

## Composite group vector space method for estimating melting and boiling point of pure organic compound

Wei Wenying<sup>\*,\*\*,\*†</sup>, Wang Zhen<sup>\*\*</sup>, Yin Yanhua<sup>\*\*</sup>, Han Jinyu<sup>\*</sup> and Xu Wen<sup>\*</sup>

<sup>\*</sup>Key Laboratory for Green Chemical Technology of State Education Ministry,  
School of Chemical Engineering & Technology, Tianjin University, Tianjin 300072, P.R.China

<sup>\*\*</sup>The 718 Research Institute of CSIC, HanDan 056027, P.R.China

(Received 1 December 2006 • accepted 5 August 2007)

**Abstract**—Composition Group Vector Space (CGVS) method for estimating melting and boiling point  $T_m$ ,  $T_b$  of organic compound has been proposed, and the principle of this method has been elucidated. The models for estimating  $T_m$ ,  $T_b$  have been established and the numerical values of relative parameters have been presented. The average percentage deviations of  $T_m$ ,  $T_b$  estimation are 7.53 and 1.58, respectively, which show that the present method demonstrates significant improvement in applicability to predict the above properties, compared to conventional group methods.

Key words: Composition Group Vector Space, Pure, Organic Compound, Melting Boiling, Boiling Point, Estimation

### INTRODUCTION

Melting and boiling points are very important data in the chemical and petroleum industries, and they are applied widely in domains such as state calculation, process simulation, and product design. It is not always possible, however, to find reliable experimental values of melting and boiling points for the compounds of interest in the literature, nor is it practical to measure the properties as the need arises. So the estimation of melting and boiling points is required.

For the estimation of the properties of pure organic compounds, simple group contribution (SGC) methods [1-4] are widely used. These methods provide the important advantage of quick estimates without requiring substantial computational resources. However, many of these methods are of questionable accuracy and utility. To overcome this limitation, composite group contribution methods (Marrero-Morejún and Pardillo-Fontdevila [5] and Constantinou and Gani [6]) have been reported in the literature. However, new problems have emerged along with the increment of regression accuracy of these methods.

Model parameters of the group method are obtained from fitting property data of a great many substances, which behave like a statistical feature. Only if the number of substances in linear data regression is much more than that of parameters in the model does the group method have a function of extrapolation predicting. Compared with the number of substances in regression, the higher the parameter number in the model, the poorer the predicting function of the model is. If the number of model parameters is more than that of the substances, the value of the model parameter solved will not be unique. For simple group methods, only a single functional group should be taken as the independent molecular structural unit. There are not more than 40 groups for various organic compounds in these methods. An addition of two or more adjacent simple groups was taken as the independent molecular structure unit for composite group methods. When only interaction between two adjacent sim-

ple groups is taken into account, the number of independent molecular structural units for all organic compounds is 223, and the parameter number in property correlation will be the same. The method proposed by Marrero-Morejún and Pardillo-Fontdevila was the group interaction contribution (GIC) method. Among for their estimation of critical properties, the number of critical volume data in regression was only 289, which was not much larger than 223. For the Constantinou and Gani method, the estimation was performed at two levels: the basic level only used the contribution from the first-level groups, while the second level increased the consideration of the second-order groups. Their 78 first-level groups were not sufficient to describe the molecules of some common compounds, and under certain circumstances, the same molecule may be described in different ways because of over complication of their method.

In this work, a composite group vector space method for estimating the melting and boiling points of organic compounds is proposed. Another purpose of this study is to develop a new estimation method limit the number of model parameters with higher accuracy.

### GROUP VECTOR SPACE FOR ORGANIC COMPOUNDS

In this work, we select 41 simple groups (see Table 1) to describe organic compounds. The molecule is considered to be in a given space, and every group in the molecule is only a point in the space. The molecules of organic compounds can be divided into three kinds: chain molecule, monocyclic molecule, and polycyclic molecule. For sake of convenience, graphs with different number of points are all expressed as graphs with five points. Thus, the organic molecules can be expressed as three topologic graphs in Fig. 1.

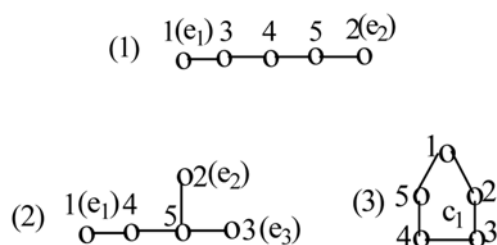
The topologic structure of a molecule can be described by the distance matrix of the molecule. The distance matrix is a square symmetric matrix. It is required to input  $n^2$  elements for the  $n$ -order square matrix. The input quantity for a large molecule is too large, and, furthermore, the distance matrix is unable to describe the molecular structure characteristics such as chains, rings, and branches sufficiently. To overcome the above limitation, the following group

<sup>†</sup>To whom correspondence should be addressed.

E-mail: wying\_wei@peric.ac.cn

**Table 1.** The values of group parameters of Eqs. (6)-(8)

No.	Group	$\Delta T_{mi}$	$\Delta T_{bi} \times 10^{-2}$	No.	Group	$\Delta T_{mi}$	$\Delta T_{bi} \times 10^{-2}$
1	-CH <sub>3</sub>	-0.511	-3.262	22	O=CH-	0.295	8.843
2	>CH <sub>2</sub>	-0.0790	2.967	23	-COOH	0.856	16.680
3	>CH-	0.158	2.259	24	HCOO-	0.0259	4.699
4	>C<	0.122	2.334	25	-COO-	0.0273	4.752
5	=CH <sub>2</sub>	-0.720	-1.306	26	=C=O	0.167	-1.672
6	=CH-	-0.077	3.367	27	-SH	-0.348	8.930
7	=C<	0.347	7.165	28	-S-	4.209	15.631
8	=C=	-0.549	-2.006	29	(-S-) <sub>R</sub>	1.566	21.565
9	≡CH	-0.411	-2.756	30	-NH <sub>2</sub>	1.368	10.779
10	≡CH-	-0.082	-1.973	31	>NH	0.900	11.352
11	(>CH <sub>2</sub> ) <sub>R</sub>	-0.412	1.263	32	(>NH) <sub>R</sub>	0.334	17.065
12	(>CH-) <sub>R</sub>	-0.178	7.218	33	>N-	3.046	38.260
13	(>C<) <sub>R</sub>	-1.163	-8.720	34	-N=	2.685	17.345
14	(=CH-) <sub>R</sub>	-0.082	2.243	35	(-N=) <sub>R</sub>	2.512	-2.665
15	(=C<) <sub>R</sub>	0.400	11.430	36	-CN	0.0282	5.909
16	-OH	0.144	11.807	37	-NO <sub>2</sub>	1.074	12.433
17	(-OH) <sub>AC</sub>	0.727	6.866	38	-F	0.111	-0.531
18	-O-	1.307	7.991	39	-Cl	0.184	2.804
19	(-O-) <sub>R</sub>	0.119	-13.124	40	-Br	0.441	3.714
20	>C=O	0.335	14.261	41	-I	0.501	9.351
21	(>C=O) <sub>R</sub>	1.625	28.171				

**Fig. 1.** Three topologic graphs expressing organic compounds.

vector space method has been proposed.

Considering the chain graph first, the dimension number of the space is equal to the number of end points ( $e_i$ ) on the chain, and one end point has determined a dimension of the space. The coordinate of an end point in the dimension determined by it is zero, while the coordinate of other points in this dimension is the distance from that point to the end point. For cyclic graph, one ring ( $c_i$ ) represents a dimension. In that dimension, the coordinate of the ring point equals the number of points on the ring, and the coordinate of the nonring point equals the sum of the distance from the point to ring and the number of points on the ring. If the route from the ring point to end point is non-unique, the shortest route should be selected. So, the dimension number  $m$  of the space for a graph is equals to the sum of the number  $k_e$  of end points ( $e_i$ ) and the number  $k_c$  of rings ( $c_i$ ) in the graph. Every point in the graph has  $m$  coordinates in the  $m$ -dimension space. The graph may be described by a space matrix, the number of rows in the matrix equals the number of points in the graph, and the number of columns equals the dimension number of the space. The space matrixes of the above three topologic graphs are as follows:

$$\begin{array}{l}
 \begin{array}{c} e_1 \\ 1 \end{array} \begin{bmatrix} 0 & 4 \\ 2 & 4 \\ 3 & 1 \\ 4 & 3 \\ 5 & 2 \end{bmatrix} ; \quad \begin{array}{c} e_1, e_2, e_3 \\ 1 \end{array} \begin{bmatrix} 0 & 3 & 3 \\ 2 & 3 & 0 \\ 3 & 3 & 2 \\ 4 & 1 & 2 \\ 5 & 2 & 1 \end{bmatrix} ; \quad \begin{array}{c} c_1 \\ 1 \end{array} \begin{bmatrix} 5 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}
 \end{array}$$

The numerals in the left of the matrix represent the numbers of points in the space, the symbols above the matrix are the dimensions of the space, and the element  $b_{ij}$  of the matrix is the distance from  $i$ -group to  $j$ -dimension. That is, the values of the row vector in the matrix are the distance from a group to all dimensions, and the values of the column vector are the distance from all groups to a dimension. The matrixes show that the space position of point  $i$  in the graph can be represented by the  $m$ -dimension vector  $(b_{i1}, b_{i2}, \dots, b_{im})$ . Then the module  $\alpha_i$  of point  $i$  vector is defined as:

$$\alpha_i = \left( \sum_{j=1}^m b_{ij}^2 \right)^{1/2} \quad (i=1, 2, 3, 4, 5)$$

The average square root of the module of some point  $i$  in the graph is defined as the module index  $\nu_i$  of this point vector. That is:

$$\nu_i = \alpha_i / \left( \sum_{j=1}^5 \alpha_j^2 \right)^{1/2} \quad (i=1, 2, 3, 4, 5)$$

The quantity  $\nu_i$  is used to describe the point  $i$  position in the space. On the analogy of this, the module index  $\nu_i$  of group  $i$  in the molecule is taken to characterize the position of that group in the molecular space. Thus, every simple group, except halogen groups, has its own independent module index. For the four halogen groups, their module indexes were determined to be the same as those of the hydrocarbon groups with which they were connected.

## GROUP-ADJACENT ATOM PAIR

In each group that expresses organic compounds, there are one or more atoms connected with other groups through their free bonds and being adjacent atoms of other groups. Because some groups of different type contain the same atom connected with the same adjacent group, if a group and adjacent atom pair are taken as the independent molecular structural unit, the sum of the structural units of molecules will be much less than that of adjacent group and group pairs. Taking group  $-\text{CH}_3$  as an example, the number of groups adjacent to it is 28 in the GLC method, while the number of atoms adjacent to it is only 11. So, the number of structural units is reduced.

## CORRELATION AND GROUP PARAMETERS

The expression of critical property  $f$  for simple group method is:

$$f = a + \sum_i n_i \Delta f_i \quad (1)$$

Where subscript  $i$  represent group type,  $\Delta f_i$  is the contribution value of  $i$ -type group,  $n_i$  is the number of  $i$ -type group in the molecule, and  $a$  is the correlation constant.

In this study, the group contribution was divided into three parts: position contribution of the group, independent contribution of the group and contribution of the group-adjacent atom pair. The property  $f$  is expressed as follows:

$$f = a + \sum_i \left( \sum_{k=1}^{n_i} \nu_{ik} \right) \Delta f_i + \sum_i n_i \Delta f_i^o + \sum_i \left( \sum_j n_{ij} \Delta f_{ij}^o \right) \quad (2)$$

Where, subscripts  $i$  and  $ij$  represent a group and a group  $i$ -adjacent atom  $j$  pair, respectively.  $\left( \sum_{k=1}^{n_i} \nu_{ik} \right)$  is the module index sum of  $i$ -type groups.  $\Delta f_i$  is the position contribution of  $i$ -type group,  $\Delta f_i^o$  is the independent contribution of the group  $i$  and  $\Delta f_{ij}^o$  is the contribution of group  $i$ -adjacent atom  $j$  pair. If group  $i$  has  $l_i$  free bonds, then

$$n_i = \sum_j n_{ij} / l_i \quad (3)$$

$$\text{Let } \Delta f_j = \Delta f_i^o / l_i + \Delta f_{ij}^o \quad (4)$$

Eq. (2) can be rewritten as in the following form:

$$f = a + \sum_i \left( \sum_{k=1}^{n_i} \nu_{ik} \Delta f_i \right) + \sum_j n_j \Delta f_j \quad (5)$$

To improve the property estimation accuracy, the power index "1" of property  $f$  in Eq. (5) is substituted by a variable, the optimum value of which can be obtained by trial computation. On doing this, the expressions of  $T_c$ ,  $P_c$ ,  $V_c$  may be written as follows:

$$T_m^{0.3} = 4.643 + \sum_i \left( \sum_{k=1}^{n_i} \nu_{ik} \right) \Delta T_{mi} + \sum_j n_j \Delta T_{mj} \quad (6)$$

$$T_b^{1.3} = 1221.52 + \sum_i \left( \sum_{k=1}^{n_i} \nu_{ik} \right) \Delta T_{bi} + \sum_j n_j \Delta T_{bj} \quad (7)$$

Where  $T_m$  and  $T_b$  are melting point and boiling point, respectively, and the units employed are Kelvin.

A great deal of experimental data in literature [7-10] has been used to optimize the values of group parameters in Eqs. (6)-(8). The simple method of least square is used for optimization with the fol-

**Table 2. The values of group-adjacent atom pair parameters of Eqs. (3)-(5)**

No.	Group-adjacent atom pair	$\Delta T_{mj}$	$\Delta T_{bj} \times 10^{-2}$	No.	Group-adjacent atom pair	$\Delta T_{ej} \times 10^3$	$\Delta T_{ej} \times 10$
1	$-\text{CH}_3 / >\text{C}<$	0.132	1.234	74	$(>\text{CH}-)_R / =\text{C}<$	-0.183	2.068
2	$-\text{CH}_3 / =\text{C}<$	0.167	0.419	75	$(>\text{CH}-)_R / \equiv\text{C}-$	-	2.909
3	$-\text{CH}_3 / \equiv\text{C}-$	0.839	-29.29	76	$(>\text{CH}-)_R / (>\text{C}<)_R$	-0.021	-0.289
4	$-\text{CH}_3 / (>\text{C}<)_R$	0.360	0.0036	77	$(>\text{CH}-)_R / (=C<)_R$	0.165	-0.258
5	$-\text{CH}_3 / (=C<)_R$	0.0270	0.0370	78	$(>\text{CH}-)_R / -\text{O}-$	0.578	-0.455
6	$-\text{CH}_3 / -\text{O}-$	0.0638	0.920	79	$(>\text{CH}-)_R / (-\text{O}-)_R$	-0.460	3.241
7	$-\text{CH}_3 / -\text{S}-$	-0.451	0.871	80	$(>\text{CH}-)_R / -\text{S}-$	-0.807	-0.323
8	$-\text{CH}_3 / >\text{N}-$	0.205	0.539	81	$(>\text{C}<)_R / >\text{C}<$	0.257	3.444
9	$-\text{CH}_3 / -\text{NO}_2$	0.165	3.444	82	$(>\text{C}<)_R / (>\text{C}<)_R$	-0.194	0.351
10	$>\text{CH}_2 / >\text{C}<$	0.0358	0.504	83	$(>\text{C}<)_R / -\text{F}$	0.393	1.304
11	$>\text{CH}_2 / =\text{C}<$	0.0294	-0.316	84	$(=\text{CH}-)_R / (>\text{C}<)_R$	-0.368	-0.433
12	$>\text{CH}_2 / \equiv\text{C}-$	0.384	-30.72	85	$(=\text{CH}-)_R / (=C<)_R$	0.0352	0.356
13	$>\text{CH}_2 / (>\text{C}<)_R$	-0.058	-0.867	86	$(=\text{CH}-)_R / (-\text{O}-)_R$	0.136	-0.446
14	$>\text{CH}_2 / (=C<)_R$	-0.170	-0.785	87	$(=\text{CH}-)_R / (-\text{S}-)_R$	-0.237	-5.831
15	$>\text{CH}_2 / -\text{O}-$	-0.003	-0.098	88	$(=\text{CH}-)_R / (>\text{NH})_R$	0.191	-0.813
16	$>\text{CH}_2 / -\text{S}-$	-0.514	0.536	89	$(=\text{CH}-)_R / (=N-)_R$	0.0239	1.745
17	$>\text{CH}_2 / >\text{N}-$	0.102	0.0392	90	$(=C<)_R / >\text{C}<$	0.142	0.830
18	$>\text{CH}_2 / -\text{F}$	-0.061	0.0200	91	$(=C<)_R / =\text{C}<$	0.0215	0.346
19	$>\text{CH}_2 / -\text{Cl}$	0.0657	1.293	92	$(=C<)_R / \equiv\text{C}-$	-0.220	28.573
20	$>\text{CH}_2 / -\text{Br}$	0.216	2.635	93	$(=C<)_R / (>\text{C}<)_R$	-0.332	-1.274
21	$>\text{CH}_2 / -\text{I}$	0.286	2.047	94	$(=C<)_R / (=C<)_R$	0.0388	0.366
22	$>\text{CH} / >\text{C}<$	-0.010	0.922	95	$(=C<)_R / -\text{O}-$	-0.274	-1.766
23	$>\text{CH} / =\text{C}<$	-0.006	0.319	96	$(=C<)_R / (-\text{O}-)_R$	0.0029	-0.013

Table 2. Continued

No.	Group-adjacent atom pair	$\Delta T_{mj}$	$\Delta T_{bj} \times 10^{-2}$	No.	Group-adjacent atom pair	$\Delta T_{ej} \times 10^3$	$\Delta T_{ej} \times 10$
24	>CH-/≡C-	0.353	-30.85	97	(=C<) <sub>R</sub> />N-	-0.031	-2.543
25	>CH-/(>C<) <sub>R</sub>	0.0859	-0.509	98	(=C<) <sub>R</sub> /(>N-) <sub>R</sub>	-	0.610
26	>CH-/(>C<) <sub>R</sub>	-0.185	-0.680	99	(=C<) <sub>R</sub> /-F	-0.126	-2.400
27	>CH-/O-	0.213	0.328	100	(=C<) <sub>R</sub> /-Cl	-0.033	-0.525
28	>CH-/>N-	0.163	14.595	101	(=C<) <sub>R</sub> /-Br	-0.046	0.391
29	>CH-/F	-0.161	-0.317	102	(=C<) <sub>R</sub> /-I	-0.028	1.007
30	>CH-/Cl	0.0091	1.397	103	O=CH-/>C<	-0.073	-0.027
31	>CH-/S-	0.148	1.218	104	O=CH-/=C<	0.0832	-0.052
32	>CH-/I	0.0773	3.138	105	O=CH-/(>C<) <sub>R</sub>	0.123	-0.570
33	>CH-/Br	-0.045	3.016	106	>C=O/>C<	0.0547	-0.251
34	>C</>C<	0.111	1.116	107	>C=O/(>C<) <sub>R</sub>	0.176	1.209
35	>C</=C<	-0.155	-0.583	108	>C=O/>CO	0.117	0.602
36	>C</≡C-	0.314	-31.32	109	>C=O/>N-	-0.313	0.602
37	>C</(>C<) <sub>R</sub>	-0.001	-0.380	110	>C=O/-S-	-	0.352
38	>C</(>C<) <sub>R</sub>	-0.054	-0.222	111	-COOH/>C<	0.293	2.602
39	>C</O-	0.438	-0.227	112	-COOH/=C<	0.509	2.041
40	>C</F	-0.092	-0.401	113	-COOH/(>C<) <sub>R</sub>	0.464	2.476
41	>C</Cl	0.0848	2.806	114	HCOO-/>C<	0.276	2.591
42	>C</Br	-0.008	0.909	115	-COO-/>C<	0.177	2.976
43	>C</S-	-0.004	0.906	116	-COO-/=C<	0.432	3.077
44	>C</=C<	0.198	-0.211	117	-COO-/(>C<) <sub>R</sub>	0.115	0.158
45	>C</=C=	0.124	-0.970	118	-COO-/(>C<) <sub>R</sub>	-0.676	-1.145
46	>C</(>C<) <sub>R</sub>	-0.024	-1.171	119	(>C=O) <sub>R</sub> /(>C<) <sub>R</sub>	-0.478	-3.970
47	=CH-/>C<	-0.070	1.239	120	-CN/>C<	-0.102	35.783
48	=CH-/=C<	0.0674	0.400	121	-CN/=C<	-0.0091	-3.434
49	=CH-/≡C-	0.395	7.128	122	-CN/≡C-	-	-3.667
50	=CH-/=C=	-0.106	-0.308	123	-CN/(>C<) <sub>R</sub>	0.511	-0.0709
51	=CH-/(>C<) <sub>R</sub>	-	-1.032	124	-CN/(>C<) <sub>R</sub>	-	0.0002
52	=CH-/(>C<) <sub>R</sub>	-0.007	0.470	125	-OH/>C<	0.213	0.192
53	=CH-/O-	-0.103	0.258	126	-OH/(>C<) <sub>R</sub>	0.145	-0.467
54	=CH-/F	-0.256	-3.419	127	(-OH) <sub>AC</sub> /(>C<) <sub>R</sub>	0.339	1.993
55	=CH-/Cl	0.0181	-0.431	128	-O-/>C<	-0.133	0.197
56	=CH-/>C<	-0.087	1.128	129	-O-/=C<	-0.0497	-0.361
57	=CH-/=C<	0.133	1.399	130	-O-/(>C<) <sub>R</sub>	0.144	1.779
58	>CH-/=C=	-0.030	-	131	(-O-) <sub>R</sub> /(>C<) <sub>R</sub>	0.113	2.826
59	>CH-/(>C<) <sub>R</sub>	0.0862	0.0206	132	-SH/>C<	0.918	-0.656
60	>CH-/F	-0.298	-2.643	133	-S-/>C<	-0.152	0.0560
61	>CH-/Cl	-0.145	-0.584	134	(-S-) <sub>R</sub> /(>C<) <sub>R</sub>	0.00015	2.615
62	≡CH/≡C-	0.335	0.771	135	(-S-) <sub>R</sub> /(>C<) <sub>R</sub>	0.101	2.247
63	≡C-/>C<	-0.428	32.396	136	-NH <sub>2</sub> />C<	-0.514	-1.361
64	≡C-/=C<	0.0427	-5.446	137	-NH <sub>2</sub> /=C<	0.191	-0.203
65	≡C-/≡C-	0.123	1.905	138	-NH <sub>2</sub> /(>C<) <sub>R</sub>	-0.236	2.239
66	≡C-/(>C<) <sub>R</sub>	0.0311	-29.17	139	>NH/>C<	-0.0535	0.353
67	=C=/=C<	0.294	2.402	140	>NH/(>C<) <sub>R</sub>	-0.197	0.408
68	(>CH <sub>2</sub> ) <sub>R</sub> /(>C<) <sub>R</sub>	0.110	0.6477	141	>N-/>C<	-0.307	-1.741
69	(>CH <sub>2</sub> ) <sub>R</sub> /(>C<) <sub>R</sub>	0.381	0.948	142	>N-/(>C<) <sub>R</sub>	-0.0645	-1.540
70	(>CH <sub>2</sub> ) <sub>R</sub> /(-O-) <sub>R</sub>	0.0906	4.338	143	(>NH) <sub>R</sub> /(>C<) <sub>R</sub>	-0.0002	-0.0033
71	(>CH <sub>2</sub> ) <sub>R</sub> /(-S-) <sub>R</sub>	-0.183	-4.483	144	(>NH) <sub>R</sub> /(>C<) <sub>R</sub>	-0.0356	0.506
72	(>CH <sub>2</sub> ) <sub>R</sub> /(>NH) <sub>R</sub>	0.307	-1.601	145	-NO <sub>2</sub> />C<	-0.127	2.940
73	(>CH-) <sub>R</sub> />C<	-0.188	2.037	146	-NO <sub>2</sub> /(>C<) <sub>R</sub>	-0.097	0.774

Note: R represents group on the ring; - show these values are not simulated because there are not any experiment data of organic compound including these Group-Adjacent Atom Pair.

lowing objective function:

$$OF = \sum_{j=1}^n \sum_i \left| \frac{Q_j^{exp} - Q_j^{cal}}{Q_j^{exp}} \right| \quad (8)$$

Where  $Q_j^{exp}$ ,  $Q_j^{cal}$  are the experimental value and the calculation value of the two properties.  $n$  is the number of compounds used in the regression. The equations parameters obtained by correlating data are listed in Table 1 and Table 2.

### METHOD COMPARISON AND ESTIMATION RESULTS

The group division in this work is the same as that of Joback and

Reid method. The Joback and Reid [1] method, the Marrero-Morejon and Pardillo-Fontdevila [5] method, and the Constantinou and Gani [6] method have been chosen as the comparative methods to examine this work. The results are listed in Table 3.

Melting points of 824 compounds and boiling points of 1171 compounds are divided into six groups; the average deviations between the calculated value by this method and experimental data are listed in Table 4.

The proposed method gives, as shown in Table 3, more accurate predictions for  $T_m$  and  $T_b$  than the compared methods. For the Constantinou and Gani method, the accuracy of the  $T_b$  estimation data is too small to illustrate the real situations. Moreover, the Constantinou and Gani method is of questionable utility.

**Table 3. Average deviations of property by the proposed method and GIC method**

Method	Melting points (K)			Boiling points (K)		
	No. of compds	AAE*/K	APE**/%	No. of compds	AAE*/K	APE**/%
Joback and Reid [1]	388	22.6	11.2	438	12.9	3.6
Marrero-Morejon and Pardillo-Fontdevila [7]	-	-	-	507	6.48	1.73
Constantinou and Gani [8]	312	14.03	7.23	392	5.35	1.42
The proposed Method	824	17.33	7.53	1171	7.74	1.58

Note: \*, \*\* represent average absolute error and average percent error respectively; - show no values of these items in the literature 7.

**Table 4. The correlation accuracy and the error distribution of  $T_m$  and  $T_b$  for 6 groups compound**

Compound	$T_m$ (K)			$T_b$ (K)		
	No. of data	Absolute deviation	No. of data	Absolute deviation	No. of data	Absolute deviation
Aliphatic hydrocarbon	157	14.3	6.99	224	5.06	1.11
Naphthenic hydrocarbon	85	14.45	7.37	95	7.84	1.44
Aromatic hydrocarbon	71	19.14	9.42	80	7.5	1.53
O,S compounds	280	17.97	6.85	461	7.84	1.65
N compounds	48	14.1	6.18	73	7.49	1.61
Halogenated compounds	183	20.53	8.73	238	10.17	1.95
Total	824	17.33	7.53	1171	7.74	1.58

**Table 5. Results of  $T_m$  and  $T_b$  estimation for 2-methyl-3-ethyl pentane**

No.	$\alpha_i$	$\nu_i$	No.	$\alpha_i$	$\nu_i$	Groups	$\sum_{k=1}^{n_i} \nu_{ik}$	$\Delta T_{mi}$	$\Delta T_{bi} \times 10^{-2}$	Group-adjacent atom pairs	$n_j$	$\Delta T_{mj}$	$\Delta T_{bj} \times 10^{-2}$
1	6.000	0.372	5	5.292	0.328	CH <sub>3</sub> -	1.604	-0.511	-3.262	-CH <sub>3</sub> />C<	4	0.132	1.234
2	6.000	0.372	6	5.292	0.328	-CH <sub>2</sub> -	0.656	-0.079	2.967	>CH <sub>2</sub> />C<	4	0.0358	0.504
3	6.928	0.430	7	4.000	0.248	>CH-	0.525	0.158	2.259	>CH-/>C<	6	-0.010	0.922
4	6.928	0.430	8	4.472	0.277								

**Table 6. Results of  $T_m$  and  $T_b$  Estimation for 1-bromo-2,4-dimethyl benzene**

No.	$\alpha_i$	$\nu_i$	No.	$\alpha_i$	$\nu_i$	Groups	$\sum_{k=1}^{n_i} \nu_{ik}$	$\Delta T_{mi}$	$\Delta T_{bi} \times 10^{-2}$	Group-adjacent atom pairs	$n_j$	$\Delta T_{mj}$	$\Delta T_{bj} \times 10^{-2}$
1	8.062	0.365	6	7.483	0.339	CH <sub>3</sub> -	0.730	-0.511	-3.262	-CH <sub>3</sub> /(<C<) <sub>R</sub>	2	0.027	0.037
2	8.062	0.365	7	6.782	0.307	(=CH-) <sub>R</sub>	0.972	-0.082	2.243	(=C<) <sub>R</sub> />C<	2	0.142	0.83
3	6.633	0.300	8	6.782	0.307	(=C<) <sub>R</sub>	0.953	0.400	11.43	(=C<) <sub>R</sub> /(<C<) <sub>R</sub>	6	0.0388	0.366
4	7.483	0.339	9	7.483	0.339	-Br	0.3340	0.441	3.714	(=C<) <sub>R</sub> /-Br	1	-0.046	0.391
5	7.348	0.333								(=CH-) <sub>R</sub> /(<C<) <sub>R</sub>	6	0.0352	0.356

Note: R represents group on the ring.

## CONCLUSION

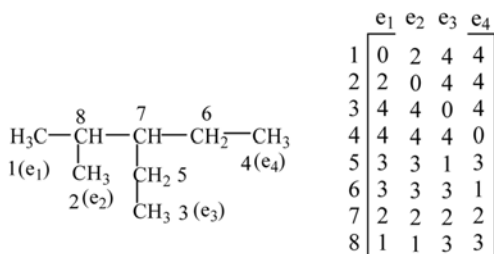
The specific environment and position of groups in the molecule have been considered, and a composite group vector space (CGVS) method for melting and boiling point of organic compounds developed. Expressions for  $T_m$  and  $T_b$  have been established and the numeric values of relative equation parameters obtained. The average percent deviations of estimation of the above two properties are 7.53 and 1.58, respectively, which shows that the present method demonstrates significant improvement in applicability to predict the above properties, compared to the conventional group methods.

## APPENDIX: ILLUSTRATIVE EXAMPLES OF ESTIMATION

The procedure for estimating the physical properties of organic compounds can be readily performed. First, a structural formula and corresponding topologic space matrix for the molecule are drawn. Then module  $\alpha_i$  of group  $i$  in the molecule and corresponding module index  $v_i$  may be calculated. Finally, when each  $\sum_{k=1}^{n_i} v_{ik}$  and  $n_i$  along with the values of the corresponding group and group-adjacent atom pairs parameters found in Table 1 and Table 2 are substituted into the relative correlation, the estimation value of the physical property can be obtained. Two examples are shown as follows.

**Example 1: 2-methyl-3-ethyl pentane.** The experimental data of the melting and boiling temperatures are 158.18 K and 388.81 K.

(a) The structural formula and the corresponding vector space matrix:



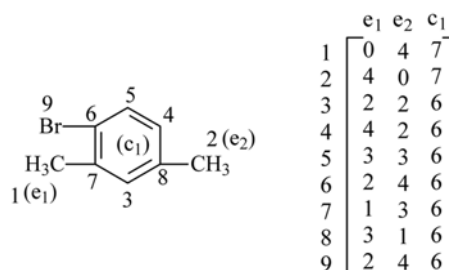
(b) The values of module  $\alpha_i$ , corresponding module index  $v_i$ ,  $\sum_{k=1}^{n_i} v_{ik}$ ,

$n_i$ , and group and group-adjacent atom pairs parameters are listed in Table 5.

(c) When the values in Table 5 are substituted into Eq. (6) and Eq. (7), the estimated values of melting and boiling temperatures are 146.60 K and 380.25 K, respectively. The percent deviations are -7.32% and -2.20%.

**Example 2: 1-bromo-2, 4-methyl benzene.** The experimental data of melting and boiling temperatures are 273.15 K and 478.15 K.

(a) The structural formula and the corresponding vector space matrix:



(b) The values of module  $\alpha_i$ , corresponding module index  $v_i$ ,  $\sum_{k=1}^{n_i} v_{ik}$ ,  $n_i$ , and group and group-adjacent atom pairs parameters are listed in Table 6.

(c) When the values in Table 6 are substituted into Eq. (6) and Eq. (7), the estimated values of melting and boiling temperatures are 286.09 K and 481.27 K, respectively. The percent deviations are 4.74% and 0.65%.

## REFERENCES

1. K. G. Joback and R. C. Reid, *Chem. Eng. Commun.*, **57**, 233 (1987).
2. D. Ambrose, *Correlation and estimation of vapour-liquid critical properties. 1. Critical temperatures of organic compound*, NPL reports in chemistry 92, National Physical Laboratory, Teddington, U.K. (1978).
3. D. Ambrose, *Vapor-liquid critical properties: II. Critical pressure and critical volume*, NPL reports in chemistry 107, National Physical Laboratory, Teddington, U.K. (1980).
4. W. J. Lyman, W. F. Reehl and D. H. Rosenblatt, *Handbook of chemical property estimation methods*, American Chemical Society, Washington, DC (1990).
5. J. Marrero-Morejón and E. Pardillo-Fontdevila, *AIChE J.*, **45**, 615 (1999).
6. L. Constantinou and R. Gani, *AIChE J.*, **40**, 1697 (1994).
7. K. R. Hall and L. B. Beach, *selected values of properties of hydrocarbons and related compounds*, Thermodynamics Research Center, Texas A & M university, Texas (1981).
8. B. E. Poling, J. M. Prausnitz and J. P. O'Connell, *The properties of gases and liquids*, 5th ed. McGraw-Hill Book Co., New York (2001).
9. R. C. West and M. J. Astle, *CRC handbook of data on organic compounds*, CRC Press Inc., Boca Raton, FL (1985).
10. V. Majer and V. Svoboda, *Enthalpies of vaporization of organic compounds*, Blackwell Scientific Publication, Oxford (1985).