated Lorentzian features in the spectrum seem to move about or change shape. The frequency of a noise line could well change and its area should decrease with increase in N_{tr}. Moreover it has been observed that such lines don't phase well. True signal lines phase well and are quite stable in ω although they may change for numerical reasons slightly with N_{tr}; the change will not always be to smaller values. As such for given N_{tr}, once K is specified the program will produce a spectrum. Then one can loop back and increase K by 1 and then by 2 and in each case get a spectrum. Now repeat this for the next two values (or perhaps two previous values if further transients are not accessable) of N_{tr} and compare the respective spectra of the same K value. If relative to the first two N_{tr} spectra, the highest N_{tr} spectra, for K, K+1, and K+2, don't show the same stable features more transients are needed to determine K. Success in this process will give high confidence in the choice of K and should do away with the need to obtain very large gaps that would be made by lowering the height of the string of pearls. Clearly, variations of the spectral consistency idea are imaginable. It costs nothing to look at a K+5 spectrum as there may be a slightly smaller feature waiting to emerge. More prior N_{tr}s may be needed to verify the stability or instability of a feature. The program will instruct the user how to graphically make such comparisons in a line by line manner. Of course, all spectra are available in tables that give the parameters which also could be used for comparison. Line widths should not be used to study stability as they are often not stable for numerical reasons.

Further Reading

A new signal processing method to observe weak ^{31}P and ^{17}O NMR peaks, S. Kunikeev, J.J. Pan, A. Kershaw, C. McKenna, H.S. Taylor, J. Org. Chem. 690, 2644-2650, (2005)

New Signal Processing Method for Faster Observation of Natural Abundance ^{15}N NMR Spectra and its application to N_5^+, S.D. Kunikeev, H.S. Taylor, T. Schroer, R. Haiges, C.J. Bigler Jones,and K.O. Christe, Inorganic Chemistry, 45, 437-442, (2005)

If any problems occur in using these programs or if any instructions are not clear please contact Allan Kershaw at: kershaw@ usc .edu and cc to: taylor@ usc.edu.We would appreciate any feedback you can supply to us concerning your experience with the software and the method.

Running the program

A single analysis is specified by the choice of two parameters. The first (M) whose default value is 2, sets any selected FID lengths at N/M, where N is the number of points in the FID. The second is the damping (LB), which defaults to the digital resolution. At this point, the analyst chooses an N_{tr} value which refers the analysis to the stored FID at N_{tr} in the Data Directory. Since this is the first FID to be examined, a window must be established. As for all N_{tr}, the FFT is shown and here provision is made for specifying the window. The program then produces internally a windowed FID at the N_{tr} and it will do so similarly whenever another choice of N_{tr} is made in the future. For this and for all future choices of N_{tr}, the program will produce automatically an SVD diagram. The operator can select an alternate logarithmic version of the diagram if it seems it is more visually revealing. Additionally, the operator can choose to increase or decrease the ordinate spacings to get a better view of the elbow if it exists. The operator then chooses a value for K. Rough high side of the elbow estimates should be made for any elbow for which the gap is not clear. Once K is chosen, a cleaned signal windowed transient averaged FFT spectrum will appear. If some features are apparent, manual phasing is available and should be applied as experience shows that this also leads to narrower lines in the HI function. An option to use a pre-existing phase is also available. The HI spectrum is displayed when phasing of the cleaned signal is finished. The HI can also be phased in an identical manner. Once phasing is completed the operator can choose to save an image of any window for various K+m, m=0, 1, 2…spectra. Usually m is chosen as 0, 1 and 2 but for cases where the situation is uncertain higher spectra for higher values of K can be also stored.

The operator now chooses the next N_{tr} and proceeds as before except for the windowing. If one is working online, the N_{tr} will presumably be chosen to increase. For comparison, several SVD diagrams from successive values of N_{tr} can be displayed and compared to see the evolution toward the gap. When a gap or near gap is observed in the elbow but is not convincing, an evolutionary sequence of spectra with the same SVD cut offs, ie; K, K+1, K+2—etc but sequential N_{tr} can be examined feature by feature. Stability, instability, bad phasing, area shrinkage, etc. are the features to look for. This can be used to support the choice of K. If it is not supportive, further N_{tr} must be collected and analyzed. For offline operation after transient collection is completed, the last three or four FIDs in the Data Directory should be studied.

Example Study 1

As an example for ^{13}C, a 0.040 molar solution of butyltrifluoroacetate was studied. As we wished to compare the SVDHI method to the FT method, a 24 hour run was made collecting 3200 transients which was needed to produce the quality of FT spectrum seen in Figure 2 where we estimate a S/N about 3. The carbonyl region was chosen to be studied and a 266 point window was placed around the expected frequency region. This region was chosen because carbonyl signals are challenging due to their quaternary nature. This particular example is even more challenging because of the splitting imposed by the adjacent trifluoromethyl group.
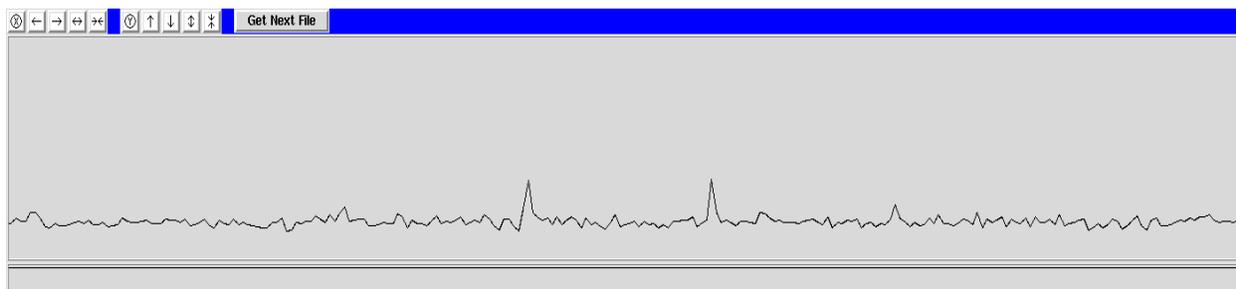
Figure 2 Standard FFT of data after 3200 scans

Since a 9 to 10 times saving in transients was hoped for N_{tr}=320 was chosen for analysis from the NDD. After that, N_{tr} values in steps of Δ, which had been chosen as 40, were examined up to N_{tr}=560. At this point, we believed that we saw the gap opening due to the evolution seen in fig 3a – 3g. The HI spectra supported this choice of K=4. In the SVD curves, no convincing gap develops in the elbow region until N_{tr}=480. This is reinforced by the curves for N_{tr}=520 and N_{tr}=560 and this can be seen in the 8-fold expansions to the right of each curve. This suggests that K=4 is the correct value. The HI spectra were calculated using K=6 so that the expected lines would show even if there were a couple of large noise lines present. The HI spectra are shown in the second column of Figure 4. The first column of Figure 4 shows the traditional FFT spectra and just where only the two large center lines of the expected quartet are visible. If no foreknowledge exists, little would be learned from the FFT in this region. If one did not know that K=4 was the expected result, then at N_{tr}=320 one might assume that the gap has formed at K=2, but a quick glance at the K=6 HI spectrum confirms that there seems to be a quartet plus two badly phased lines. As we proceed down the second column of Figure 4, you can see the instability in frequency of the two noise lines and the increase in area of the quartet lines.
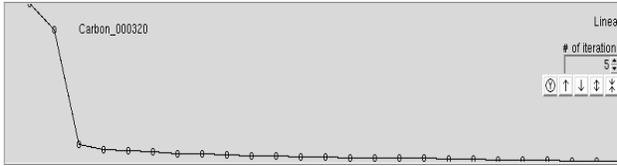
Figure 3a 320 scans
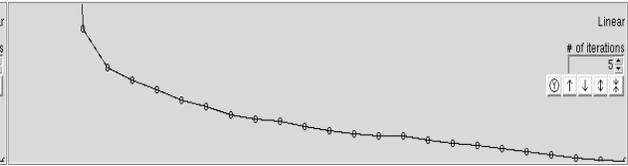
Figure 3a 8-fold expansion
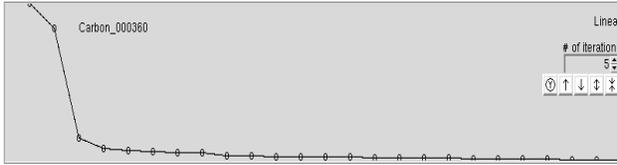
Figure 3b 360 scans
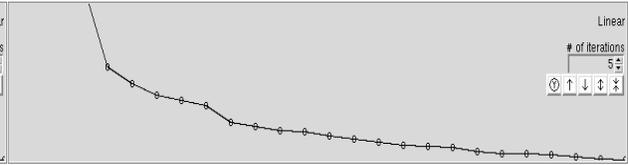
Figure 3b 8-fold expansion
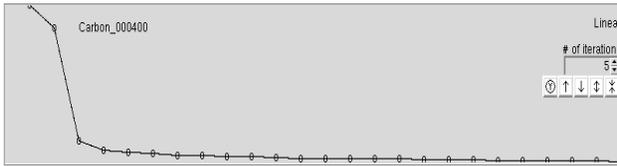
Figure 3c 400 scans
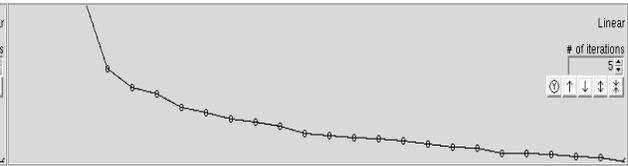
Figure 3c 8-fold expansion
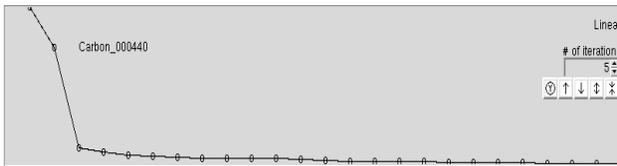
Figure 3d 440 scans
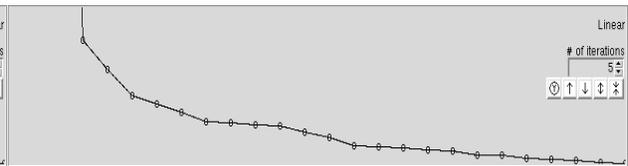
Figure 3d 8-fold expansion
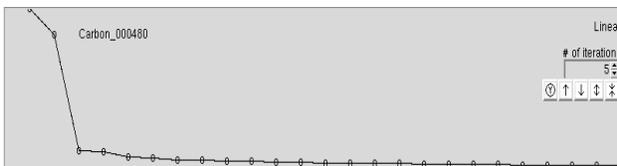
Figure 3e 480 scans
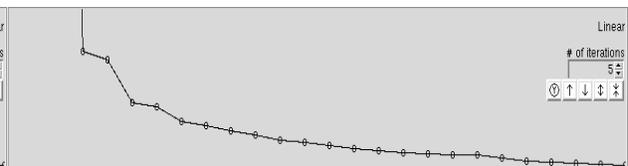
Figure 3e 8-fold expansion

Figure 3f 520 scans
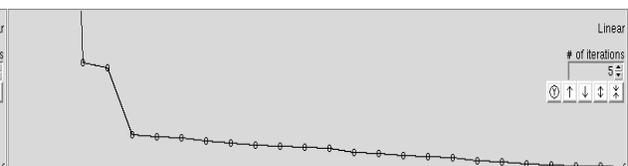
Figure 3f 8-fold expansion

Figure 3g 560 scans
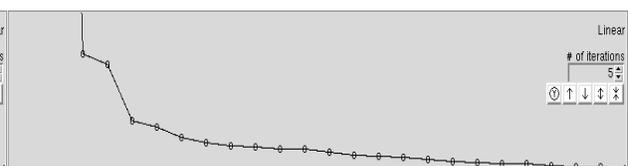
Figure 3g 8-fold expansion

9
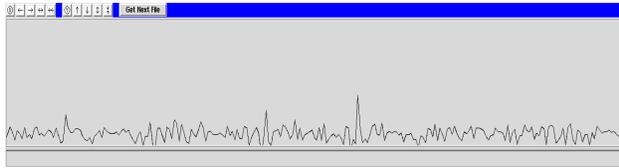
Figure 4a FFT after 320 scans
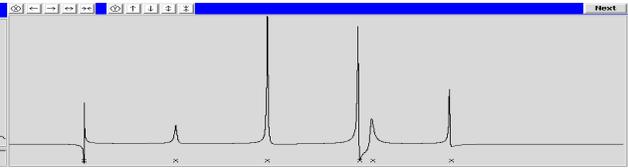
Figure 4a HI after 320 scans
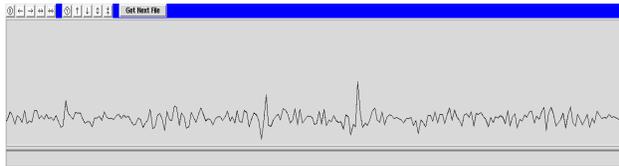
Figure 4b FFT after 360 scans
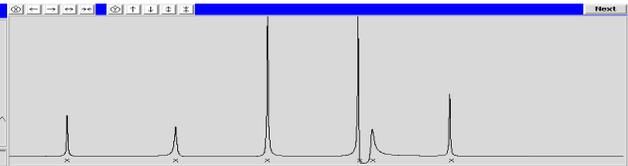
Figure 4b HI after 360 scans
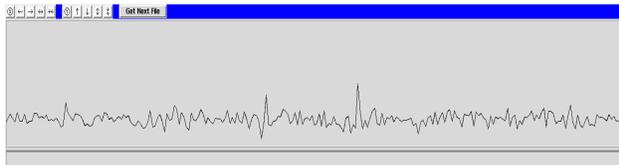
Figure 4c FFT after 400 scans
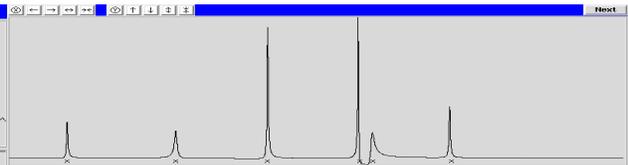
Figure 4c HI after 400 scans
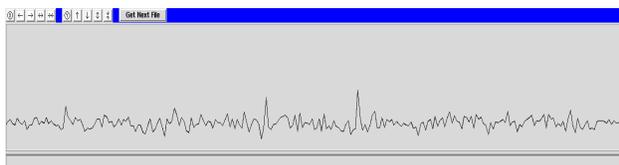
Figure 4d FFT after 440 scans
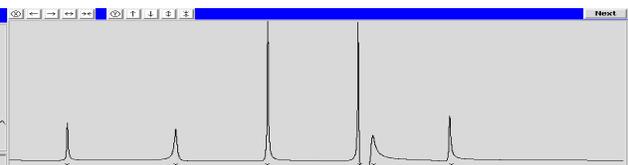
Figure 4d HI after 440 scans
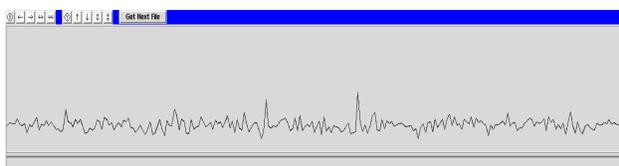
Figure 4e FFT after 480 scans
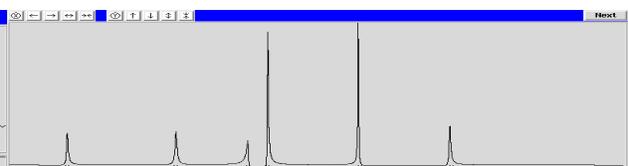
Figure 4e HI after 480 scans
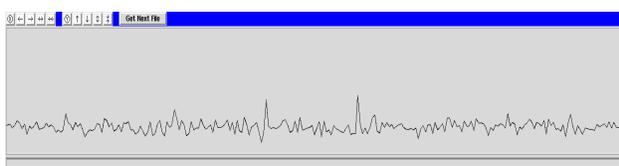
Figure 4f FFT after 520 scans
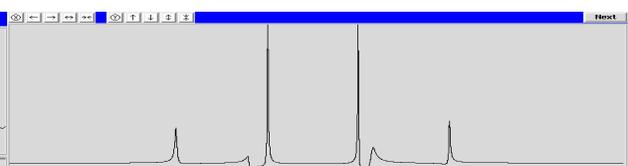
Figure 4f HI after 520 scans
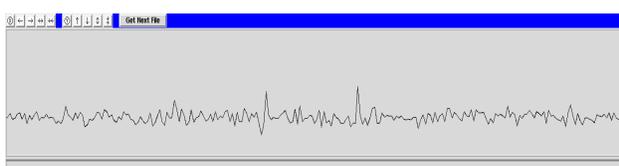
Figure 4g FFT after 560 scans
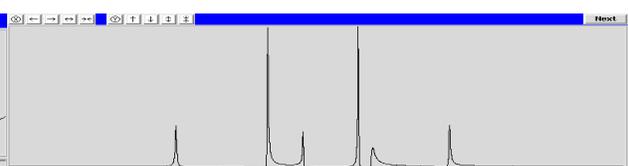
Figure 4g HI after 560 scans

This can be checked with the d values for these scans. Re-evaluation of the HI at N_{tr}=560 scans with K=4 shows a clear quartet with the expected 1:3:3:1 intensity pattern which is depicted in Figure 5.

Figure 5 Final HI of 560 scans with K=4 Denoised spectra

The final HI spectrum can now be compared to the FFT spectrum at 3200 transients (Figure 2). A clear 8 to 10-fold saving in data collection time has been demonstrated.

A question: What if 360 transients were all that could be collected? Do you believe that at least you have found the splitting correctly? Your call, just as one often does in the FT method.

Estimations made without strong indications of a gap need be treated very carefully to be believed. Consider for example the spectrum at 440 scans. Certainly do not accept feature 1 as true even if at this point it seems to phase well and seems stable. Use only features that make up good triplets quartets etc. as noise features are unlikely to mimic them. Don't set K=4 to get a final result as peak 1 is weightier than 2 and the latter would vanish.Considerations like these need to be takin in.

Conclusion

It has been demonstrated that the SVDHI method has a greatly improved sensitivity over the "traditional" FFT Method. It also yields high resolution results. For low sensitivity nuclei as ^{17}O, ^{15}N, ^{13}C etc, or for low concentrations of other nuclei and for problems where scale, that is dynamic range, is a problem, long collection times and/or enrichment may be avoided or reduced significantly.

Example Study 2

This sample is a solution of $100\mu l$ of natural abundance ^{15}N CH_3NO_2 in 1.0 ml $CDCl_3$ ($7.0\mu M$ of ^{15}N-CH_3NO_2). The spectra were recorded in a standard Pyrex 5mm NMR tube on a Varian 400MR spectrometer locked and without spinning.
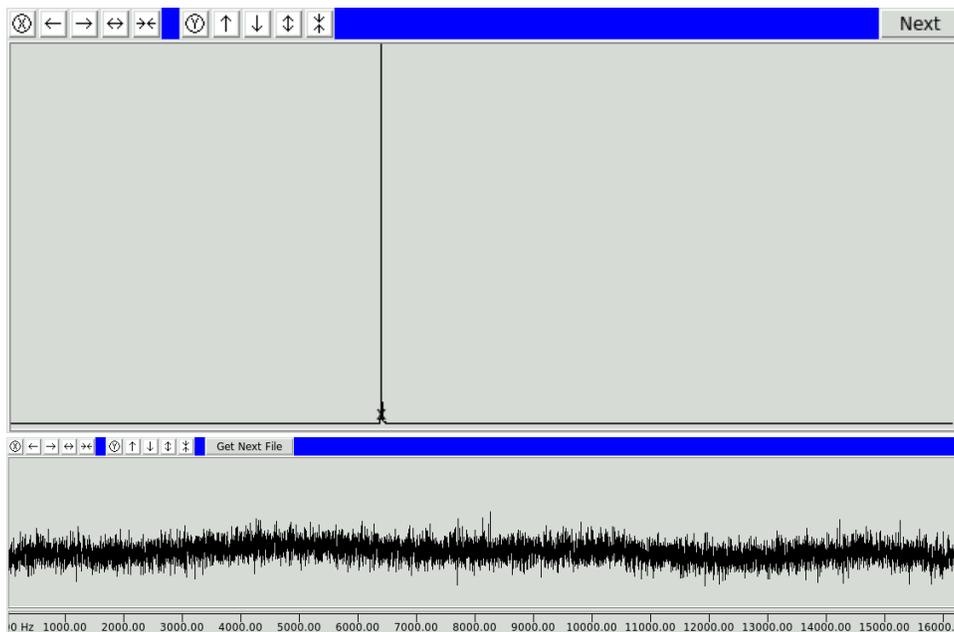
Figure 6 ^{15}N NMR spectra after 256 transients with a line broadening of 0.99 Hz and K=1. The upper is the HI spectrum, the lower is the traditional FT spectrum.
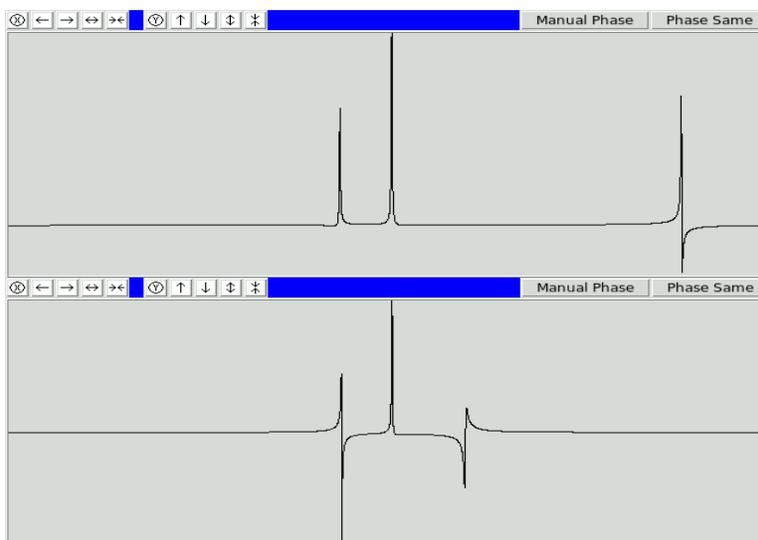
Figure 7 HI spectra with K=3: after 256 transients (upper) and after 128 transients (lower).
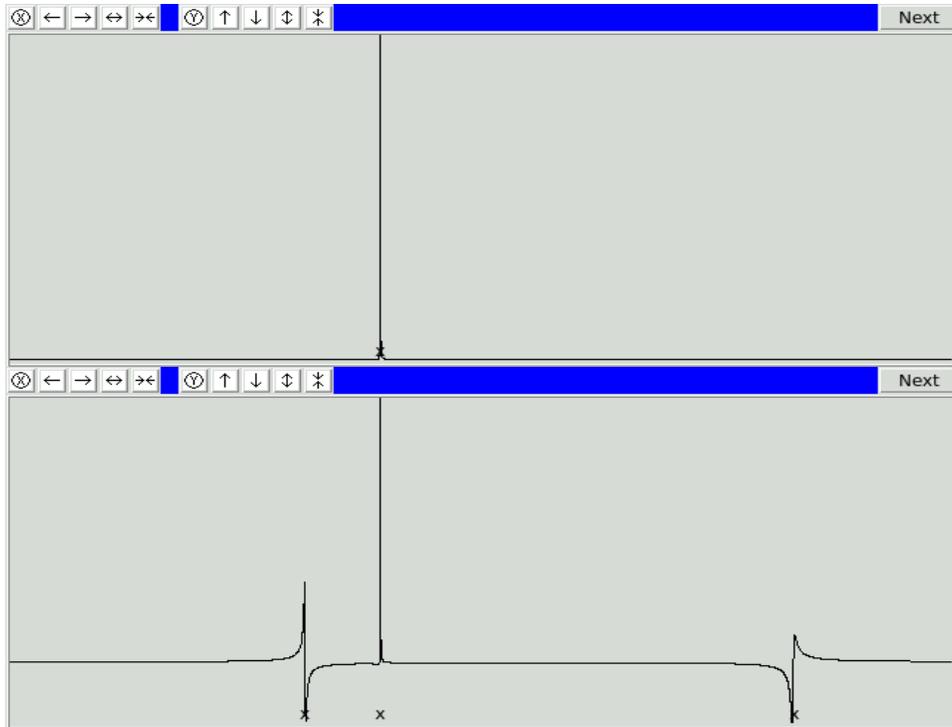
Figure 8 HI spectra after 256 transients: K=1 (upper) and K=3 (lower)
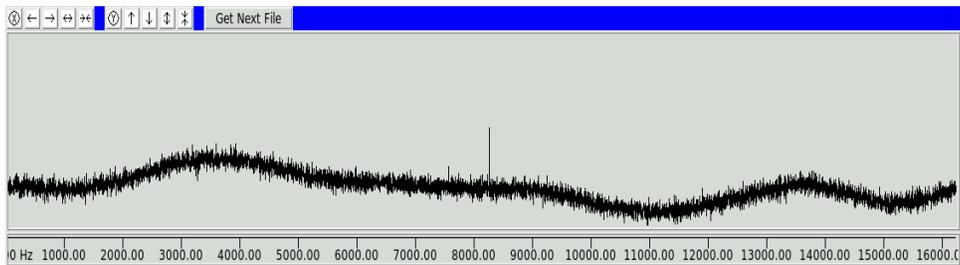
*Figure 9 Traditional FT spectrum after 2592 transients, LB=0.99Hz.
Signal at 8196Hz is an artifact.*
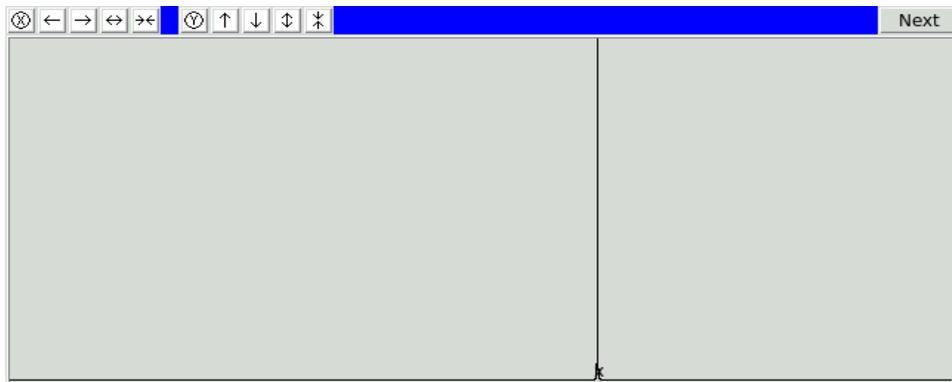
Figure 10 HI spectrum after 32 transients, LB=0.99Hz and K=1

13

Example Study 3

N_γ

N_β N_β

N_α N_α
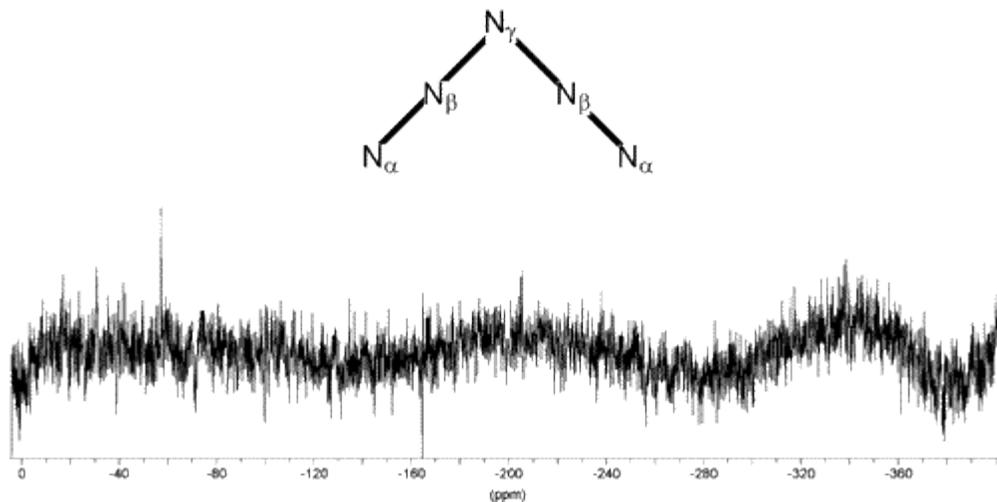
Figure 11   Natural Abundance ^{15}N NMR spectrum of $[N_5]^+[SbF_6]^-$ in aHF (c=0.19M) at $-20^\circ C$ after 11,600 transients (0.703 mM ^{15}N)
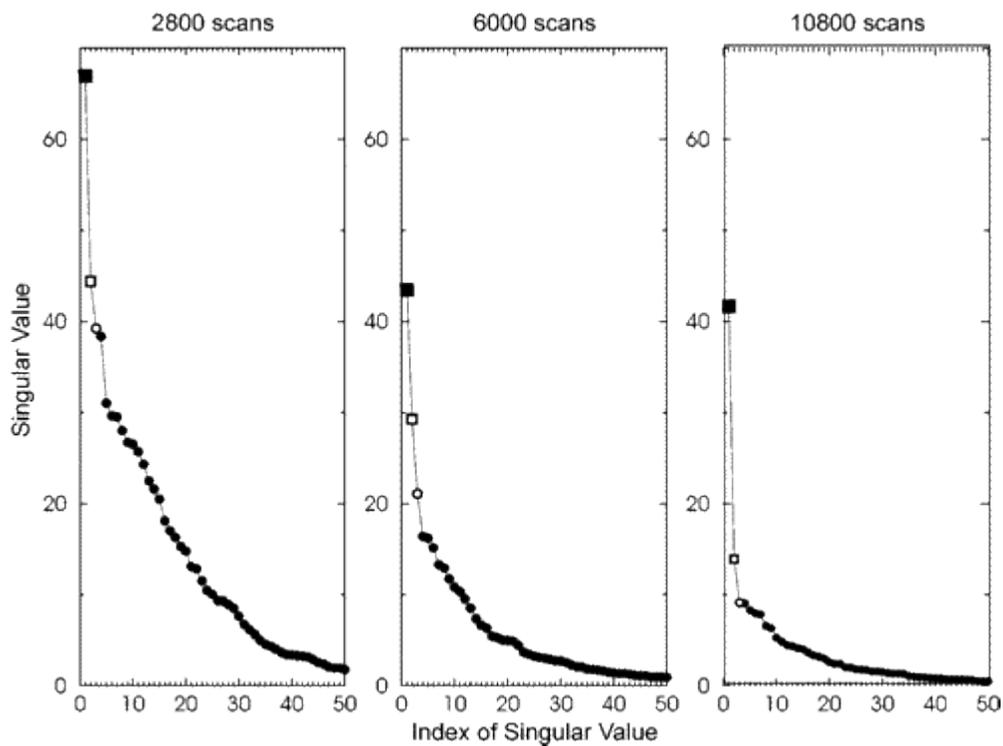
Figure 12   Distribution of the singular values (beta nitrogen) after different number of transients for the -145 to -185 ppm window
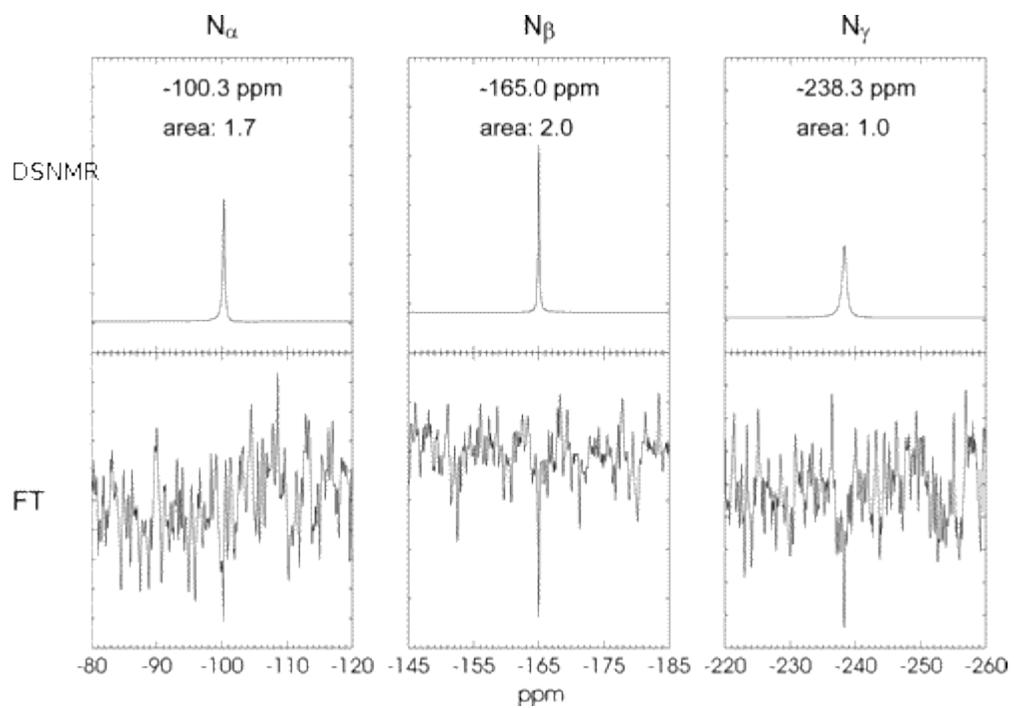
Figure 13 After 10800 scans

Example Study 4

31P example the *E* isomer of methyl α-(hydroxyimino)phosphonoacetic acid, dicyclohexylammonium (DCHA) salt (**I**)
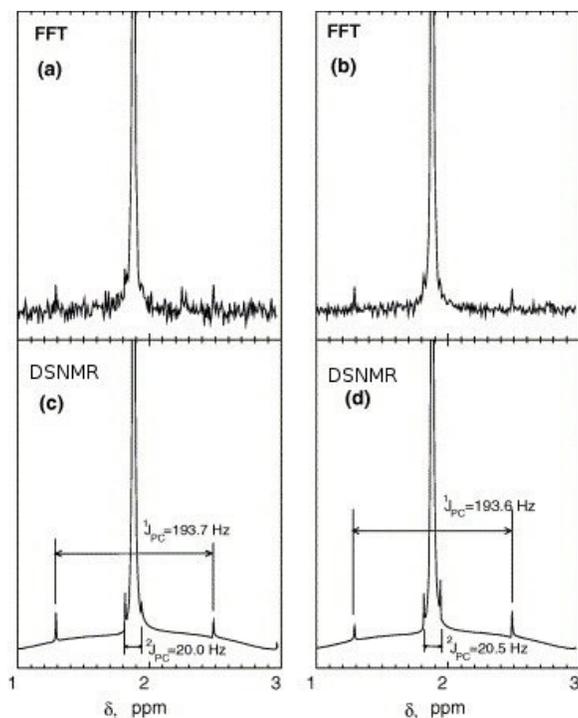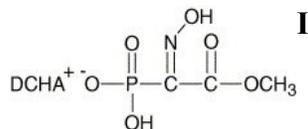
*Figure 14 ³¹P NMR spectra of **I**, 67 mM in methanol-d4/methanol(1/4). (a) FFT, after 0.5 h, (b) FFT, after 2 h, (c) DSNMR, after 0.5 h, (b) DSNMR, after 2 h.*
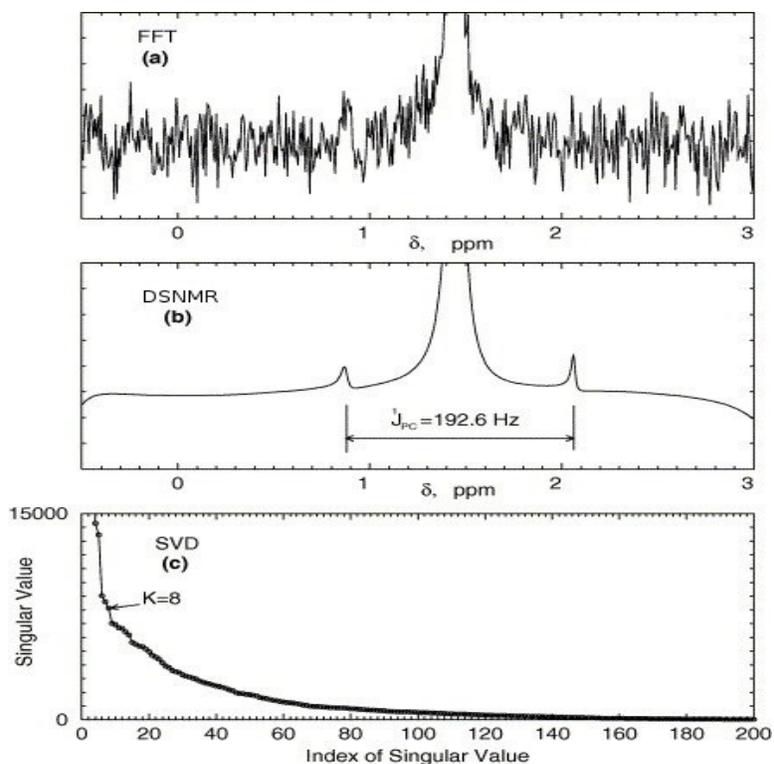
*Figure 15 ^{31}P NMR spectra of **I**, 8.4 mM in methanol-d4/methanol(1/4) N_{tr} = 40,000 (20 h). (a) FFT without damping. The one-bond ^{13}C P-C satellites are difficult to distinguish from random noise peaks with (not shown) or without damping, as predicted from the results in Figure 1 by the formula, $S/N \propto C \sqrt{N_{tr}}$ (S/N, signal to noise ratio; C, concentration; Ntr, number of transients (b) DSNMR-derived spectrum (c)Singular value curve corresponding to (b). The 8th singular value is indicated (K = 8); the top three SV's are above 15,000.*
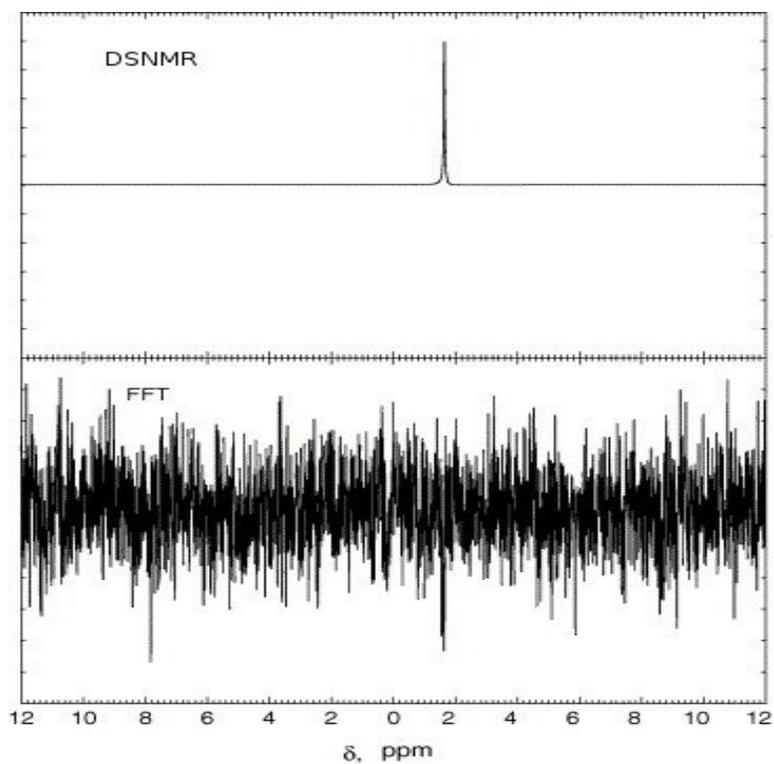
Figure 16 ^{31}P *NMR spectra of **I**, 0.058 mM in methanol-d4/methanol(1/4)*
N_{tr} *= 40,000 (20 h). DSNMR-derived spectrum (upper),FFT spectrum (lower).*

Example Study 5

An ^{17}O example.
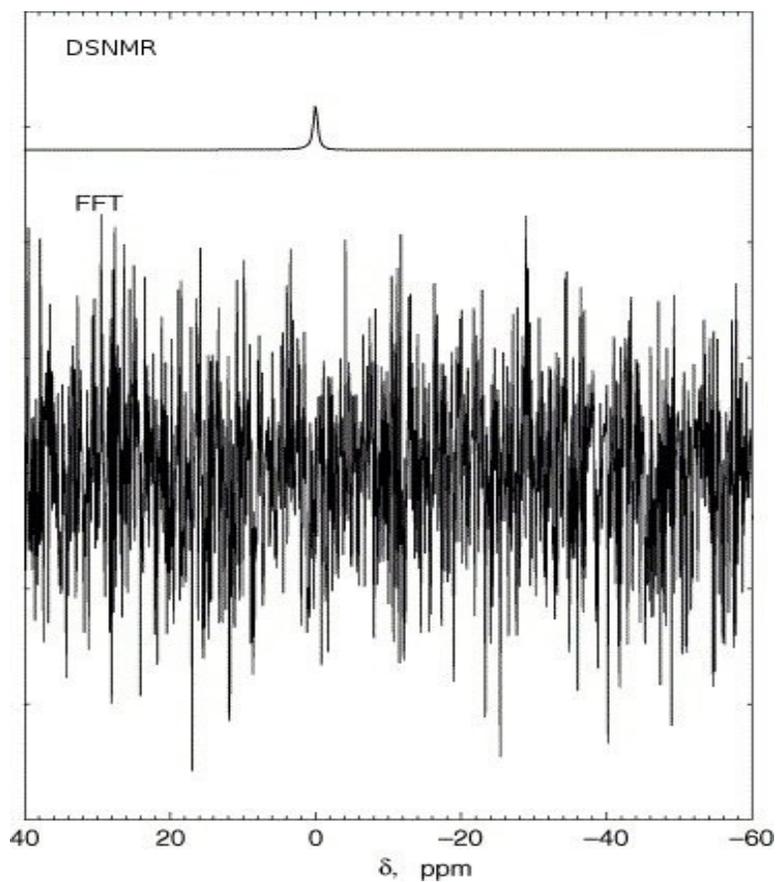
Figure 17 ^{17}O NMR spectrum of H_2O (natural abundance, 0.037%) in pyridine, concentration 170 mM (63μM of ^{17}O) at 92°C after 8000 scans. DSNMR spectrum (upper), FT spectrum (lower). The chemical shift of the observed DSNMR peak was arbitrarily set to 0.0ppm.
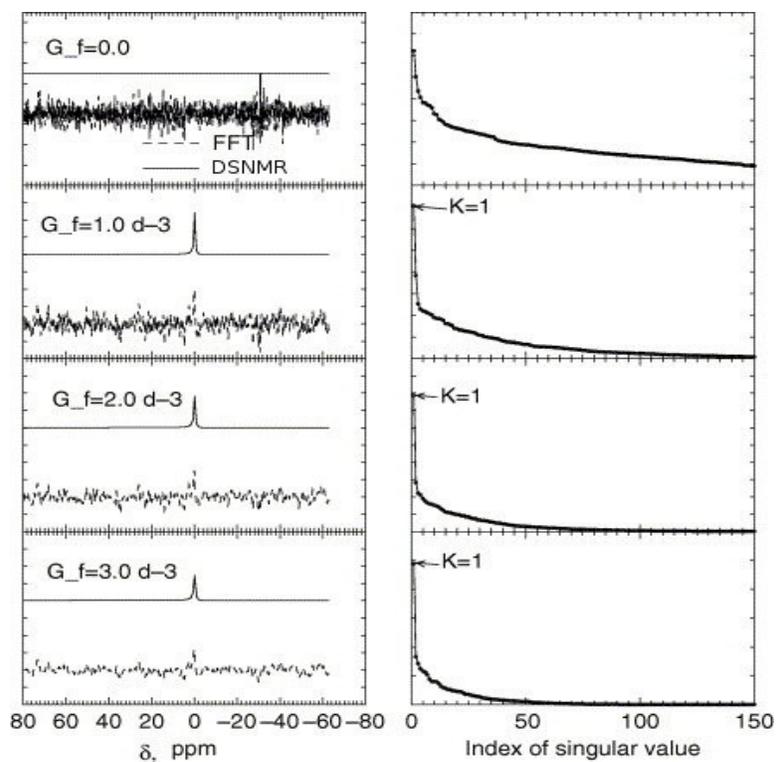
Figure 18 Comparison of FFT (lower, broken trace) and DSNMR(upper, solid trace) of ^{17}O NMR spectra (left) and the singular value plots (right) of H_2O in pyridine, concentration 68mM (25 μM of ^{17}O) at 92°C. N_{tr}=58,000 (18.3h).