

Use of Neural Networks for Quantitative Measurements in Ion Mobility Spectrometry (IMS)

Zvi Boger* and Zeev Karpas

Nuclear Research Center-Negev, P.O. Box 9001, Beer-Sheva, Israel 84190

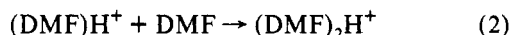
Received June 24, 1993*

The use of a neural network (NN) to derive quantitative information from the mobility spectra of dimethylformamide, bromine, and hydrogen fluoride is described. Unlike traditional methods, where the areas or heights of known and identified peaks are used for calibration, employment of NN does not require detailed knowledge of the ion chemistry of the measured system. Efficient NN training algorithms were employed, allowing quick learning with hundreds of input channels. It was also found that the size of the required training set was much smaller than expected from theoretical considerations, due to the interactions between input channels. Thus, this technique may be easily implemented in online or offline analysis of IMS spectra.

INTRODUCTION

Ion mobility spectrometry (IMS) is a sensitive method that can be applied to monitoring pollutants, at sub-parts per million (sub-ppm) concentrations, in ambient air.¹⁻⁸ The technique is based on atmospheric pressure chemical ionization of pollutant molecules in the reaction region of the drift tube. A packet of ions is then injected into the drift region of the tube. As the ions drift through a buffer gas in an electric field, they are separated according to their mobilities.¹⁻⁴ Complex sequences of ion-molecule reactions, which are not always clearly understood, precede ion separation, so that interpretation of mobility spectra to derive quantitative information is generally not straightforward. IMS is particularly suitable for compounds that have a high proton affinity, like organophosphorus compounds, amines, or ketones, or compounds that have a high electronegativity, such as nitro compounds and halogenated chemicals. The former group readily forms positive ions, generally the protonated molecule, while in the latter, negative ions are readily formed.

For example, dimethylformamide (DMF) molecules, like most amines and amides,⁹ readily undergo protonation in the reaction region of the IMS. The reaction proceeds mainly through proton transfer from reactant ions, such as H_3O^+ , and its clusters which are formed in moist air, thus forming protonated DMF ions (reaction 1), designated here as (DMF)- H^+ . The number of (DMF) H^+ ions formed increases initially with the concentration of the DMF in the sampled air, but beyond a certain level, further ion-molecule reactions occur, in which a protonated DMF dimer, (DMF) $_2\text{H}^+$, is formed through reaction 2. As the intensity of the peak due to (DMF)-



H^+ ions in the mobility spectrum begins to decrease, a peak due to (DMF) $_2\text{H}^+$ gradually appears and becomes the dominant peak in the spectrum. Thus, a method for determining the DMF concentration from the mobility spectrum, that relies on correlation with the intensity of the protonated DMF peak, would be valid only at lower concentrations and seriously in error at concentrations in which reaction 2 is not negligible. The above example is of a relatively simple system, the gas-phase ion chemistry of which is understood.

Neural networks (NN) have been applied to identify correlations in complex systems, in which isolation of factors and understanding how they affect the apparent behavior of measurable parameters are difficult.¹⁰ An attempt to combine NN methods for interpretation of IMS data has been reported,¹¹ but few details were given. In another work,¹² we have shown that neural networks can be used to process signals from a noisy spectrum to identify ions in mobility spectra. In a recent work, Bell *et al.*¹³ have described the application of a method based on a connectionist hyperprism neural network to interpret a large data set of IMS spectra. Their network had high-speed training and flexibility to adapt the training parameters to minimize undesirable false negatives, but had a sensitivity to redundant spectral information.

In this work, we present a demonstration of the capability of fast NN techniques to derive quantitative information from IMS data.

EXPERIMENTAL SECTION

The Ion Mobility Spectrometer. The mobility measurements were carried out on the Rotem Prototype PT-IMS instrument, described in detail previously.¹⁴ The drift tube consisted of segmented guard rings made of stainless steel, divided by alternate Teflon rings. The total length of the drift region, in which the ions were separated, was 5.56 cm. A focusing electric field of about 250 V cm^{-1} was applied along the entire cell to move the ions toward the center of the drift tube. The samples were diluted in air, and air was also used for both carrier and drift gas streams. The analog output from the electrometer at the far end of the drift region was processed by an IBM/PC compatible, with the aid of a National Instruments A/D interface and customized software (written by A. Matmor). To enhance the signal to noise ratio, 64 spectra were averaged before being displayed or saved for further processing.

The Dynamic Dilution System and Sample Preparation. The homemade dynamic dilution system used to generate the desired concentration of the pollutant in the air was also described previously.¹⁴ Briefly, it consisted of a 25-L mixing vessel, through which a fixed and regulated stream of air flowed. A known amount of the desired analyte could be injected through a septum into a port upstream of the vessel. The air stream flowing out of the vessel was split so that a stream with a fixed flow rate entered the IMS reaction region while the superfluous flow was exhausted into a hood.

* Abstract published in *Advance ACS Abstracts*, March 15, 1994.

Typically, the total flow rate through the vessel was 600 mL/min, of which 150 mL/min was introduced into the IMS and the remainder exhausted through a fume hood.

The DMF and the bromine used in this study were commercially available (Merck). A few milliliters of the liquid sample were injected into an ampule equipped with a valve and a silicone septum. The sample was frozen in a liquid nitrogen bath, and the ampule was evacuated with a rotary pump. After three freeze-pump-thaw cycles, the sample was allowed to warm to room temperature. A calibrated gas-tight syringe was inserted through the septum and a vapor sample drawn. This sample was then introduced into the dynamic dilution system described above through the injection port. The quantity of vapors introduced was calculated from the vapor pressure of the liquid sample and the injected volume.

Gaseous hydrogen fluoride (Matheson) was mixed with purified air in a stainless steel bulb equipped with a septum and a valve, so that a 10% (by volume) mixture was obtained. Use of glass vessels was avoided when working with hydrogen fluoride in order to eliminate reactions of HF with the vessel.

The ratio between the volume of the dilution vessel and the volume of the air stream during a 4-min measurement was about 10:1, when the flow rate was 600 mL/min. For the calculation of the sample concentration, it was therefore assumed that the concentration of the pollutant remained approximately constant during a measurement and that dilution effects were negligible during a set of experiments.

DESCRIPTION OF NEURAL NETWORK LEARNING

Neural network ideas were developed in the 1950's but went into decline in the 1960's when it was proved that the linear architecture used at the time was unable to solve nonlinear problems. Only with the advent of new network architectures and teaching algorithms in the mid-1970's and early 1980's did their true application potential become evident. As a recent detailed review of NN theory and applications to chemistry is available,¹⁰ only a brief description of the neural network concepts will be given here. The rigorous name of NN is "artificial neural network", showing the similarity of concepts to the neural networks of our brain cells whose axons and dendrites are connected with each other through synapses. In the computer, processing nodes (neurons) are linked to each other by variable strength connections through use of software or special hardware. The sum of all the inputs to a neuron, if larger than a set bias, activates the neuron. The neuron output, multiplied by the connection weights, is transferred to all neurons receiving inputs from this neuron. Although many neural architectures are possible, the most common one is the forward network with one input layer and two processing layers. One of these is called the "hidden layer" and the other one is the output layer. The sigmoidal-shaped activation equation usually used by a processing node is

$$y = 1/(1 - e^{-x}); \quad x = b + \sum w^*y_p \quad (3)$$

where y_p values are the outputs of the previous layer neurons. w is the connection weight leading to a neuron with b bias, and y is the neuron output. The most common learning algorithm is the supervised error back-propagation algorithm, in which a data set of system inputs and outputs is presented to neural net having initial connection weights. The difference between the network output and the known output is the network error, and the connection weights are modified to decrease the sum of the squared errors, as shown schematically in Figure 1. The process is repeated, until the neural network

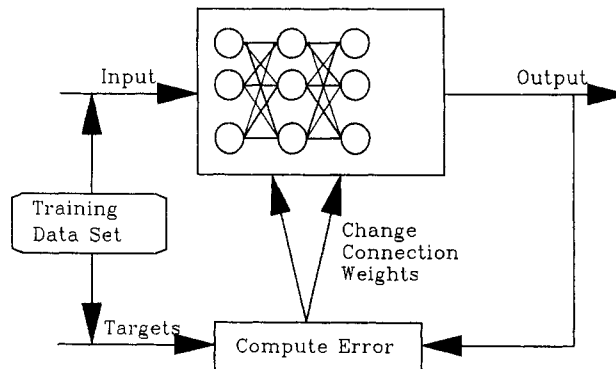


Figure 1. Schematic diagram of neural network training.

is considered to be "taught", as tested by a previously unseen data set which gives an acceptable small error.

Although the error back-propagation method is effective, the learning process is slow even on fast computers, especially when the NN contains many inputs and outputs. The reasons are that random numbers are chosen as initial connection weights, and there is no exact rule to specify how many hidden neurons are needed (the number of input and output neurons is fixed by the problem definition). Thus, the NN may easily converge to a local minimum in an n -dimensional space, and even if not, this has to be verified by starting over with new random initial weights and maybe a different number of hidden neurons. It was because of this slow learning that NN applications were limited to small- to medium-sized systems.

Several methods were recently proposed to overcome this limitation in processing multichannel data typical of instrumental analysis. Bos *et al.*¹⁵ used parallel computer hardware (transputer) and special software to reduce the training time. However, they still needed several thousand epoches (presentations of the data set) to train a relatively small neural network of 24 input channels, 4 neurons in the single hidden layer, and one output channel (a 24-4-1 NN) in 1-2 h. Borggaard and Thodberg¹⁶ transformed NIR data by Principal Component Analysis (PCA) and presented only the major coefficients to the NN, which also contained direct connections to the output neurons. They succeeded to train a neural network in a relatively short time (10-30 neurons on a 20-MHz personal computer). However, the use of PCA precludes knowledge acquisition from the trained NN, as no direct relation between the results and original data exists.

An algorithm developed by Guterman¹⁷ appears to solve the problem of training time. PCA technique, used to determine the number of major orthogonal axes in the data hyperspace, is applied to the training data. As the number of neurons in the hidden layer is believed to match the number of "concepts" in the data, the PCA analysis results determine the suggested number of hidden layer neurons. A further proprietary statistical analysis algorithm gives a good first guess for the values of the initial connection weights. A learning speed improvement by a factor of 20-50 over conventional learning is achieved by this algorithm, as it almost always converges to a global minimum in a relatively small number of data set presentations during the learning process.

The generality of learning, i.e., the ability of the NN to predict correctly system outputs from new inputs, depends theoretically on the ratio of learning cases to the number of connections in the NN, otherwise "over-fitting" will occur. Thus, a large NN should require a large number of training cases, which are not always available, to achieve reliable learning. Even if this requirement is met, the resulting NN may contain a considerable number of small connection

weights, which obscure the important relations that may be learned from the NN. Several methods of eliminating the weak connections without sacrificing NN accuracy were reported, but these methods require retraining after each connection is eliminated and thus are not practical for large NN. An algorithm that analyzes the statistical relative importance of each input to the NN was developed by Boger *et al.*¹⁸ and applied to various industrial modeling problems.¹⁹⁻²⁰ In these papers it was shown that the teaching of a reduced-size NN from the significant inputs only leads to NN with better accuracy and reliability.

The procedure for analysis of the IMS spectra was as follows: The whole mobility spectrum was preprocessed to give an array of vertical-scale-independent data. Usually about 80% of the data was randomly chosen for the training set, while the remaining 20% was left for testing the taught NN. The teaching algorithm gave the number of hidden neurons and the initial weight set, then proceeded to train the NN by the conjugate gradient error back-propagation method.²¹ Once the NN error was below the required criterion, the test data set was used to check the generalization capacity of the NN. If the test error was acceptable, the reduction algorithm was applied to the inputs, selecting the most significant inputs for further training. The teaching, testing, and reduction steps were repeated while the learning or testing error was acceptable (within a few percent). The reduced NN outputs were postprocessed to reverse the preprocessing and the results plotted.

The training was done in an IBM 386/33 compatible machine, and up to several hundred epoches were needed to train each NN. The whole training and reduction session typically took about less than 1 h, using the nonoptimized development version of software. A software package that incorporates the new algorithms mentioned in these paragraphs, operating in Windows 3.1 environment, is now available.²²

RESULTS AND DISCUSSION

Dimethylformamide (DMF). The gas-phase positive ion chemistry of DMF, under the atmospheric pressure ionization conditions pertaining to the IMS, was described above. The mobility spectrum of DMF is shown in Figure 2, with two peaks at drift times of 5.9 and 7.8 ms corresponding to the protonated DMF monomer and dimer, respectively. The relative intensities of the two peaks changed with the DMF concentration. At a concentration of 25 ppb (Figure 2a), the protonated monomer was the dominant peak in the mobility spectrum. As the concentration of DMF was increased to 44 ppb (Figure 2b), the intensity of the protonated dimer was augmented, and at 109 ppb, the protonated dimer dominated the spectrum, while the monomer peak was not noticeable (Figure 2c). Conventionally, a quantitative calibration curve may be prepared either from measurement of the integrated peak areas or, if the peaks have a Gaussian shape, by determining the peak heights. In Figure 3 the peak areas of the protonated monomer, protonated dimer, and the sum of the two (traces a, b, and c, respectively) are plotted as a function of the DMF concentration. Thus, using traditional methods, a reasonably good correlation was obtained between the peak areas in the mobility spectra and the DMF concentration, as evident from traces b and c in Figure 3.

Application of the neural network to interpret the mobility spectra was done according to the procedure described in the previous section. Eight DMF mobility spectra in the 25–125-ppb range were used. Seven spectra served as the training

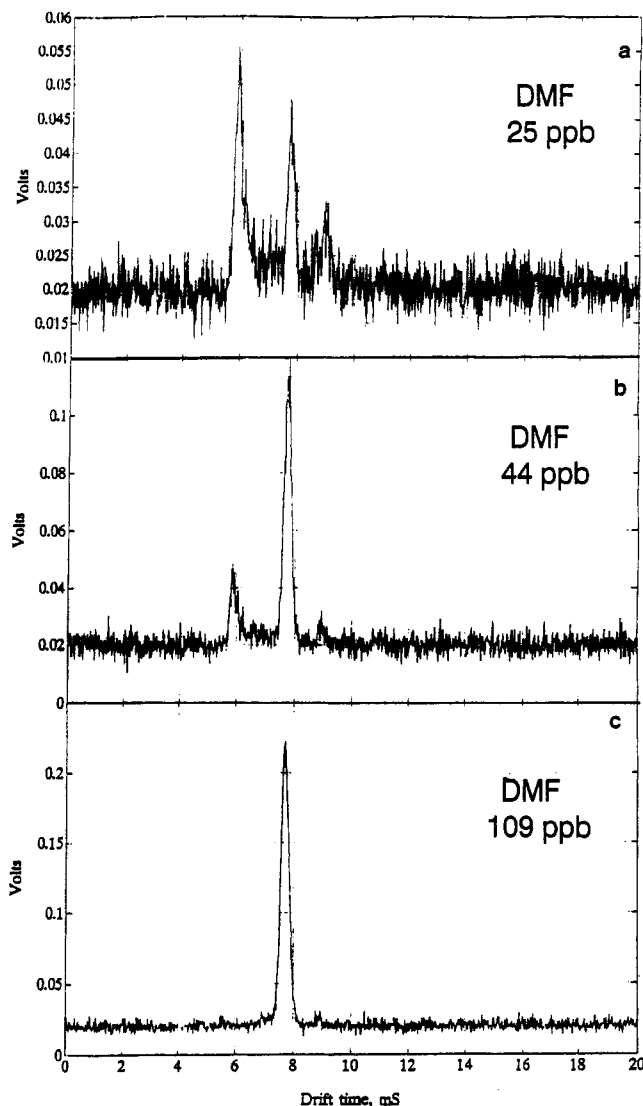


Figure 2. Mobility spectra of dimethylformamide (DMF) in air at concentrations of 25, 44, and 109 ppb, shown in a, b, and c, respectively. Note the protonated monomer and dimer peaks at drift times 5.9 and 7.8 ms, respectively.

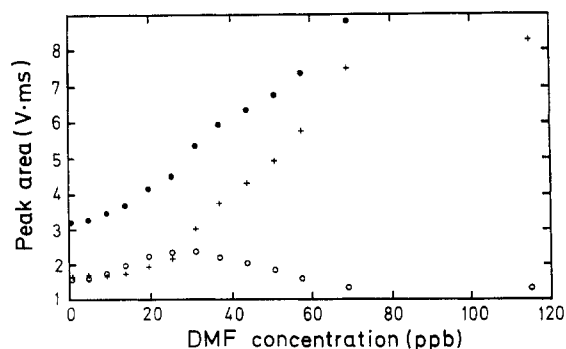


Figure 3. Peak areas of protonated DMF monomer (O, trace a) and protonated dimer (+, trace b) and the sum of the two ions (●, trace c) as a function of the DMF concentration.

set and the remaining one as the test point. Five NN reduction cycles were carried out, and only 14 of the 200 original input channels remained in the final 14-2-1 NN. The results are presented in Figure 4, with both the training set data and the test point compared with the known DMF concentrations. The processed mobility spectra are shown in Figure 5, where it is evident that the prediction error was quite small and that most of the information was retained in channels 371, 375, 386, 389, 394, and 396 (denoted 5–10 in Figure 5).

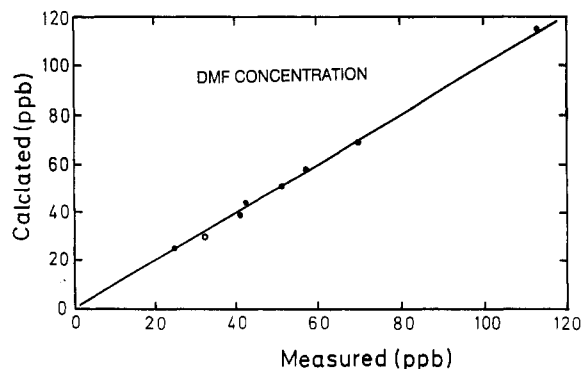


Figure 4. Results of neural network processing of the DMF mobility spectra plotted versus the calibrated concentration in the range 25–125 ppb. ● represents the training set and ○ the test points.

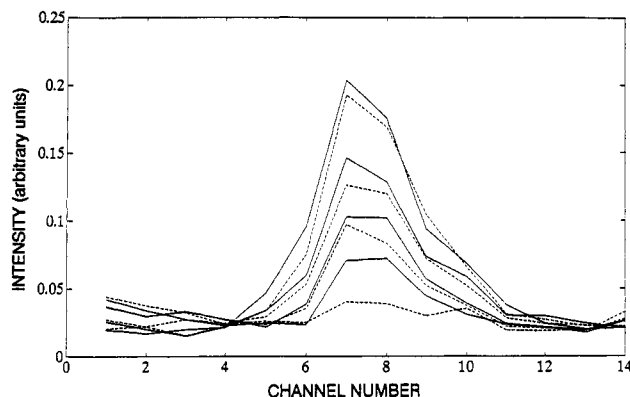


Figure 5. Processed ion mobility spectra of DMF in the 25–125-ppb range showing only the outputs of the channels selected by neural network analysis. Channel numbers 1–14 in this plot correspond respectively to channel numbers 301, 308, 311, 342, 371, 375, 386, 389, 394, 396, 401, 406, 417, and 452 in the original mobility spectra.

Bromine. The gas-phase negative ion chemistry of bromine in air, under conditions of the IMS reaction region, has been described in detail previously.¹⁴ As in solution chemistry, bromide ions that are formed through dissociative electron attachment processes can cluster with a bromine molecule to form Br_3^- ions. In the gas phase, charge transfer and electron capture reactions in which Br_2^- ions are formed are also observed. Thus, the equilibrium between Br^- , Br_2^- , and Br_3^- is strongly dependent on the concentration of bromine vapors in the sampled air. The conventionally depicted correlation between the peak area ratios $[\text{Br}_3^-]/([\text{Br}_3^-] + \text{Br}^-)]$ and bromine concentration was shown previously.¹⁴ Application of the NN to the analysis of bromine mobility spectra was already reported.¹⁹ Of the nine available mobility spectra in the concentration range 30–460 ppb, eight were used for the training set and one as the test point. Five reduction cycles were made, so that only nine of the 330 original input channels remained in the final 9-5-1 NN. The results are summarized in Table 1, where an error of 2.7% (5 ppb) was found in the test case and several of the training points had zero error.

Hydrogen Fluoride. Although hydrogen fluoride is a very simple molecule, its gas-phase negative ion chemistry in the reaction region of the IMS is complex.^{6,7} Spangler and Epstein⁶ showed in their IMS/MS (ion mobility spectrometry/mass spectrometry) study that several species that contained F^- or O_2^- core ions, clustered with a combination of HF, N_2 , H_2O , and CO_2 molecules that were present in the IMS drift tube, were formed. They demonstrated that addition of a reagent gas (methyl salicylate) was greatly helpful, if not vital, for obtaining quantitative information from the mobility spectra. In our own measurements we noticed that, by

Table 1. Prediction Error of Bromine Concentration from Ion Mobility Spectra by Reduced NN

sample number	Br ₂ concentration (ppb)		error	
	actual	calcd	ppb	%
1	30	32	2	7.3
2	70	70	0	0
3	82	82	0	0
4	115	115	0	0
5 ^a	190	195	5	2.7
6	360	360	0	0
7	415	415	0	0
8	435	435	0	0
9	460	458	-2	-0.5

^a Sample 5, the test point, was not used in training the NN.

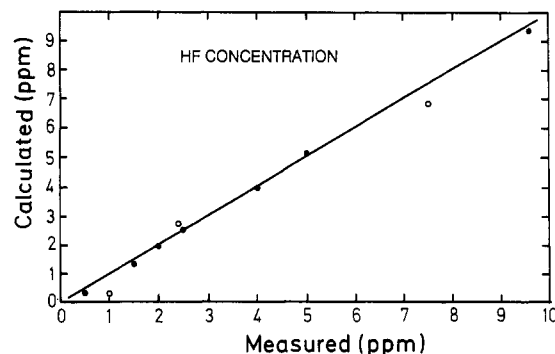


Figure 6. Results of the neural network processing of the mobility spectra of HF in air plotted as a function of the calibrated concentrations in the 0.5–10-ppm range. ● represents the training set and ○ the test points.

measuring the area of the three peaks with drift times of 5.62, 6.16, and 6.72 ms, we could correlate the HF concentration to the areas of these peaks, without the addition of a reagent gas.

Application of the NN gave good results for the mobility spectra of HF in the concentration range 0.5–10 ppm. One reduction cycle was made, and 116 of the 330 input channels remained in the final 116-5-1 NN. Seven mobility spectra were used for the training, and three test points were examined. The results are depicted in Figure 6 with relatively small absolute errors in the test cases, although the test point at an HF concentration of 1 ppm gave a large relative error.

CONCLUSION

The great advantage of using the neural network is that complete understanding (or, for that matter, any understanding) of the ion chemistry is not needed for data interpretation. In fact, the list of the significant input channels may help the researcher identify interesting features and ion–molecule reactions. Just feeding the mobility spectra from the learning set into the computer was sufficient for obtaining good results for the test set. This was examined for positive ion mobility spectra (DMF) and for mobility spectra of negative ions (HF and bromine). On the other hand, conventional methods that use areas of known and identified peaks can be applied only to systems where the gas-phase ion chemistry is understood, like the protonated monomer and dimer in DMF in one of the examples presented here.

The training set was very small compared with the theoretical requirement that it should be at least equal to the number of connections. Nevertheless, even these small training sets were sufficient for good learning. The explanation is that the theoretical requirement assumes that no relation exists between inputs, which apparently is not the case for mobility

spectra. A similar behavior was also found in industrial process data.²⁰ Thus, the utilization of neural network methods for IMS spectra analysis, and possibly other instrumental spectra, does not require as many examples as unrelated data sets do.

An additional benefit of the NN method is that neural hardware components, already available from several sources, may be incorporated into portable or even hand-held instruments. The trained NN could instantly give the correct results for air pollutant concentration, even in the parts per billion range, thus providing a reliable tool to help meet environmental monitoring requirements.

In summary, use of the neural networks for deriving quantitative information from mobility spectra was demonstrated for three chemicals in air, at concentrations in the parts per billion to parts per million ranges. The agreement between the calculated concentration of the test points and the "known" value determined from the parameters of the dynamic flow system was within a few percent in all cases. The main advantage of using NN techniques is that detailed knowledge of the ion chemistry of the sampled species was not necessary, while in conventional data processing methods, such knowledge is vital. Also, correlating measured test points with the real concentration of the chemical was straightforward and rapid, once the learning set had been processed by the NN. Thus, online application of NN is a distinct possibility for IMS instruments.

REFERENCES AND NOTES

- (1) St. Louis, R. H.; Hill, H. H., Jr. Ion Mobility Spectrometry in Analytical Chemistry. *Crit. Rev. Anal. Chem.* **1990**, *21*, 321-355.
- (2) Hill, H. H., Jr.; Siems, W. F.; St. Louis, R. H.; McMinn, D. G. Ion Mobility Spectrometry. *Anal. Chem.* **1990**, *62*, 1201A-1209A.
- (3) Eiceman, G. A. Advances in Ion Mobility Spectrometry: 1980-1990. *Crit. Rev. Anal. Chem.* **1991**, *22*, 17-36.
- (4) Roehl, J. E. Environmental and Process Applications for Ion Mobility Spectrometry. *Appl. Spectrosc. Rev.* **1991**, *26*, 1-57.
- (5) Dam, R. J. Analysis of toxic vapors by plasma chromatography. In *Plasma Chromatography*; Carr, T. W., Ed.; Plenum Press: New York, 1984; Chapter 6, pp 177-214.
- (6) Spangler, G. E.; Epstein, J. Detection of HF Using Atmospheric Pressure Ionization (API) and Ion Mobility Spectrometry (IMS). The 38th Conference on Mass Spectrometry and Allied Topics, Tucson, Arizona, 1990.
- (7) Bacon, A. T. Ion Mobility Spectrometry: A New Method of Monitoring for Hydrogen Fluoride, Ammonia, and Other Industrial Gases. 41st Pittsburgh Conference and Exposition of Analytical Chemistry and Applied Spectroscopy, New York, NY, 1990.
- (8) Brokenshire, J. L.; Dharmarajan, V.; Coyne, L. B.; Keller, J. Near Real Time Monitoring of TDI Vapor using Ion Mobility Spectrometry. *J. Cell. Plast.* **1990**, *29*, 123-142.
- (9) Karpas, Z. Ion mobility spectrometry of aliphatic and aromatic amines. *Anal. Chem.* **1989**, *61*, 684-689.
- (10) Zupan, J.; Gasteiger, J. Neural networks: A new method for solving chemical problems or just a passing phase. *Anal. Chim. Acta* **1991**, *248*, 1-30.
- (11) Bell, S. C.; Eiceman, G. A. Hand-held GC-ion mobility spectrometry for on-site analysis of complex organic mixtures in air or vapors over waste sites. 2nd International Symposium on Field Screening Methods for Hazardous Wastes and Toxic Chemicals, Las Vegas, NV, 1991.
- (12) Boger, Z.; Karpas, Z. Application of neural networks for interpretation of ion mobility and X-ray fluorescence spectra. *Anal. Chim. Acta*, in press.
- (13) Bell, S. E.; Mead, W. C.; Eiceman, G. A.; Jones, R. D.; Ewing, R. G. Connectionist hyperprism neural network for the analysis of ion mobility spectra: an empirical evaluation. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 609-615.
- (14) Karpas, Z.; Pollevoy, Y.; Melloul, S. Determination of bromine in air by ion mobility spectrometry. *Anal. Chim. Acta* **1991**, *249*, 503-507.
- (15) Bos, A.; Bos, M.; van der Linden, W. E. Artificial neural networks as a tool for soft-modelling in quantitative analytical chemistry: the prediction of the water content of cheese. *Anal. Chim. Acta* **1992**, *256*, 133-144.
- (16) Borggaard, C.; Thodberg, H. H. Optimal minimal neural interpretation of spectra. *Anal. Chem.* **1992**, *64*, 545-551.
- (17) Guterman, H. Unpublished work.
- (18) Boger, Z.; Guterman, H.; Kramer, M. A. Unpublished work.
- (19) Boger, Z. *Isr. Chem. Eng.* **1991**, *20*, 6.
- (20) Boger, Z. Application of neural networks to water and wastewater treatment plant operation. *ISA Trans.* **1992**, *31*, 25-33.
- (21) Leonard, J.; Kramer, M. A. Improvement of the backpropagation algorithm for training neural networks. *Comput. Chem. Eng.* **1990**, *14*, 337-341.
- (22) TURBO-NEURON 1.1, NEXSYS TEMED-Neural Expert Systems, Ltd., Temed Park, D. N. Arava, Israel 86800.