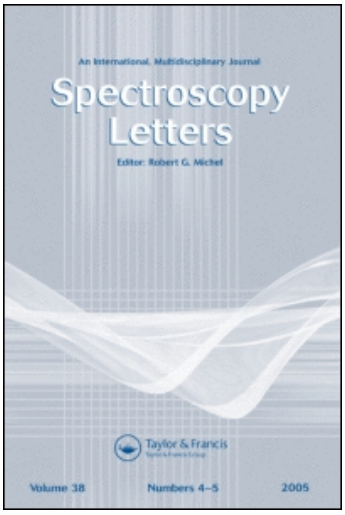


Spectroscopy Letters
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597299

Bayesian Estimation of NMR Spectral Parameters Under Low Signal-to-Noise Conditions

Ronald F. Evilia[a]; Rosemary Effiong[ab]; Scott L. Whittenburg[a]
[a] Department of Chemistry, University of New Orleans, New Orleans, LA [b] Department of Chemistry, Xavier University, New Orleans, LA

To cite this Article Evilia, Ronald F. , Effiong, Rosemary and Whittenburg, Scott L.(1993) 'Bayesian Estimation of NMR Spectral Parameters Under Low Signal-to-Noise Conditions', Spectroscopy Letters, 26: 8, 1559 — 1570
To link to this Article: DOI: 10.1080/00387019308011634
URL: http://dx.doi.org/10.1080/00387019308011634

SPECTROSCOPY LETTERS, 26(8), 1559-1570 (1993)

BAYESIAN ESTIMATION OF NMR SPECTRAL PARAMETERS UNDER LOW SIGNAL-TO-NOISE CONDITIONS

Key Words: Bayesian Estimation, NMR, Low Signal-to-Noise Ratio

Ronald F. Evilia, Rosemary Effiong* and Scott L. Whittenburg**

Department of Chemistry
University of New Orleans
New Orleans, LA 70148

* Present Address: Department of Chemistry
Xavier University
New Orleans, LA 70125

** Corresponding Author

Abstract The Bayesian statistical method of spectral estimation is applied to NMR free induction decay signals at various values of signal-to-noise ratio (SNR). The frequency and amplitude estimates from the Bayesian calculations are more accurate than those from the commonly used fast Fourier transformation (FFT) of the same data sets. Both real and synthetic data sets are examined with the Bayesian results being superior in all cases. In addition to the superior performance at low SNR the Bayesian derived amplitudes and frequency estimates were not as affected by signal decay as in Fourier Transformed spectra. Finally, the amplitudes obtained are equal to the FFT integrated intensities resulting in an apparent frequency domain signal-to-noise ratio (SNR) greater than the FFT SNR by a factor proportional to the FFT frequency domain linewidth. For typical high resolution spectra this improvement was approximately a factor of 2.5. Even greater improvement is obtained when rapidly decaying signals are analyzed. Bayesian computation time for the 6 line p-chloroanaline and chloroform spectrum was approximately 12 minutes on a modern computer work station.

Introduction

NMR Spectroscopy is well known to be insensitive relative to other forms of spectroscopy due to the nearly equal Boltzmann populations at the low energies characteristic of even the highest field instruments. Because of wide-spread interest in the information available from NMR spectroscopy, numerous attempts have been made to maximize the signal-to-noise ratio of NMR instruments and to extract the information as efficiently as possible.

The principal means by which SNR has been increased is through increases in the magnet field strength. Although increases in field strength result in increases in the Boltzmann distribution leading to signal increases that are approximately proportional to the square of the magnet field strength, non-ideal factors that become significant at higher frequencies limit SNR improvements to considerably less than the squared relationship predicts[1,2]. Indeed, the current state-of-the-art in magnet field strength is approaching a value at which non-idealities and relaxation effects will result in poorer performance at higher fields for ^{1}H spectra of even modest sized molecules[2,3,4]. In addition to the high field magnets, SNR improvements have been made by refinements in probe technology, RF circuit optimization and the use of very low noise preamplifiers. Along with the hardware improvements, a variety of sophisticated pulse techniques such as INEPT and HMQC have been employed for those situations in which the methods are applicable. These efforts and careful consideration of all the experimental and interpretation factors affecting SNR's have greatly improved the inherent sensitivity of the NMR technique compared to even a few years ago.[5,6,7]

The hardware, signal processing and pulse sequence developments are approaching the point of diminishing returns, however, and it is unlikely that future refinements in these areas will result in improvements comparable to those obtained in the past. While the data acquisition part of the NMR sensitivity problem has been well developed, the data analysis aspect has been less well developed. Recent advances in alternative computational methods have shown some promise for dramatic spectral improvements compared to Fourier transformation for spectral calculation[8-17]. These alternative computational methods rely on statistical theory and appear to have distinct advantages for extraction of signal from noise. Despite the promise that these approaches show, they have not received wide spread acceptance because of problems of artifact generation, they are computationally intensive and of the common belief that the Fourier transformed result is the "correct" result and cannot be improved.[8] Furthermore, the mathematical foundation and principles of these methods are complex and difficult to understand.

In this paper we report on the application of the Bayesian method of spectral estimation to NMR signals of low signal-to-noise ratio. The theory of Bayesian spectral estimation has

been elegantly developed and reported by Bretthorst and others [9-16]. In addition to Bayesian methods, other alternative computational approaches have been suggested such as the maximum entropy and maximum likelihood methods. Bayesian methodology has the advantage that prior information can be incorporated into the procedure. These methods have recently been reviewed and their application to experimental NMR and other forms of spectroscopy recently reported.[8,11,18-22] In this paper we will show that improved spectral estimation in both amplitude and frequency is obtained and that the calculations can be performed in a reasonable time period with a typical modern computer work station.

EXPERIMENTAL

Synthetic data sets were generated according to equations 1 and 2 for the real (in-phase) and imaginary (quadrature) signals at each frequency respectively:

$$f_R(t) = B_1\cos(\omega t)\exp(-\alpha t) + B_2\sin(\omega t)\exp(-\alpha t) \quad (1)$$

$$f_I(t) = B_2\cos(\omega t)\exp(-\alpha t) - B_1\sin(\omega t)\exp(-\alpha t) \quad (2)$$

1024 point in-phase and quadrature FID's were generated by incrementing t every 0.004 second corresponding to a 500 Hz spectral width if Fourier transformed. After the calculation of each time domain data point a random number generated by software was added to the data value. The random numbers were scaled to provide various peak-to-peak noise values. In all synthetic data sets reported in this paper B_1 was set equal to 1.00 and B_2 was set equal to 0.00 to simulate a perfectly phased data set. Because the model function contains both cosine and sine terms for both in-phase and quadrature channels the Bayesian results account for the phase of the signal. Signals with random phase are handled as easily as those with no phase error. Perfectly phased data sets were employed to make possible ready comparison between the Bayesian and Fourier transformed results without concern about proper phasing of the FT spectrum. The value of α used was varied as reported in the results and discussion section. Peak-to-peak noise values were varied from 0.20 to 4.0. All data sets were analyzed by the program ASAP which performed both Fourier transform and Bayesian analysis. For the analysis of the data reported here, Fourier transformation was performed first and the peaks observable in the FT spectrum used as initial estimates for the Bayesian calculation. This was done to speed the calculation. After selection of these initial estimates for peak locations, the Bayesian program computes the most probable frequencies and decay rates for the model function to fit the data using a simplex or conjugate gradient method. The amplitude for each frequency determined to be present in the FID is then back calculated. The resulting frequency, amplitude and decay estimates comprise the Bayesian "spectrum". In addition to the values for each of these parameters, a statistical uncertainty is reported that reflects the SNR of the data set. In the results reported below, the Bayesian output is compared to the frequencies and intensities obtained by examination of the FT spectrum. This

comparison is shown both graphically and in tabular form. By using a procedure that utilizes the FT as prior information the Bayesian calculation produces estimates for the frequencies, amplitudes and decay rates for the resonances obtained from the FT that are more accurate than possible from the FT plot, but it does not locate signals below the FT noise level. This limitation can be overcome by utilizing a Bayesian estimate at every possible frequency to locate signals below the FT noise at a considerable increase in computation time. The purpose of this study is to evaluate the accuracy of the Bayesian estimations for those cases where the FT results have high uncertainties because of noise or rapid signal decay.

The p-chloroanaline spectrum was acquired on a Varian UNITY 400 NMR spectrometer on an approximately 1% solution in deuterochloroform. The FID consisted of 20K real and 20K imaginary data points. The Fourier transform was performed on the first 4,266 of these data points zero filled to 8K and apodized with a decaying exponential function with decay rate 8.5 sec^{-1}. The Bayesian calculation was performed on only the first 1,024 data points without zero filling or apodization. The spectral width acquired was 2,941 Hz (0.34 msec sampling rate). The transmitter pulse width and power were set to very small values (approximately a 1° rf pulse) to generate a spectrum with a poor signal-to-noise ratio.

Calculations were performed on either a SUN Microsystems 4/65 (SPARC 1+ with 8 MB memory) or 4/75 (SPARC 2 with 32 MB memory) computer. The longest calculation time required for any of the spectra reported here was approximately 12 minutes for the p-chloroanaline spectrum.

Results and Discussion

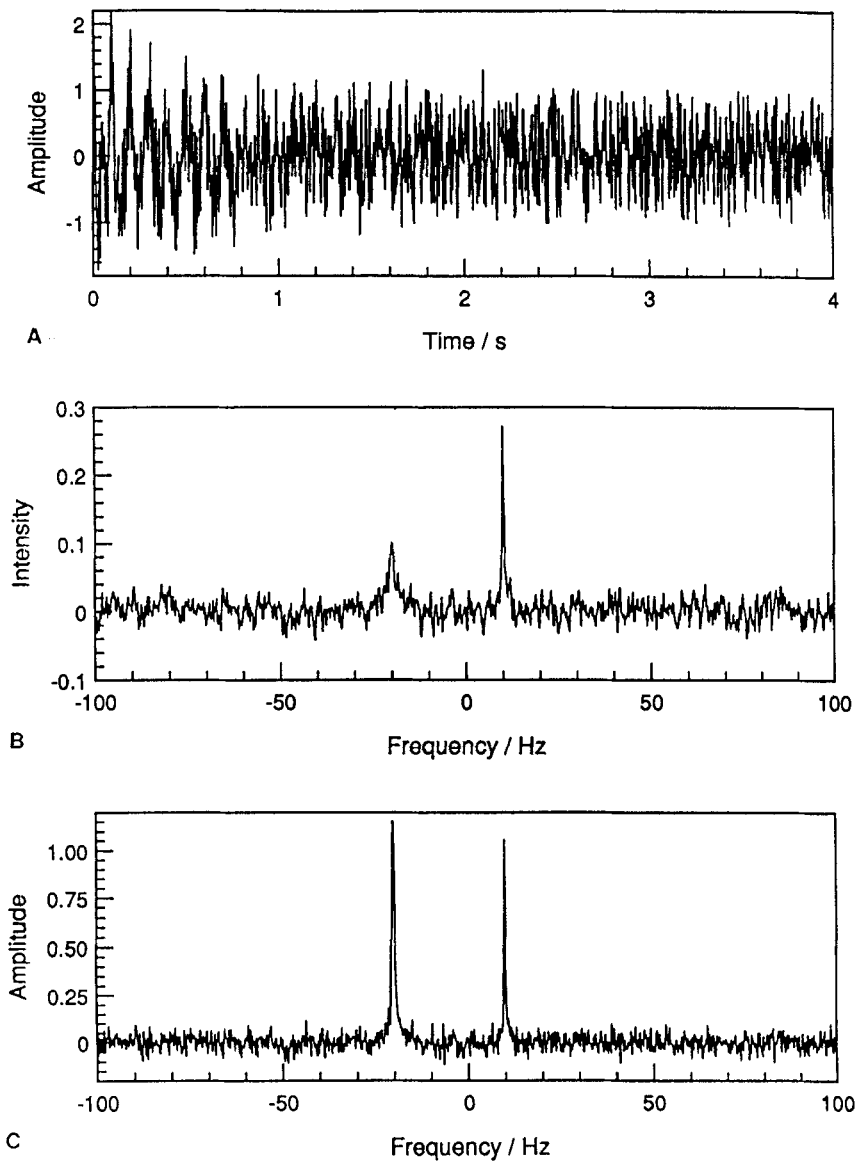
Synthetic Data Sets: To evaluate the potential of the technique data sets of varying SNR were analyzed. The synthetic data considered consisted of two exponentially decaying cosine waves having equal amplitudes but different decay rates. The data consisted of a -20 Hz signal with decay rate of 5 sec^{-1} and a signal at +10 Hz with a decay rate of 1 sec^{-1} where frequencies are relative to the center of the spectrum.

A direct Bayesian/Fourier signal-to-noise ratio comparison is not straightforward because the Bayesian estimate does not produce a spectrum *per se* but, rather, provides a table of probable frequencies, amplitudes, decay rates and statistical uncertainties in these quantities. For the operational mode used in the study reported here, all frequencies in the vicinity of the operator chosen values are evaluated regardless of their evidence level. For an explanation of evidence level see Bretthorst[16]. Frequencies chosen for evaluation that are not actually present in the data set result in calculated amplitudes whose error estimates place them in the noise level and decay rates that are very nearly zero with large uncertainties. Frequencies actually present

have small uncertainties in amplitudes and non-zero decay rates. All frequencies satisfying these requirements are optimized and reported. The Fourier transform frequencies and intensities are estimated by inspection of the frequency domain plot for comparison to the Bayesian calculated values. Since the Bayesian calculation does not require equally spaced data points while the FT does, the final Bayesian optimized value does not correspond to the FT estimate but is a more accurate value. The error in the FT frequency estimate results from the fact that only values corresponding to equally spaced frequencies across the spectrum are calculated by FT and, if the actual frequency does not happen to correspond to one of those points, an error is introduced. Obviously, the greater the number of data points used, the smaller this error is. Thus, one often "zero fills" in FT methods to reduce this source of error. All frequencies other than the ones found in the optimization are assigned zero intensity and, thus, there is no "noise", as such, in the Bayesian "spectrum". In the Bayesian analysis, increasing noise levels are manifested as increasing uncertainties in the values of the frequency, amplitude and decay parameters. In the spectra shown in Figures 1 and 2, the plotted Bayesian linewidths are equal to the uncertainties in the frequency estimates (1 standard deviation).

In order to compare the two techniques a noise level had to be designated for the Bayesian method. For the purpose of comparison, the Bayesian noise level was determined by subtracting the Bayesian determined signals from the experimental FID and performing a Fourier transformation on the resulting residual data. The noise level so obtained was found to be equal to that of the FT spectrum noise of the original data set if all of the resonances were found. Any resonances not found appear as peaks in the residual spectrum. For the cases reported here, the Bayesian calculation correctly found all of the resonances actually present and did not report any false positive results.

A typical two frequency data set is shown in Figure 1 along with its Bayesian and Fourier spectra. The Bayesian spectrum was constructed for display purposes by addition of the spectral lines to the residual noise evaluated as explained above. Examination of this figure indicates a greater SNR for the Bayesian spectrum because its signal intensity is greater than the FT intensity while the noise is the same. While no computation technique can actually change the SNR of the data, the apparent improvement here is the result of the fact that the FT intensity is inversely proportional to the decay rate while the Bayesian estimate is independent of decay rate. Thus, when shown as a plot, the Bayesian spectrum appears to have a greater SNR. An alternative explanation for this intensity difference is that the Bayesian analysis gives the true signal intensity at the assigned frequency while the Fourier transform "spreads out" the intensity over a substantial linewidth. Therefore, it is necessary to integrate the FT result to obtain the



1.) Bayesian/Fourier Comparison: Two Frequency Data Set. Peak-to-peak noise = 1.00. Frequency components -20.00 (amplitude 1.00, decay time 0.2 second); +10.00 (amplitude 1.00, decay time 1.00 second). A.) FID and apodization function. B.) Spectrum after Fourier transformation with exponential apodization with 0.2 second time constant. C.) Spectrum obtained by Bayesian estimation of unapodized data as described in text.

TABLE I

Two Frequency Data Sets, 0.2 and 1 Second Decay Times

P-P Noise	Frequency[2]	Amplitude[3]	Decay[4]	S/N B[5]	S/N F[6]
	-20.00	1.11	0.17	154	14
0.2	10.00	0.99	1.00	138	67
	-19.97	1.09	0.18	76	7
0.4	10.00	1.00	1.00	69	33
	-19.95	1.24	0.16	34	2
1.0	9.99	0.98	1.01	27	13
	—	—	—	—	—[7]
2.0	9.99	0.98	1.02	13	6
	—	—	—	—	—[7]
4.0	10.04	0.71	0.49	5	3

1) P-P Time Domain Noise Level.
2) Bayesian Determined Frequencies (correct values -20.00 and 10.00).
3) Bayesian Determined Amplitude (correct values 1.00).
4) Bayesian Determined Decay Rate (correct values 0.2 and 1.00 sec.).
5) Bayesian Frequency Domain Signal-to-Noise Ratio (see text).
6) Fourier Transform Frequency Domain Signal-to-Noise Ratio.
7) No Peak at -20Hz found above noise in FT spectrum.

true intensity of the signal, while Bayesian estimation yields all of the intensity at the correct frequency. Therefore, spectral integration is unnecessary.

One may wonder how an amplitude equal to the signal strength can be obtained if the calculation involves only probabilities. First of all, the amplitude of the signal is integrated out as a "nuisance parameter" (see Bretthorst[16] for a discussion of the theory of this procedure) and the most probable frequencies determined. In this way the evidence for the presence of a particular frequency is not dependent upon the amplitude of that frequency. Following assignment of the most probable frequencies, the value of B_1 and B_2 that best fit the data for each probable frequency are calculated and the signal amplitude determined by equation 3:

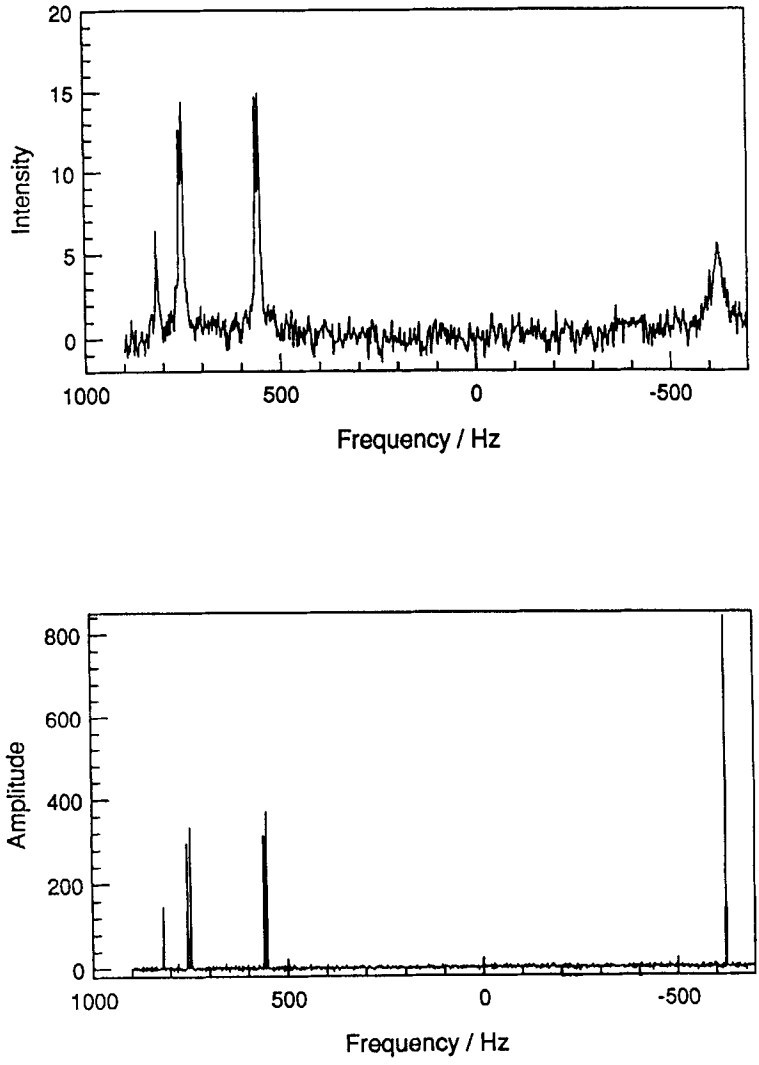
$$\text{Amplitude} = (B_1^2 + B_2^2)^{1/2} \tag{3}$$

Table I summarizes the results of the Bayesian/Fourier comparison for two frequency data sets on unapodized data. The Bayesian SNR was calculated by dividing the amplitude found by

the program by the Fourier transform peak-to-peak frequency domain noise level and multiplying by the usual factor of 2.5.[5] The Fourier transformed SNR was calculated from the maximum value in the data set evaluated by graphical estimation of the peak amplitude. As can be seen from an examination of these tables, the Bayesian results were consistently superior to the Fourier results. Although the Bayesian SNR, as defined above, exceeds that of the FT, it cannot "find" signals that are below the FT noise level in this implementation because of the requirement that the FT peak locations be used as prior information. In principle, Bayesian analysis can be performed at arbitrarily close frequency points in the spectrum and frequencies having evidence levels above some threshold value chosen for amplitude evaluation.[16] This type of implementation does not require FT prior information and should locate signals below the FT noise levels at a considerable increase in calculation time. All frequencies having evidence levels above 37 are considered to be present in the spectrum and their amplitudes back calculated in this work.

The effect of applying an apodization function to the data set prior to analysis is to further increase the Bayesian SNR relative to the FT value. This effect does not reflect superior performance of the Bayesian calculation, however, because the apparent advantage arises from the way in which we have defined the SNR in this study rather that any actual difference in the calculated results. Briefly, the Bayesian results are not improved by apodization. The amplitude found is the same whether the data set is apodized or not and the uncertainties are the same. Because the frequency domain noise level is decreased by the apodization, the Bayesian SNR as defined in this study, increases in proportion to the decrease in the noise level (but the uncertainties in the results are the same). The Fourier transformed SNR is improved by the decrease in noise level as well, but by a lesser factor because the linewidth is broadened by the apodization process resulting in a lower amplitude at the peak. Therefore, when the comparison is made with apodized data sets the relative benefit appears greater, but in reality, the reliability of the Bayesian result has not been improved while the FT result can be improved by a matched filter weighting function. Thus, in our experimental spectrum, we compare the Bayesian derived spectrum to the apodized FT spectrum. Obviously, matched filter apodization can not be done in cases where signals of differing decay rates are present. In that case some compromise value must be used or the FT performed multiple times to emphasize different decay rates.

Experimental Data Analysis: Figure 2 shows a comparison of the results obtained for the proton spectrum of p-chloroanaline taken under low signal-to-noise ratio conditions. This figure shows in dramatic fashion the superior frequency and amplitude estimating ability of the Bayesian calculation. First, consider the FT spectrum. If one wishes to know the chemical shifts



2.) Bayesian/Fourier Comparison: ^1H spectrum of p-chloroanaline in $CDCl_3$.

TABLE II

p-Chloroanaline Spectral Comparison

BAYESIAN ESTIMATE				FFT	
Line #	Frequency Hz	Amplitude	Decay Sec^{-1}	Frequency Hz	Peak Height
1	817.58 ± .08	146 ± 33	91.3 ± .20	818.1	6.4
2	756.83 ± .05	293 ± 40	13.3 ± .14	756.7	11.9
3	748.35 ± .04	335 ± 39	13.0 ± .12	750.2	13.3
4	560.27 ± .04	313 ± 37	11.5 ± .12	561.3	13.7
5	552.34 ± .05	369 ± 40	13.6 ± .14	553.4	14.2
6	-624.97 ± .05	888 ± 123	122 ± .14	-623.8	5.0

Errors in Bayesian estimate equal to one standard deviation. Bayesian amplitude and Fourier transform peak height in arbitrary units.

of the lines, they must be estimated from the plot and those estimates are effected by the overlap in the doublets at 560 and 750 Hz as well as the broadening of the signal at -625 Hz. Also, the intensities of the resonances cannot be estimated from the peak heights but, rather, require integration of the spectrum. Overlap within the doublets precludes accurate integration of their individual intensities. Finally, if one is interested in the decay rate of the individual lines, only the peak at -625 Hz is sufficiently resolved to allow measurement of its linewidth.

The Bayesian result yields accurate frequency estimates, not complicated by overlap and linewidth; accurate amplitudes for each line in the spectrum also not affected by overlap and accurate estimates of the decay rates of each of the spectral lines. Table II summarizes the quantitative comparison of the data in Figure 2. The small peak at 817 Hz is residual $CHCl_3$ in the deuterochloroform solvent.

Thus, it appears that Bayesian spectral estimation can be a viable alternative to FT for data analysis. At the current level of implementation, improved characterizations of resonances

observed in the FT spectrum are obtained with modest computation time. In future implementations it should be possible to extract and characterize signals that are not observable at all in the FT spectrum at an increased cost of computation time or to resolve closely spaced lines that are not resolved by FT analysis without degradation of the SNR.

The studies reported here involved simulations or data acquisition under conditions typically utilized in Fourier transform NMR spectroscopy. Because of differing requirements of the Bayesian method, alternative data acquisition schemes may be employed that will provide even greater improvement over the FT spectrum. For example, since no artifacts such as sinc wiggles are introduced, one need not continue an acquisition until the signal completely decays away allowing the use of smaller data sets. A future paper will examine the data acquisition ramifications of Bayesian spectral estimation.

The superior frequency and amplitude estimating ability suggest future applications to situations in which rapid signal decay introduces excessive line broadening in the FT spectrum or where the greater computational time is offset by decreased acquisition time of low SNR samples. Some examples of potential applications include spectra of quadrupolar relaxed nuclei such as ^{14}N, ^{17}O and ^{33}S where linewidths are a serious problem and studies of chemical systems undergoing chemical exchange at intermediate rates where extreme broadening can also occur.[25]

The ability to determine amplitudes from noisy signals suggests applications involving quantitative measurements of dilute components where the excessive acquisition time required for FT analysis compensates for the increased Bayesian computation time. Examples of applications to each of these problem areas will be reported in the near future.

Acknowledgements: Support of this work in part by NSF EPSCoR Grant EHR 9108765 is gratefully acknowledged. LEQSF grant (1990-91)-ENH-53 for purchase of the NMR spectrometer is also gratefully acknowledged. Presented in part at Pittcon'92, New Orleans, La., March, 1992.

References

1. Pople, J. A.; Schneider, W. G.; Bernstein, H. J. *High-resolution Nuclear Magnetic Resonance*, McGraw-Hill Book Co.: New York, 1959, Chapter 3.

2. Bothner-By, A. A.; Dadok, J *NMR and Biochemistry*, Opella, S. J.; Lu, P. Eds., Marcel Dekker: New York, 1979, Chapter 12.

3. Pitner, T. P.; Walter, R.; Glickson, J. D. *Biochem. Biophys. Res. Commun.* **1976**, *70*, 746-751.

4. McLaughlin, A. C.; Cullis, P. R.; Hemminga, M. A.; Hoult, D. I.; Raddu, G. K.; Ritchie, G. A.; Seeley, P. J.; Richards, R. E. *FEBS Lett.* **1975**, *57*, 213-220.

5. Ernst, R. R. *Advances in Magnetic Resonance*, Vol. 2, Waugh, J. S., Ed.; Academic Press: New York, 1966, Chapter 1.

6. Turner, C. J. *Prog. NMR Spect.* **1984**, *16*, 311-370.

7. Becker, E. D.; Ferretti, J. A.; Gambhin, P. N. *Anal. Chem.* **1979**, *51*, 1413-1420.

8. Stephenson, D. S. *Prog. NMR Spect.* **1988**, *20*, 515-626.

9. Gesmar, H.; Led, J. J.; Abildgaard, F. *Prog. NMR Spect.* **1990**, *22*, 255-288.

10. Jaynes, E. T. *Maximum Entropy and Bayesian Spectral Analysis Problems*; Smith, C. R.; Erickson, G. J., Eds.; Reidel, Dordrecht, The Netherlands, 1987.

11. Bretthorst, G. L.; Hung, C. -C.; D'Avignon, D. A.; Ackerman, J. J. H. *J. Magn. Reson.* **1988**, *79*, 369-394.

12. Bretthorst, G. L. *Maximum Entropy and Bayesian Methods*; Foregere, P., Ed.; Kluwer, Academic, Dordrecht: The Netherlands, 1989.

13. Bretthorst, G. L. *J. Magn. Reson.* **1990**, *88*, 533-551.

14. Bretthorst, G. L. *J. Magn. Reson.* **1990**, *88*, 552-570.

15. Bretthorst, G. L. *J. Magn. Reson.* **1990**, *88*, 571-595.

16. Bretthorst, G. L. *Bayesian Spectrum Analysis and Parameter Estimation*, Vol. 48; Springer-Verlag, New York, 1988.

17. Sibisi, S.; Skilling, J.; Brereton, R. G.; Laue, E. D.; Staunton, J. *Nature* **1984**, *311*, 446-447.

18. Meier, J. E.; Marshall, A. G. *Anal. Chem.* **1990**, *62*, 201-208.

19. Meier, J. E.; Marshall, A. G. *Anal. Chem.* **1991**, *63*, 551-567.

20. Doerschuk, P. C. *J. Opt. Soc. Am.* **1991**, *8A*, 1207-1221.

21. Doerschuk, P. C. *J. Opt. Soc. Am.* **1991**, *8A*, 1222-1232.

22. Whittenburg, S. L.; Evilia, R. F. *Paper P80, 32nd ENC Meeting*, St. Louis, Mo. April 1991.

23. Vines, K. S.; Evilia, R. F.; Whittenburg, S. L. *J. Magn. Reson.* **1992**, *100*, 195-201.

24. Grant, D. M.; Hirst, R. C.; Gutowsky, H. S. *J. Chem. Phys.* **1963**, *38*, 470-487.

25. Vines, S. K.; Evilia, R. F.; Whittenburg, S. L. *J. Phys. Chem.* In Press.

Date Received: May 17, 1993
Date Accepted: June 21, 1993