ELSEVIER

Journal of Fluorine Chemistry 73 (1995) 107–111

JOURNAL OF FLUORINE CHEMISTRY

# Estimation of the properties of hydrofluorocarbons by computer neural networks

Andrei A. Gakh [a,*], Elena G. Gakh [a], Bobby G. Sumpter [a], Donald W. Noid [a], Lee D. Trowbridge [b], David A. Harkins [b]

[a] *Chemistry Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6197, USA*

[b] *Martin Marietta Energy Systems, INC., Oak Ridge, TN 37831-7266, USA*

Received 12 August 1994; accepted 20 November 1994

**Abstract**

A simple computational scheme which utilizes computational neural networks was developed and used for estimating physical properties of hydrofluorocarbons. Testing of the computational method has demonstrated that thermodynamic and physical characteristics (boiling point, density, critical temperature, heat of evaporation) could be predicted with an average error of 3–5%.

*Keywords:* Hydrofluorocarbons; Computer neural networks

## 1. Introduction

The impending ban on the production of common Freons (chlorofluorocarbons, CFCs) has generated a considerable boost of research activity in fluoro-organic chemistry, especially in the field of synthetic chemistry of small fluoro-organic molecules. The main priorities of this research, however, have focused on the synthesis of hydrofluorocarbons (HFCs) with given properties to match the properties of the discontinued CFCs as closely as possible.

Here we present a computational scheme based on neural network computing which is capable of estimating properties of HFCs based on their structures. Using commercially available and inexpensive software, it allows the focus of synthetic efforts on the most promising compounds instead of wide screening of all possible candidates. We utilized a user-friendly neural-network simulator installed on an IBM personal computer, which makes these computations intelligible and reproducible even for a non-specialist in computational chemistry.

## 2. Results and discussion

Application of computational neural networks in solving chemical problems has been receiving growing attention from computational chemists for use as a powerful tool of approximation, computation and pattern recognition [1–3]. Such applications have been particularly successful in the area of prediction of physical properties of compounds based on their structures [4–6].

In our previous paper we demonstrated one of the possible schemes of neural network computing by predicting the properties of hydrocarbons [6]. The main idea was to encode structural parameters by using a set of topological invariants, and then to use these invariants as input for neural-network computing [6]. Another possible approach entails direct utilization of chemical structure features of the compound as a source of the structural information for the neural networks.

Hydrofluoroethanes and -propanes were chosen as basic compounds since they are the most likely substitutes for Freons R-11 and R-114 which are currently being sought by different organizations, as they are less likely to destroy the ozone layer. The final choices will be made based on the match of the physical parameters of the CFC alternatives with those of the parent compounds. Among the most important parameters are boiling point, heat of evaporation, density, critical temperature, atmospheric lifetime, and ozone-depletion potential. Using our technique we have tried to predict some of the above properties. Different sources of physical data including reference materials and primary

* Corresponding author.


*SSDI* 0022-1139(94)03215-7

publications as well as computer databases were used for training in this study [7–13]. In the cases of ambiguous data, preference was given to the most recent publications. A user-friendly commercial neural-network simulator [14] was used in this study, which makes these computations intelligible and reproducible even for the non-specialist in computational chemistry (in this study we have used backpropagation-type neural networks included in this program package).

Since HFCs consist of only carbon, fluorine and hydrogen, they are ideal objects for a technique which generates numeric input vectors directly from structural formulas. This method is also capable of some generalization thus allowing us to utilize one neural network for all compounds: hydrofluoroethanes, -propanes and -butanes.

The neural network consists of eight input neurons, one hidden layer and one output neuron (see Fig. 1). For simplicity, bias was omitted in the figure. The quantity of neurons in a hidden layer vary (2–6) and were determined during a standard optimization procedure using leave-k-out and other methods [14]. As an input we have used a quantity of fluorine and hydrogen atoms at each carbon unit of the alkane framework. An example of a sample-input preparation is presented in Fig. 1. Since our scheme cannot distinguish structures of diastereoisomers, their physical data were averaged. Different numeration techniques were used for propanes and ethanes since their numeric vectors must contain six and four positions, respectively. In these cases we added 00 or 00 00 vectors to make them compatible with the numeric vectors for butanes. These zero values indicated to the neural network that the corresponding methylene fragments were missing. A full data set including formulas, experimental and predicted properties of hydrofluoroethanes and hydrofluoropropanes is presented in Table 1.

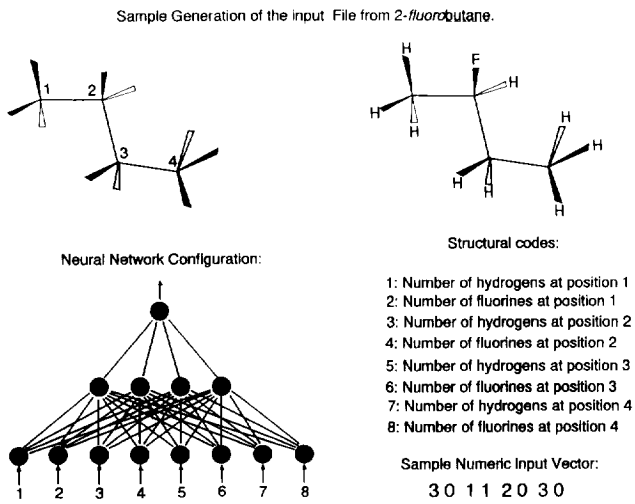


Fig. 1. Sample generation of the input file, structural codes and neural-network configuration.

Table 1
Physical properties of hydrofluorocarbons

| Formula | $B.p.$ (°C) | $T_c$ (°C) | $D_4^{25}$ (g cm$^{-3}$) | $\Delta H_{evap}^{bp}$ (kJ mole$^{-1}$) |
|---|---|---|---|---|
| $CH_3CH_2F$ | −37.4 | 102.2 | 0.709 [b] | 18.40 [c] |
| $CH_3CHF_2$ | −24.7 | 113.4 | 0.901 | 18.92 [c] |
| $CH_2FCH_2F$ | 10.5 | 196.0 | 0.979 [b] | 20.37 [a] |
| $CH_3CF_3$ | −47.6 | 73.2 | 0.962 | 14.19 [c] |
| $CH_2FCF_2H$ | 5.0 | 149.1 | 1.095 [a] | 20.81 [a] |
| $CH_2FCF_3$ | −26.1 | 100.0 | 1.216 [b] | 20.12 [a] |
| $CHF_2CHF_2$ | −23.0 | 99.1 | 0.972 [b] | 20.12 [a] |
| $CHF_2CF_3$ | −48.6 | 68.8 | 1.200 [a] | 18.69 [a] |
| $CF_3CF_3$ | −78.2 | 19.8 | – | 16.16 |
| $CFH_2CH_2CH_3$ | −2.5 | 145.4 | 0.719 [b] | 24.20 |
| $CH_3CHFCH_3$ | −9.4 | 139.1 [a] | 0.860 [b] | 23.60 |
| $CF_2HCH_2CH_3$ | 8.0 | 156.9 | 0.892 [a] | 25.27 [a] |
| $CFH_2CHFCH_3$ | 15.0 | 167.2 [a] | 0.912 [a] | 24.87 [a] |
| $CFH_2CH_2CFH_2$ | 41.3 | 191.0 [a] | 1.006 | 25.27 [a] |
| $CH_3CF_2CH_3$ | −0.4 | 149.0 [a] | 0.920 | 24.44 [a] |
| $CF_3CH_2CH_3$ | −13.0 | 118.9 | 0.959 [a] | 26.59 [a] |
| $CF_2HCHFCH_3$ | 18.7 | 168.2 [a] | 1.054 [a] | 26.30 [a] |
| $CF_2HCH_2CFH_2$ | 45.0 | 191.7 [a] | 1.131 [a] | 26.59 [a] |
| $CFH_2CHFCH_2F$ | 70.0 | 206.5 [a] | 1.202 [a] | 26.30 [a] |
| $CH_3CF_2CH_2F$ | 21.7 [a] | 171.8 [a] | 1.000 [a] | 25.98 [a] |
| $CF_3CFHCH_3$ | −1.0 | 138.8 [a] | 1.130 [a] | 26.81 [a] |
| $CF_3CH_2CFH_2$ | 29.4 | 175.5 [a] | 1.213 [a] | 27.05 [a] |
| $CF_2HCHFCFH_2$ | 4.5 | 146.6 [a] | 1.146 [a] | 26.82 [a] |
| $CHF_2CH_2CHF_2$ | 27.0 | 173.1 [a] | 1.207 [a] | 27.05 [a] |
| $CHF_2CF_2CH_3$ | −0.8 | 140.5 [a] | 1.131 [a] | 24.86 |
| $CH_2FCF_2CFH_2$ | 45.3 [a] | 191.0 [a] | 1.134 [a] | 26.50 [a] |
| $CF_3CF_2CH_3$ | −18.3 | 108.5 | 1.176 | 23.60 |
| $CF_3CHFCH_2F$ | 20.0 | 163.0 [a] | 1.300 [a] | 26.53 [a] |
| $CF_3CH_2CHF_2$ | 15.3 | 157.5 | 1.324 | 28.00 |
| $CF_2HCHFCHF_2$ | 18.6 [a] | 161.2 [a] | 1.254 [a] | 26.53 [a] |
| $CF_2HCF_2CH_2F$ | 26.0 | 178.4 | 1.336 | 29.20 |
| $CF_3CF_2CFH_2$ | −1.4 | 130.1 | 1.290 [b] | 25.60 |
| $CF_3CFHCF_2H$ | 6.5 | 141.1 | 1.390 | 26.83 |
| $CF_2HCF_2CF_2H$ | 12.6 | 155.2 | 1.480 [b] | 26.60 |
| $CF_3CH_2CF_3$ | −1.1 | 130.6 | 1.371 | 25.60 |
| $CF_3CF_2CF_2H$ | −16.3 | 106.3 | 1.375 [b] | 23.69 |
| $CF_3CHFCF_3$ | −18.3 | 102.8 | 1.409 [b] | 22.30 |
| $CF_3CF_2CF_3$ | −39.0 | 71.0 | 1.337 [b] | 20.50 |

[a] Neural Network Estimations.
[b] Extrapolations from different temperatures.
[c] At 25 °C.

## 2.1. Boiling point

There are several reports on predicting boiling points of CFCs and HFCs. The most recent one is based on data for 256 halogenated alkanes, mostly propanes, ethanes and methanes, and includes equations found by regression analysis which correlate boiling point with a set of topological indices and stoichiometric data [15]. However, the best accuracy was achieved by a modified boiling-point numbers scheme [16], which was initially developed to establish relationships between stoichiometry and boiling point. The modification included

some additional rules which were established in the series of halogenated ethanes on the basis of structural fragment analysis [16].

Our neural network performance data are presented in Table 1 and in Fig. 2. These results were achieved by a neural network having six neurons in a hidden layer (hyperbolic tangent) after 128 032 epochs requiring 27 min of real-time computations on an IBM PC XP-486 (50 MHz) [14]. The average error of the test set is 10.8 °C, which is noticeably worse than what has been reached by regressional analysis (5–6 °C) [15,16]. The main reason for this is utilization of the generalized input file (ethane, propane, butane derivatives all together), and the well-known dependence of the boiling points of HFCs on dipolar interactions [16]. The last parameter is the most important one since we have already shown that a similar approach for predicting boiling points of hydrocarbons gives an average error as low as 2–3 °C [6]. It appears that the neural network experienced difficulties in recognition of these dipole–dipole interactions (not directly represented in our scheme). The last problem can be easily recognized by observation of the S-shaped curve on the plot of the predicted values versus observed values, where the lower boiling points (weak or no dipole interactions) are estimated higher, and the higher boiling points (strong dipole interactions) are predicted lower. The overall performance was sufficiently better (average error 4.7 °C) when we excluded the extremes and used a limited temperature range of −40 to +40 °C.

### 2.2. Critical temperature

Our initial attempts of critical-temperature estimation for HFCs showed that although our scheme developed for boiling points was also applicable, the average error could not be reduced below 10 °C. The main reasons for this are that an insufficient volume of experimental data was available for us and the complex structure–property relationships (see discussion on boiling point above). Other developed methods, for example, employing the correlation between topological indices and critical parameters, have also shown significantly less accuracy for critical temperature. Much better results (average accuracy 4.5% or 5.9 °C) were achieved using experimental boiling points of HFCs as an additional input. In this case the neural network utilizes known relationships between boiling point and critical temperature for the error reduction. Although boiling points are not known for all HFCs, this method works best for approximating the critical temperature. An optimized neural network (leave-k-out and other methods) contains one hidden layer (four neurons, sigmoid function). The performance results are shown in Fig. 3 and Table 1. They were achieved after 12 min of computation on an IBM PC XP-486 (50 MHz) over 52 170 epochs [14].

### 2.3. Heat of vaporization

The heat of vaporization (at the normal boiling point or 25 °C and equilibrium pressure) represents an important parameter for HFC application, especially in refrigeration and air-conditioning systems. Since this parameter is much less sensitive to the specific features of the molecule, the results are usually good, even in the case of HFCs. We were able to achieve accuracy as good as 1.1 kJ mole$^{-1}$ average error which is good enough for most applications. The optimized architecture contained one hidden layer (four neurons, sigmoid function), and the training cycle was completed after 44 467 epochs. The results are presented in Fig. 4 and Table 1.

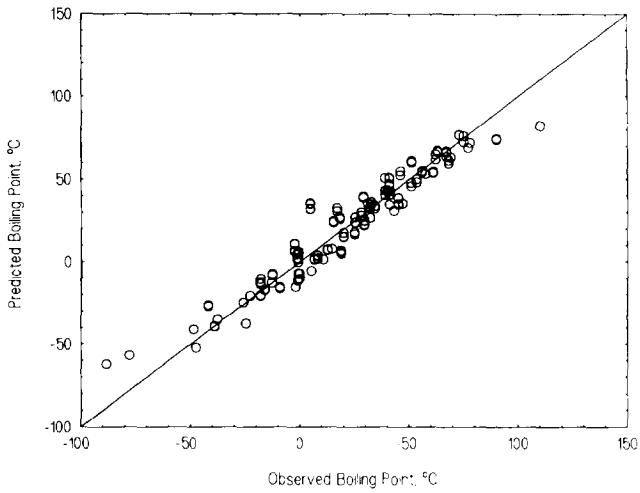


Fig. 2. Experimental versus neural-network predicted values of HFC boiling points for the testing (cross-validation) set.

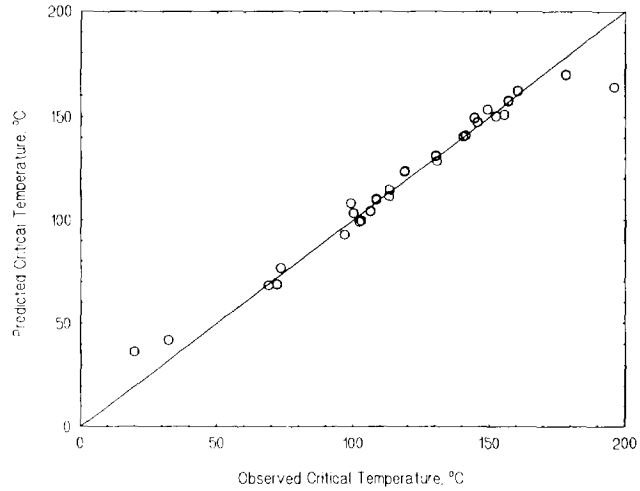


Fig. 3. Experimental versus neural-network predicted values of HFC critical temperatures for the testing (cross-validation) set.

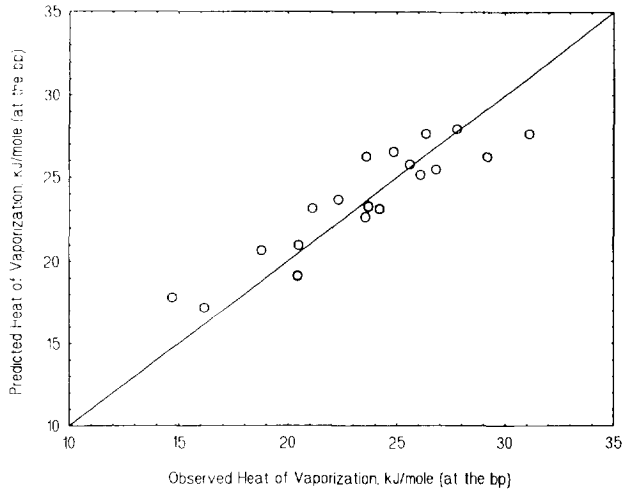


Fig. 4. Experimental versus neural-network predicted values of HFC heats of vaporization for the testing (cross-validation) set.

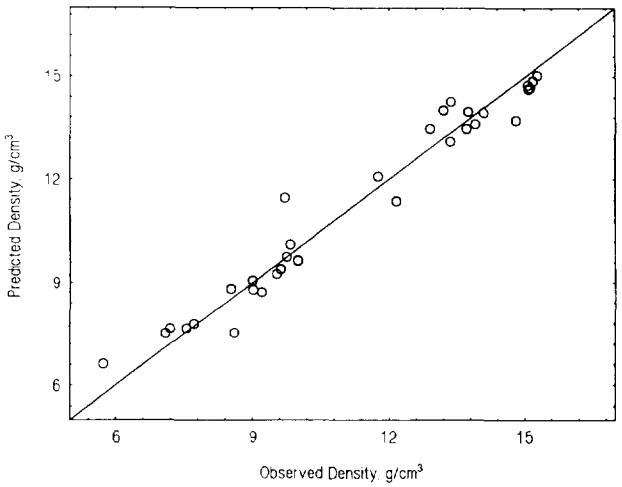


Fig. 5. Experimental versus neural-network predicted values of HFC liquid densities for the testing (cross-validation) set.

### 2.4. Liquid density

Density is another important parameter for HFC applications, and it appears to be the best parameter for the predictions based on neural-network computations. As we have already shown, the density of alkanes could be predicted from their structures with an average accuracy better than 1% [6]. The density of HFCs can be predicted with lower, but still very good, accuracy (2.7% or 0.03 g $cm^{-3}$), and is satisfactory for most applications. These results were achieved with one hidden layer containing two neurons (sigmoid function) after 30 021 epochs. The graphic plot of the results is presented in Fig. 5 and Table 1.

## 3. Conclusions

We have shown that neural-network computing can be effectively used for estimation of unknown physical and thermodynamic properties of HFCs based on their chemical structures. This simple computational scheme has an average accuracy of 3–5% and does not require any prior knowledge of property–structure relationships for this class of compounds. In addition, all necessary computations can be performed on a personal computer within 1 h.

## Acknowledgements

Research sponsored by the Division of Materials Sciences, Office of Basic Energy Sciences, US Department of Energy, under contract DE-AC05-84OR21400, with Martin Marietta Energy Systems, Inc, and in part by an appointment to the ORNL Postdoctoral Research Associates Program administered jointly by the ORNL and the Oak Ridge Institute for Science and Education.

## References

[1] G.M. Maggiora, D.W. Elrod and R. G. Trenary, Computational neural networks as model-free mapping devices, *J. Chem. Inf. Comput. Sci., 32* (1992) 732–741.
[2] J. Gasteiger and J. Zupan, Neural networks in chemistry, *Angew. Chem. Int. Ed. Engl., 32* (1993) 503–527.
[3] J. Zupan and J. Gasteiger, Neural networks: a new method for solving chemical problems or just a passing phase?, *Anal. Chim. Acta, 248* (1991) 1–30.
[4] N. Bodor, A. Harget and M-J. Huang, Neural network studies. 1. Estimation of the aqueous solubility of organic compounds, *J. Am. Chem. Soc., 113* (1991) 9480–9483.
[5] L.M. Egolf and P.C. Jurs, Prediction of boiling points of organic heterocyclic compounds using regression and neural network techniques, *J. Chem. Inf. Comput. Sci., 33* (1993) 616–625, and references cited therein.
[6] A.A. Gakh, E.G. Gakh, B.G. Sumpter and D.W. Noid, Neural network–graph theory approach to the prediction of the physical properties of organic compounds, *J. Chem. Inf. Comput. Sci., 34* (1994) 832–839.
[7] Physical and thermodynamical properties of non-hydrocarbons, *American Petroleum Institute Research, Project 44*, National Bureau of Standards, 1947–1991.
[8] Beilstein Database File, STN International.
[9] A.L. Beyerlein, D.D. DesMarteau, S.H. Hwang and N.D. Smith, Physical properties of fluorinated propane and butane derivatives as alternative refrigerants, *ASHRAE Trans., 99*(1) (1993) 368–379.
[10] W.L. Kopko, Beyond CFCs: extending the search for new refrigerants, *Rev. Int. Froid, 13* (1990) 79–85.
[11] N.V. Vityuk, Calculation and prediction of critical parameters of chloroalkanes on the basis of a topological model, *Zh. Fizicheskoi Khim., 66* (1992) 2665–2670.
[12] J. Burdon, S.T. Ezmirly and T.N. Huckerby, The fluorination of butane over cobalt trifluoride, *J. Fluorine Chem., 40* (1988) 283–318.
[13] T. Hudlicky, R. Fan, J.W. Reed, D.R. Carver, M. Hudlicky and E.I. Eger II, Practical preparation of some potentially anesthetic fluoroalkanes: regiocontrolled introduction of hydrogen atoms, *J. Fluorine Chem., 59* (1992) 9–14.

[14] *NeuralWorks Professional II/PLUS and NeuralWorks Explorer,* NeuralWare, Inc., Penn Center West, Building IV, Pittsburgh, PA 15276-9910 (for PC and mainframe computers). For more details of neural-network computing see, for example, Refs. [1–3,6], and references cited therein.

[15] A.T. Balaban, N. Joshi, L.B. Kier and L.H. Hall, Correlations between chemical structure and normal boiling points of halogenated alkanes C1–C4, *J. Chem Inf. Comput. Sci., 32* (1992) 233–237.

[16] A.A. Woolf, Boiling point relations in the halogenated ethane series, *J. Fluorine Chem., 50* (1990) 89–99.