

# Development of 2D BTEM for 2D NMR analysis of multicomponent mixtures and reactive mixtures

Liangfeng Guo<sup>1</sup>, Anette Wiesmath<sup>2</sup>, Peter Sprenger<sup>3</sup> and Marc Garland<sup>1\*</sup>

<sup>1</sup>Department of Chemical and Biomolecular Engineering, 4 Engineering Drive 4, National University of Singapore, Singapore, 117576

<sup>2</sup> Institute of Chemical and Engineering Sciences, Singapore

1 Pesek Road, Jurong Island  
SINGAPORE 627833

<sup>3</sup> Bruker Biospin AG

Keywords: deconvolution, chemometrics, NMR

Prepared for Presentation at the 2004 Annual Meeting AIChE

Copyright © Liangfeng Guo, Anette Wiesmath, Peter Sprenger and Marc Garland, National University of Singapore

August 2004

Unpublished

AIChE Shall Not Be Responsible For Statements or Opinions Contained in Papers or Printed in its Publications

\* Tel : +65 6874 6617, Fax : +65 6779 1936, E-mail: [chemvg@nus.edu.sg](mailto:chemvg@nus.edu.sg)

## Abstract

Spectral reconstruction from multi-component spectroscopic data is the frequent primary goal in chemical system identification and exploratory chemometric studies. Various methods and techniques have been reported in the literature. However, few algorithms/methods have been devised for spectroscopy recovery without the use of any *a priori* information. In the present studies, a higher dimensional entropy minimization method based on the BTEM algorithm (Widjaja, E.; Li, C.; Garland, M. *Organometallics*, 2002, 21, 1991-1997.) and related techniques were extended to large scale arrays, namely, 2D NMR spectroscopy. The performance of this novel method had been successfully verified on various real experimental mixture spectra from a series of randomized 2D NMR mixtures (COSY NMR and HSQC NMR). With the new algorithm and raw multi-component NMR alone, it was possible to reconstruct the pure spectroscopic patterns and calculate the relative concentration of each species without recourse to any libraries nor any other *a priori* information. The potential advantages of this novel algorithm and its implications for general chemical system identification of unknown mixtures are discussed.

## 1. Introduction

In the last thirty years, NMR spectroscopy has experienced a dramatic development in both the sophistication of the instrumentation as well as in the signal processing. 2D NMR was proposed by Jeener<sup>1</sup> in 1971 and later demonstrated by Ernst *et al* leading to a tremendous increase in the capability of NMR and the subsequent explosion in experimental techniques for two dimensions. 2D techniques have been widely used for the analysis of structurally complex molecules, including the structural determination of bio-molecules such as proteins, peptides and nucleic acids.

The analysis of multi-component solutions, particularly those containing unknown constituents is a general problem in spectroscopy, and it presents numerous analytical difficulties. Considerable effort has been focused on the 1-dimensional FTIR, Raman, and NMR data using a variety of chemometric techniques. In addition, second-order blind identification has been used for both 1D and 2D NMR<sup>2</sup>. However, progress on really complex blind source problems<sup>3</sup>, has been severely limited due to the frequent need for some form of a priori system information such as an estimate of the number of species present, and/or the frequent intractability of systems containing more than circa 3 components. Recently, a new technique Band-Target Entropy Minimization (BTEM) was developed and applied to 1D spectra of multi-component systems, using FTIR<sup>4,5</sup>, Raman<sup>6</sup>, and MS<sup>7</sup>. The primary

utility of BTEM comes from (1) the fact that no *a priori* estimate of the number of species present is needed, (2) considerable noise reduction can be obtained and (3) its goal-oriented approach; the user targets a single spectral feature of interest, and the algorithm returns the full-range deconvoluted spectrum.

## 2. Experimental

The samples for NMR were prepared by dissolving varying amounts of 1,5 chloro-1-pentyne (Aldrich), 4-Nitrobenzaldehyde (Aldrich) and 3-Methyl-2-butenal (Aldrich) and topping with  $\text{CDCl}_3$  to achieve a total volume of  $500\ \mu\text{l}$ . 7 solutions for NMR measurements were prepared and each solution contained all 3 solutes. The approximate concentrations of solutes varied over the range 1.0-2.0 weight %. Consequently, the variation in solute concentrations from sample to sample can be considered quite low. 2 spectra of each solution were measured. The use of a constant liquid phase volume in all sample preparations was crucial in the quantitative aspects of this study.

## 3. Computation

### 3.1 Pattern Recognition and Information Entropy

In 1983, Sasaki, *et al* proposed a new approach based on entropy minimization.<sup>8</sup> The concept of signal entropy was first introduced by Claude Shannon<sup>9</sup>. Watanabe has mentioned the association between entropy minimization and pattern recognition<sup>10</sup>. It is now recognized that minimization of entropy is closely associated with the *principle of simplicity*<sup>11</sup>. The development of Band-Target Entropy Minimization was made possible after the implementation of a number of additional considerations including, the concept of band-targeting, signal enhancement and signal noise reduction, and experimental design. Although entropy expressions similar to those used by Sasaki *et al* were originally used, simplified spectral measures have been used recently to expedite computation, particularly for the analysis of very large data sets. In the present contribution, the objective function used is the modified form. However, the name “entropy minimization” has been retained due the original problem formulation and the fact that the goal remains a search for spectral simplicity.

### 3.2 The Formulation of 2D BTEM

The ultimate objective of 2D BTEM is to obtain accurate estimates of the mean pure component absorptivities. This is achieved by transforming the abstract  $V^T$  information into pure component absorptivity approximations  $\hat{a}_{v_{xv}}(\overline{T, P, x_i})$ , one estimate at a time. The computation can be performed on either the right singular vectors in  $V_{jx(v_{xv})}^T$  or the right singular matrices in the 3-array  $V_{jxv_{xv}}^T$ . It is convenient to refer to the two formulations as vector-wise 2D-BTEM and matrix-wise 2D-BTEM. Since some information is in a sense “lost” during concatenation, matrix-wise 2D-BTEM will prove to be the preferred computational route since higher quality spectral estimates generally result.

There are two issues about the optimal determination of the  $j$  unknowns : the first is the repeated evaluation of the entropy of the term  $T_{1xj} V_{jx(v_{xv})}^T$  or  $T_{1xj} V_{jxv_{xv}}^T$  and the second is the search for the final value of  $T_{1xj}$  such that the global entropy minimum is obtained. Consistent with the original formulation of Shannon entropy, a definition for the entropy  $H$  used in vector-wise 2D BTEM can take the form of Eq 1, where the probability distribution  $p$  is consistent with Sasaki et al’s original suggestion that a second derivative  $m=2$  of the spectral data can be used Eq 2.

$$H = \sum_v h_v = -\sum_v p_v \ln p_v \quad \text{Eq 1}$$

$$p_v = \frac{\left| \frac{d^m \hat{a}_v}{dv^m} \right|}{\sum_v \left| \frac{d^m \hat{a}_v}{dv^m} \right|} \quad \text{Eq 2}$$

In a similar sprite, we can now formulate a definition for the entropy  $H$  used in matrix-wise 2D BTEM Eq 3, where the term  $p$  takes into consideration the smoothness of the spectra in 2 dimensions Eq 4, 5.

$$H = -\sum_{v_1} \sum_{v_2} h_{v_1 v_2} = -\sum_{v_1} \sum_{v_2} p_{v_1 v_2} \ln p_{v_1 v_2} \quad \text{Eq 3}$$

$$p_{v_1} = \frac{\left| \frac{d^m \hat{a}_{v_1 v_2}}{dv_1^m} \right|}{\sum_v \left| \frac{d^m \hat{a}_{v_1 v_2}}{dv_1^m} \right|} \quad \text{Eq 4}$$

$$p_{v_1 v_2} = \frac{\left| \frac{d^m p_{v_1}}{dv_2^m} \right|}{\sum_v \left| \frac{d^m p_{v_1}}{dv_2^m} \right|} \quad \text{Eq 5}$$

The repeated evaluation of the log terms can require significant computational time. We have previously tested expressions for H which omit the log term, and the functions H' often yield good quality pure component spectra. Accordingly, for vector-wise and matrix-wise the following two expressions are used.

$$H' = \sum_v p_v \quad \text{Eq 6}$$

$$H' = \sum_{v_1} \sum_{v_2} p_{v_1 v_2} \quad \text{Eq 7}$$

The above mentioned functions H' are needed in the global optimization. Specifically, they are included in the objective function  $F_{obj}$  along with a penalty function P and a term for peak integration Eq 8. Further details concerning P and PeakInt can be found ref 4,5. The use of PeakInt in the objective function of 2D BTEM is often favoured. The minimization of the integrated estimate is consistent with spectral simplicity. The global minimization of the objective function is achieved using simulated annealing which is a stochastic search technique.<sup>12</sup>

$$\min F_{obj} = H + P + PeakInt \quad \text{Eq 8}$$

## 4.0 Results and Discussion

**4.1 HSQC Data** The 14 2D HSQC measurements form a 3-array of dimension  $A_{14 \times 1024 \times 1024}$ . The SVD on an array of this size can be problematic, even on a high end workstation with considerable RAM. In order to decreasing the computational burden, only 4 rectangular regions containing the real physical spectral features (peaks) were taken. About 90 percent of the original 2D spectral data contained did not contain useful physical information. The small rectangular regions were assembled into a new concatenated data array  $A_{14 \times (539 \times 107)}$ . SVD was performed to obtain the 14 meaningful right singular vectors in  $V^T_{14 \times (539 \times 107)}$ . Concatenation was undone. Both vector-wise and matrix-wise 2D-BTEM were performed using  $V^T_{7 \times 539 \times 107}$  by targeting observable features in the 7 visualized right singular matrices. Exhaustive searches produced only 3 2D spectral patterns. These patterns were then

imbedded into matrices with 1024X1024 channels. The results of the matrix-wise 2D-BTEM after imbedding, to yield the 2D HSQC estimated pure component absorptivities  $\hat{a}$ . The spectral estimates appear quite good when compared to authentic experimental references. Moreover, the dual problem for relative concentrations can be solved. The calculated concentration profiles are very similar.

#### 4.2 COSY Data

The 2D COSY measurements were analyzed in a similar manner. As only 12 samples were available, the dimension of the 3-array was  $A_{12 \times 1024 \times 1024}$ . In order to decreasing the computational burden, only 8 rectangular regions containing the real physical spectral features (peaks) were taken. About 90 percent of the original 2D spectral data did not contain useful physical information. The small rectangular regions were assembled into a new concatenated data array  $A_{12 \times (468 \times 150)}$ . Similar approach was executed, and the result appear quite satisfying.

#### 5. Conclusion

An advanced entropy minimization based algorithm 2D-BTEM has been proposed and verified with real experimental data from multi-component COSY and HSQC 2D nuclear magnetic resonance spectroscopy. The quality of recovered spectra is good when compared with authentic experimental references obtained from pure component measurement.

#### References

- 
- <sup>1</sup> Jeener, J.; Ampere Int. Summer School 11. Basko Polje, Yugoslavia, 1971
  - <sup>2</sup> Nuzillard, D.; Bourg, S.; Nuzillard, J.M. *J. Magn. Reson.* 1998, 133 (2), 358-363.
  - <sup>3</sup> Hua, K.B. *Circ. Syst. Signal. Pr.* 2002, 21 (1), 91-108.
  - <sup>4</sup> Chew, W.; Widjaja, E. and Garland, M. *Organometallics*, 2002, **21**, 1982-1990
  - <sup>5</sup> Widjaja, E.; Li, C. & Garland, M. *Organometallics*. 2002, 21, 1991-1997.
  - <sup>6</sup> Ong, L.R.; Widjaja, E.; Stanforth, R. & Garland, M. *J. Raman Spectrosc.* 2003, **34**(4), 282-289.
  - <sup>7</sup> Zhang, H. J.; Garland, M.; Zeng, Y. Z. & Wu, P. *J. Am. Soc Mass Spectrum*, 2003, 14, 1295-1305.
  - <sup>8</sup> Sasaki, K.; Kawata, S.; Minami, S. *Appl. Opt.* 1984, 23, 1955-1959
  - <sup>9</sup> Shannon, C.E. *Bell System Tech. J.* 1948, 3, 379-423.
  - <sup>10</sup> Watanabe, S. *Pattern Recognitio.* 1981, 13, 381-387.
  - <sup>11</sup> Kanpur, J. N. *Maximum-Entropy Models in Science and Engineering*. Wiley Eastern Ltd: New Delhi, 1993; p 3.
  - <sup>12</sup> Corana, A.; Marchesi, M.; Martin, C. and Ridella, S. *ACM. Trans. Math. Softw.* 1987, 13, 262-280