

Wiener Matrix Invariants

M. Randić,* Xiaofeng Guo,† T. Oxley, H. Krishnapriyan, and L. Naylor

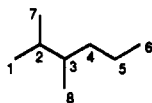
Department of Mathematics and Computer Science, Drake University, Des Moines, Iowa 50311

Received May 18, 1993*

We report on some properties of Wiener matrices as illustrated for alkanes. We start by considering the problem of graph reconstruction from Wiener matrix. Next, we consider the eigenvalues of the Wiener matrices. We also consider the row sums of the Wiener matrices, from which an index JJ , similar to the connectivity index χ and Balaban's J index, was constructed. The new invariants were tested for suitability for the structure-property relationship by analyzing regressions of novel invariants for selected properties of octanes.

INTRODUCTION

Recently we proposed a novel matrix, the Wiener matrix (W), as a potential source of structural invariants for need of QSAR (quantitative structure-activity study).¹ The (i, j) entry in the Wiener matrix is based on enumeration of paths in a graph which contain the path (i, j) . If we restrict the attention solely to paths of length one (i.e., bonds) we obtain, in the case of 2,3-dimethylhexane



the following matrix:

$$W_0 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{matrix} & \begin{pmatrix} 0 & 7 & 0 & 0 & 0 & 0 & 0 & 0 \\ 7 & 0 & 15 & 0 & 0 & 0 & 7 & 0 \\ 0 & 15 & 0 & 15 & 0 & 0 & 0 & 7 \\ 0 & 0 & 15 & 0 & 12 & 0 & 0 & 0 \\ 0 & 0 & 0 & 12 & 0 & 7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 7 & 0 & 0 & 0 \\ 0 & 7 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 7 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{matrix}$$

This is a sparse matrix with all entries being zero except those corresponding to the adjacent vertices. When the entries above the main diagonal in this matrix are added, one gets the Wiener number W . This way of deriving W is tantamount to the rules

Multiply the number of vertices on each side of a bond and add all such contributions.

This is precisely the procedure for getting the W number as outlined already by Wiener in his seminal paper on structure-property relationships.² The sparse matrix W_0 contains all the information on a graph, that is, the graph can be reconstructed from the W_0 matrix. This is obvious since the matrix is but a weighted adjacency matrix. Hence by replacing all nonzero entries by 1, one gets the adjacency matrix A of a graph.

Our prime interest is to find different structural matrices for a graph and later study the invariants of such matrices. Although sparse matrices such as W_0 are of interest, we will focus attention to dense matrices (i.e., matrices with nonzero entries, except on the main diagonal). It is always possible

to reduce a dense matrix to a weighted adjacency matrix; the procedure is trivial if one knows the adjacency relationships. The opposite, to reconstruct a graph from a dense matrix when adjacencies have yet to be deduced, could be computationally intense.

Here we will continue to explore the properties of the Wiener matrix. By definition the entries on the main diagonal of the matrix are zero. Off-diagonal entries count various path subgraphs. The element (i, j) counts path subgraphs in G that contain the path from vertex i to vertex j . Since every path intersects itself, the entries corresponding to paths between terminal vertices are necessarily equal to 1. For any other pair of vertices (i, j) the entry in the matrix gives the number of paths of which the path (i, j) is a subgraph. In the case of 2,3-dimethylhexane we obtain the following dense Wiener matrix:

$$W = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{matrix} & \begin{pmatrix} 0 & 7 & 5 & 3 & 2 & 1 & 1 & 1 \\ 7 & 0 & 15 & 9 & 6 & 3 & 7 & 3 \\ 5 & 15 & 0 & 15 & 10 & 5 & 5 & 7 \\ 3 & 9 & 15 & 0 & 12 & 6 & 3 & 3 \\ 2 & 6 & 10 & 12 & 0 & 7 & 2 & 2 \\ 1 & 3 & 5 & 6 & 7 & 0 & 1 & 1 \\ 1 & 7 & 5 & 3 & 2 & 1 & 0 & 1 \\ 1 & 3 & 7 & 3 & 2 & 1 & 1 & 0 \end{pmatrix} \end{matrix}$$

The entries of the matrix can be easily found by using a rule:

The element (i, j) is given as a product of the number of vertices on each side of the path (i, j) .

For example, to get the element $W_{2,4}$ we multiply 3 by 3, since at the terminal vertices of path 2,4, the vertices 2 and 4, there are three vertices. Hence, $W_{2,4} = 3 \times 3 = 9$.

Recently, we examined Wiener matrices for alkanes and reported several regularities for the particular sequences constructed from W_{ij} elements. The members of the sequences were obtained by summing all W_{ij} elements corresponding to vertices at equal (graph) distance, i.e., by summing contributions of paths of a same length.³ In the case of 2,3-dimethylhexane one obtains the sequence

$$70, 42, 21, 8, 2$$

The first entry in the sequence is the Wiener number W . Hence, we may refer to the other numbers in the above sequence as the "higher" Wiener numbers, $^k W$. The sum of entries in the sequence gives a "hyper-Wiener" number WW . Properties of such sequences and their construction by a recursive approach have been already outlined.³

† On leave from the Department of Mathematics, Xinjiang University, Wulumuqi Xinjiang 830046, People's Republic of China.

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

In this contribution we want to focus attention on the eigenvalues of $\langle W \rangle$. We will also examine structural information in the row sums of the Wiener matrices. First we will consider the graph reconstruction problem.

GRAPH RECONSTRUCTION

Given a matrix $\langle W \rangle$ one of the first problems to consider is reconstruction of the graph. If reconstruction is possible, there is no loss of information in representing a graph by the $\langle W \rangle$ matrix. If this would not be possible, or if the answer is ambiguous, i.e., there are two different graphs with the same $\langle W \rangle$ matrix, then the representation of a graph by the Wiener matrix involves a significant loss of information. The adjacency matrix A and a graph are in 1-1 correspondence. The distance matrix D can be simply reduced to A (by replacing all non-1 entries by 0). Thus in both cases the reconstruction is trivial; that is, the equivalence of information in a graph diagram and (0,1) adjacency matrix is obvious. It is not obvious that $\langle W \rangle$ also allows a reconstruction. If the matrix W can be reduced to the sparse matrix W_0 , we are assured that W contains all the necessary information. The reconstruction is not trivial since we do not know which vertex is next to which. Given a matrix (such as matrix W shown above for 2,3-dimethylhexane), the task is to draw the graph belonging to such a matrix, that is, to show the adjacencies.

By looking in the Wiener matrix shown, and pretending that we do not know that it was derived from 2,3-dimethylhexane, we can immediately recognize vertices 1, 6, 7, and 8 as terminal vertices (of degree $d = 1$). Only paths between terminal vertices can result in a contribution of 1 for the matrix element of $\langle W \rangle$. If we would know the distances between the terminal vertices, the reconstruction would be possible. Zaretskij⁴ has proved a theorem which states that, in trees, information on distances between terminal vertices suffices for graph reconstruction.

Rows of the matrix corresponding to terminal vertices have all entries less than n , the number of vertices in G , or the number of rows or columns in the matrix. In other rows there will be at least one entry greater than n because of the partitioning of the number n into two element partitions other than $1 + (n - 1)$. On the other hand an entry $(n - 1)$, (in our case 7), shows a terminal bond, because only terminal bonds partition the vertices into $1 + (n - 1)$ partition. Hence, looking at the matrix W we deduce that the following pairs of vertices (1, 2), (2, 7), (3, 8), and (5, 6) are adjacent. In each case the terminal vertex has been emphasized.

To continue the reconstruction, consider now vertices next to terminal vertices. Vertex 2 has terminal neighbors 1 and 7, hence, vertex 2 with the terminal vertices 1, 7 makes a group of three. Since the entry (2, 3) in $\langle W \rangle$ is 15, this means that vertex 2 has valence three, hence it is connected, in addition to vertices 1 and 7, only to vertex 3. However, vertex 3 is already connected to vertex 8, so by being adjacent also to 2 (which is connected to 1 and 7), this leads to partitioning $5 + 3$. Thus vertex 4, a neighbor of vertex 3, will have entry 15. Similarly, we find that vertex 5 is next to 4, which completes the reconstruction.

The reconstruction of a graph from the Wiener matrix would be simple if the following conjecture proves true:

Conjecture: Consider one row or one column at a time. Identify the largest entry (i, j) in the row or the column and replace that entry by 1. After completing all the rows, make the matrix symmetrical by introducing entries 1 where necessary. The resulting matrix is the adjacency matrix of the graph (which allows reconstruction trivially).

Table 1. Selected Eigenvalues of the Wiener Matrix for Octane Isomers

isomer	λ_1	λ_2	λ_{-1}
n-octane	57.1698	5.7986	-17.1147
2M	52.6122	3.9487	-17.1287
3M	48.4059	4.5405	-17.4004
4M	46.6606	5.3477	-17.6574
3E	42.2041	4.2759	-18.0148
22MM	44.4713	2.7559	-17.3299
23MM	42.0589	3.5154	-17.6893
24MM	43.4185	3.0505	-17.4100
25MM	47.7238	2.3133	-17.1421
33MM	38.5332	3.9135	-17.6695
34MM	39.2901	3.6392	-17.7238
2M3E	37.4277	2.7202	-18.0437
33ME	34.1415	2.7202	-17.8681
223MMM	34.9935	2.1639	-17.6780
224MMM	39.1411	1.3908	-17.3416
233MMM	33.4679	2.4556	-17.7117
234MMM	37.0246	1.9282	-17.7197
2233MMMM	30.3305	0.6241	-17.6241

Reconstruction of a large matrix does not introduce novel difficulties. A presence of several branches with bridging vertices will require several steps before initially separate fragments in the process of reconstruction are joined. Of course, we have not proved that reconstruction is possible; rather we illustrated how such reconstruction can be made. In graph theory proofs of reconstruction (i.e., equivalence of the information) are generally difficult. Reconstruction corresponds to resurrection of a graph from fragmentary information. Here we have a complete matrix, but lack the explicit information on the connectivity. However, we were able to derive adjacencies by identifying terminal vertices and their successive neighbors. The smallest and the largest entries in each row offer useful information. The smallest entry in each row directly shows the partitioning of n . For 2,3-dimethylhexane in the second row we find the smallest entry 3. Hence, vertex 2 is associated with the partitioning $(3 + 5)$. Factoring the largest entry, similarly, one may deduce the partitioning of the vertices.

EIGENVALUES OF $\langle W \rangle$

In Table I we have listed selected eigenvalues of $\langle W \rangle$ for the 18 isomers of octane. The first eigenvalue of a matrix, associated with the everywhere positive eigenvector (i.e., the eigenvector the coefficients of which all have a same sign), often has some structural interpretation. Lovasz and Pelikan⁵ suggested that the first eigenvalue of the adjacency matrix is an index of the molecular branching. For a graph embedded on a regular hexagonal grid it was suggested that the first eigenvalue of the distance/distance matrices⁶ is an indicator of graph folding, that is, the departure from the strictly (geometrically) linear structure.

The first eigenvalue of the Wiener matrix of octane isomers varies from 57.17 for n-octane to 30.33 for the highly branched 2,2,3,3-tetramethylbutane. In contrast, the "last" eigenvalue shows an unusual constancy: all the values lie between -17.11 and -18.04. This constancy of the last eigenvalue has some practical use: It can serve as check on the accuracy of a Wiener matrix. If for a family of isomers the last eigenvalue departs from the almost constant value, there is probably an input error in the matrix. The correlation between the first eigenvalue λ_1 of $\langle W \rangle$ and the first eigenvalue of the adjacency matrix A is very good, the two quantities parallel each other (R , the coefficient of regression is 0.958, and the F ratio is 176). Hence, the first eigenvalue of $\langle W \rangle$ offers an alternative index for the molecular branching.

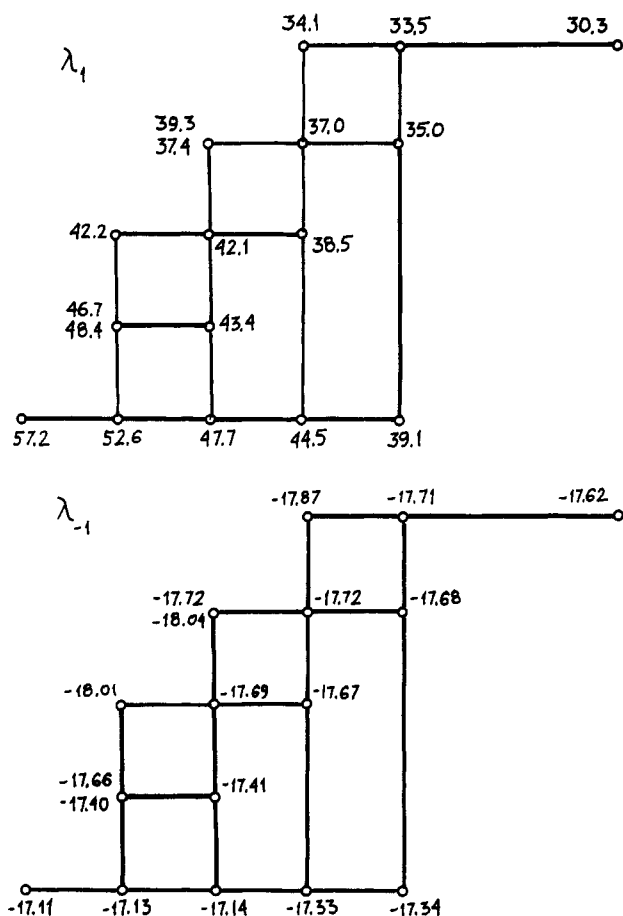


Figure 1. Regular variations of the first eigenvalue (λ_1) and the last eigenvalue (λ_{-1}) of the Wiener matrix displayed on the (p_2, p_3) grid.

In Figure 1 we have depicted the values for λ_1 and λ_{-1} (the first and the last eigenvalues of $\langle W \rangle$) on a coordinate grid derived from the count of paths of length 2 and paths of length 3. The (p_2, p_3) coordinate grid illustrates regularities in isomeric variations of various physicochemical properties for alkanes.⁷ However, one can extend such considerations also to purely mathematical properties. Moreover, in such instances one need not be concerned with uncertainties due to experimental errors in measurements. As we see, both selected eigenvalues show a regular variation along the coordinate axes: both λ_1 and λ_{-1} decrease along p_2 and p_3 axes (by λ_{-1} being negative, the absolute values of λ_{-1} increase).

JJ INDEX

Another invariant considered here is an index referred to as the JJ index. This index is derived in analogy with the construction of the connectivity index χ from the adjacency matrix⁸ and Balaban's index J from the distance matrix D .⁹ First one derives all row sums, R_i , and these numbers are used in a formula $1/(R_i R_j)^{1/2}$ for all pairs of adjacent vertices (i.e., for all bonds). By adding all bond contributions in adjacency matrix we obtain the connectivity index, and for the distance matrix we obtain Balaban's J index. In the case of the Wiener matrix index in this way we get a novel index JJ.

In Table 2 in the first numerical column we have listed JJ values for the 18 isomers of octane. Balaban's J index has shown considerable discrimination among graph isomers. The smallest trees having the same J index are found among trees on $n = 12$ vertices.¹⁰ In view of the fact that the entries of the Wiener matrix are larger (and more diverse) than those

Table 2. JJ Index and Higher JJ Indices for the Octane Isomers

	¹ JJ	² JJ	³ JJ	⁴ JJ
<i>n</i> -octane	0.136 39	0.014 705	0.001 554 1	0.000 161 86
2M	0.148 50	0.022 295	0.001 739 1	0.000 175 65
3M	0.160 51	0.022 734	0.002 981 5	0.000 245 64
4M	0.163 78	0.023 262	0.002 867 4	0.000 423 27
3E	0.181 56	0.024 779	0.004 061 3	0.000 520 58
22MM	0.174 05	0.040 950	0.002 468 5	0.000 309 90
23MM	0.182 67	0.033 284	0.004 632 3	0.000 389 45
24MM	0.175 80	0.032 396	0.003 471 1	0.000 472 97
25MM	0.164 57	0.031 121	0.002 160 9	0.000 249 64
33MM	0.194 79	0.039 214	0.005 480 5	0.000 508 61
34MM	0.193 63	0.033 589	0.005 752 6	0.000 492 05
2M3E	0.202 66	0.035 664	0.005 871 2	0.000 806 36
33ME	0.215 65	0.043 307	0.008 943 3	0.000 581 99
223MMM	0.212 23	0.056 432	0.009 287 3	0.000 506 38
224MMM	0.194 53	0.053 272	0.002 970 1	0.000 813 65
233MMM	0.230 69	0.056 211	0.008 060 6	0.000 366 40
234MMM	0.204 62	0.045 274	0.006 886 2	0.000 593 30
2233MMMM	0.234 69	0.085 903	0.001 148 0	0.000 000 00

of the distance matrix, we expect for the JJ index an even higher discrimination power. Indeed, the preliminary work shows that the smallest pair of trees having some the JJ index occur for graphs with 18 vertices.¹¹

The structural interpretation of indices such as χ , J , and JJ, depends on the structural meaning of the row sums R_i of the matrices from which they have been constructed. In the case of the connectivity index the row sums of the adjacency matrix give the valence of vertices. The weighting function $1/\sqrt{d_i d_j}$ gives more weight to the "exposed" bonds, (e.g., terminal bonds) and less weight to the internal "hidden" bonds. In the distance matrix the row sums correspond to partitioning of the Wiener number W (in fact partitioning of $2W$). Here the "central" vertices have smaller row sums, thus the "inside" bonds make a greater contribution to the molecular descriptor than the peripheral bonds. We assumed the same weighting algorithm (i.e., the reciprocal square root function) for JJ.

In the case of the Wiener matrix the peripheral vertices (see Figure 2) have smaller row sums and thus make a larger contribution to JJ than the centrally located bonds associated with larger row sums. The difference between J and JJ reflects the difference between the distance matrix and the Wiener matrix. The entries in the distance matrix increase with the separation of the vertices, while in the Wiener matrix it is just the opposite. Since for the structure-property relationship the influences of close neighbors are more important than those more distant neighbors, one may anticipate that in modeling the structure-property relationship the Wiener matrix may be of more interest than the distance matrix.

Some regularities for the magnitudes of the row sums in octane isomers, the R_i numbers, can be seen from Figure 2. For example, among terminal vertices the row sums (R_i) are smaller if the vertex is closer to a branching site and the row sums are larger if the terminal vertex is farther from a branching site. Thus we find the smallest $R_i = 16$ in 2,2,3,3-tetramethylbutane and the largest $R_i = 28$ in *n*-octane. In *n*-alkanes the R_i values for terminal vertices are a simple function of the length of the path. They are the triangular numbers of 1, 3, 6, 10, 15, 21, 28, ... (with the successive difference of 1, 2, 3, 4, ...) for molecules from ethane to octane, respectively. Observe that if R_i values are known for *n*-alkane then R_i values in the "tail" part (i.e., the part formed by bridging vertices of degree 2) of other isomers can be often simply constructed by subtracting successively constant multiples from the corresponding values for *n*-alkane. For example, the "tail" part in 2-methylheptane and 3-methylheptane can be obtained

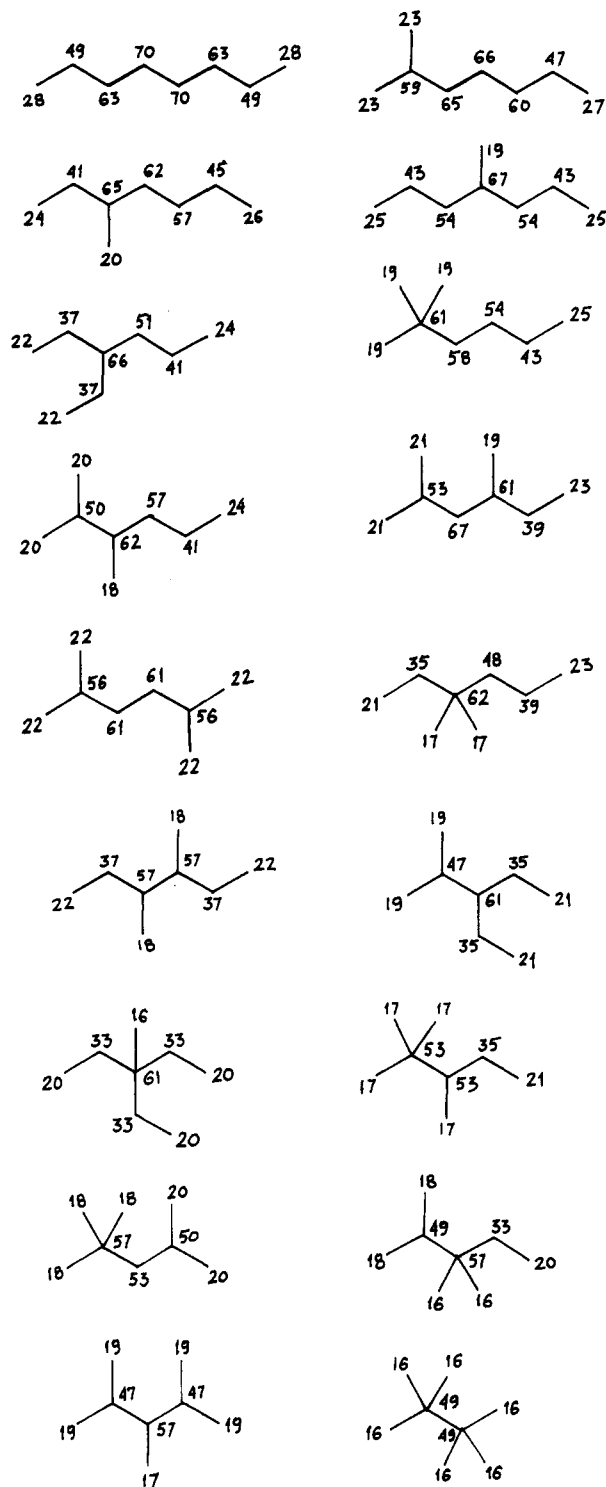


Figure 2. Variations of atomic ID numbers (the row sums in the corresponding Wiener matrices) among isomers of octane.

from *n*-octane, (the tail being 28, 49, 63, 70, 70), by subtracting 1, 2, 3, 4, 5, and 2, 4, 6, 8, respectively.

STRUCTURE-PROPERTY CORRELATION

Regressions Based on a Single Descriptor. Recently a comprehensive study of numerous properties of octane isomers based on the use of a single molecular descriptor was reported.^{12,13} The main conclusion of those studies was that when one considers dependence of physicochemical properties for molecules of a same size the properties do not mutually correlate. This means that different physicochemical properties depend in a different way on the inherent structural

Table 3. Statistical Information for Single Variable Regressions of Selected Physicochemical Properties of Octanes

		descriptors						
		<i>W</i>	<i>WW</i>	<i>WW/W</i>	x_1	x_2	x_{-1}	<i>JJ</i>
bp	<i>S</i>	5.32	5.31	5.40	5.65	3.90	6.24	5.73
	<i>R</i>	0.538	0.540	0.514	0.443	0.786	0.140	0.417
	<i>F</i>	6.5	6.6	5.8	3.9	25.8	0.3	3.4
CT	<i>S</i>	10.03	10.03	10.04	10.02	9.09	8.87	9.99
	<i>R</i>	0.027	0.041	0.013	0.059	0.425	0.469	0.097
	<i>F</i>	0.0	0.0	0.0	0.1	23.5	4.5	0.2
CP	<i>S</i>	0.771	0.742	0.729	0.661	1.29	1.03	0.555
	<i>R</i>	0.846	0.858	0.864	0.889	0.447	0.702	0.923
	<i>F</i>	37.7	41.9	44.0	56.8	3.7	14.5	86.9
<i>S</i>	<i>S</i>	2.47	2.23	1.88	2.62	2.41	4.92	2.46
	<i>R</i>	0.848	0.878	0.853	0.826	0.856	0.263	0.849
	<i>F</i>	41.0	53.7	37.4	34.3	43.9	0.12	41.3
DH	<i>S</i>	0.233		0.184	0.264	0.161	0.393	0.271
	<i>R</i>	0.807		0.803	0.743	0.914	0.106	0.728
	<i>F</i>	30.0		25.4	19.8	80.7	0.32	18.1
<i>R</i> ²	<i>S</i>	0.082	0.085	0.075	0.094	0.096	0.168	0.102
	<i>R</i>	0.899	0.890	0.911	0.861	0.858	0.439	0.837
	<i>F</i>	67.2	60.8	67.8	45.9	44.6	0.38	37.5
dens	<i>S</i>	0.0094	0.0093	0.0091	0.0085	0.0116	0.0069	0.0080
	<i>R</i>	0.594	0.597	0.622	0.681	0.103	0.805	0.723
	<i>F</i>	8.2	8.3	9.5	13.0	0.2	27.6	16.5
<i>H_v</i>	<i>S</i>	1.44	0.141	1.47	1.59	0.943	2.09	1.62
	<i>R</i>	0.725	0.738	0.6712	0.648	0.892	0.014	0.633
	<i>F</i>	17.8	19.2	16.4	11.6	62.5	0.0	10.7
<i>H_f</i>	<i>S</i>	1.13	1.11	1.40	1.19	0.827	1.23	1.20
	<i>R</i>	0.486	0.506	0.468	0.393	0.767	0.290	0.374
	<i>F</i>	5.0	5.5	4.5	2.9	22.9	1.5	2.6
DF	<i>S</i>	0.784	0.779	0.769	0.746	0.868	0.660	0.720
	<i>R</i>	0.441	0.453	0.475	0.579	0.116	0.655	0.564
	<i>F</i>	3.9	4.1	4.7	5.9	0.2	12.0	7.5
<i>H_c</i>	<i>S</i>	1.13	1.12	1.15	1.19	0.83	1.24	1.20
	<i>R</i>	0.489	0.509	0.471	0.396	0.769	0.288	0.379
	<i>F</i>	5.0	5.6	4.6	3.0	23.2	1.4	2.7
<i>V_c</i>	<i>S</i>		0.0105	0.0109	0.0010			0.0108
	<i>R</i>		0.744	0.699	0.740			0.701
	<i>F</i>		17.3	14.3	18.2			14.5
<i>T_c</i>	<i>S</i>	11.73	11.76	11.77	11.81	11.45	11.73	11.83
	<i>R</i>	0.129	0.110	0.101	0.064	0.251	0.130	0.010
	<i>F</i>	0.3	0.2	0.2	1.1	0.2	0.274	0.0
<i>P_c</i>	<i>S</i>	0.625	0.595	0.576	0.523	1.03	0.904	0.436
	<i>R</i>	0.865	0.878	0.886	0.908	0.560	0.687	0.937
	<i>F</i>	47.5	54.0	58.6	74.6	7.3	14.3	114.3

factors of a molecule. Therefore it should not be surprising that *different* mathematical descriptors have been proposed as the best to describe different properties. Previous studies have revealed that some molecular properties are "easy" to describe; i.e., there are many descriptors that give a satisfactory correlation for such selected properties.¹² However, some other properties of the same molecules are apparently difficult to describe, being "elusive". They clearly do not correlate well with any or most of the available molecular descriptors. This fact ought to be the prime stimulus for the continuing search for novel molecular descriptors, topological indices, and graph invariants in general.

Here we will test several of the descriptors derived from the Wiener matrix. First we will use them as a single descriptor. Of course, molecular descriptors can be used in multiple regression analysis. Thus the deficiency of any single descriptor can be compensated by properties of another descriptor. We will discuss in the next section that aspect of regression analysis based on Wiener matrix invariants.

We will confine our attention how novel descriptors based on the Wiener matrix perform on selected properties of octane isomers. In Table 3 we have listed the standard errors (*S*), the coefficients of regression (*R*), and the *F* ratios (*F*), for 14 properties of octanes. We tested the following descriptors: the Wiener number *W*, the hyper-Wiener number *WW*, the first eigenvalues x_1 and the second eigenvalues x_2 , and the index *JJ*. In Table 5, we have listed the numerical values for

Table 4. Statistical Information for Multiple Regressions of Selected Physicochemical Properties of Octanes Based on Two Descriptors

		descriptors						
		WW, W	WW, x_1	WW, JJ	W, x_1	W, JJ	x_1, JJ	x_1, x_2
bp	S	4.96	2.81	4.12	4.47	4.20	5.63	3.82
	R	0.648	0.902	0.775	0.727	0.763	0.502	0.810
	F	5.4	32.8	11.2	8.4	10.5	2.5	14.3
CT	S	9.29	7.51	8.69	8.45	8.41	10.07	7.45
	R	0.444	0.690	0.545	0.579	0.585	0.237	0.695
	F	1.8	36.8	3.2	3.8	3.9	0.4	7.0
CP	S	0.75	1.11	0.46	0.63	0.49	0.54	0.61
	R	0.864	0.691	0.957	0.906	0.943	0.932	0.914
	F	20.6	6.8	66.8	32.0	55.9	46.1	35.4
S	S	2.34	2.48	2.50	2.34	2.27		2.06
	R	0.874	0.857	0.854	0.874	0.881		0.904
	F	24.2	20.8	20.2	24.2	26.0		33.4
DH	S	0.210	0.148	0.216	0.159	0.199		0.160
	R	0.858	0.932	0.848	0.921	0.874		0.917
	F	20.8	49.3	19.2	42.0	24.1		41.1
R^2	S	0.066	0.070	0.078	0.065	0.064		0.074
	R	0.858	0.932	0.916	0.941	0.944		0.923
	F	56.2	49.4	38.9	57.8	60.8		43.0
dens	S	0.0090	0.0051	0.0064	0.0074	0.0063		0.0066
	R	0.669	0.907	0.847	0.791	0.852		0.836
	F	5.7	32.4	17.8	11.7	18.6		16.3
H_v	S	1.28	0.82	1.31	1.18	1.17		0.97
	R	0.804	0.918	0.777	0.838	0.840		0.892
	F	13.7	32.7	10.6	17.6	18.0		29.3
H_t	S	1.04	0.57	0.92	0.81	0.943		0.79
	R	0.662	0.887	0.677	0.793	0.707		0.808
	F	4.7	25.9	5.9	12.8	7.5		14.1
DF	S	0.80	0.56	0.63	0.73	0.67		0.68
	R	0.455	0.779	0.710	0.585	0.674		0.652
	F	2.0	10.8	7.1	3.9	6.2		5.6
H_c	S	1.05	0.57	0.92	0.95	0.95		0.79
	R	0.625	0.887	0.677	0.709	0.707		0.809
	F	4.8	25.9	5.9	7.6	7.5		14.2
V_c	S	0.009		0.011	0.009	0.010		
	R	0.808		0.706	0.822	0.769		
	F	13.2		6.9	14.6	10.1		
T_c	S	11.7	10.6	9.6	11.2	10.4		11.6
	R	0.292	0.540	0.617	0.393	0.526		0.308
	F	0.7	2.9	4.6	1.4	2.9		0.8
P_c	S	0.63	0.37	0.27	0.65	0.37		0.50
	R	0.871	0.948	0.978	0.861	0.958		0.920
	F	23.6	62.8	167.4	21.5	83.2		41.9

the selected properties examined to help those interested in verifying the numerical details and wishing to reproduce the results.

Let us first focus attention on the individual rows in Table 3. They tell us how different properties of octanes are described when different descriptors are used. From Table 3 we see that Wiener-based descriptors lead to a fair regression (R above 0.900) only for the critical pressures (CP and P_c), and the enthalpy of vaporization (DH). For other properties W as a single descriptor only approaches a fair regression. The coefficient R is above 0.800 for the entropy S , the mean radius R^2 , the liquid density, and H_v . For the remaining properties, including the boiling point (bp), none of the indices used alone produce a satisfactory regression.

If we focus attention on the columns of Table 3 (i.e., on the individual descriptors), then we see first that the eigenvalues x_2 and x_{-1} are the only two columns that dramatically differ in overall performance (good or bad) compared with the remaining descriptors. In particular we find x_2 to be good a descriptor for the enthalpy of vaporization (DH). Among the novel descriptors we find the JJ index overall as the best descriptor for the critical pressure. Notice that all descriptors correlate with R^2 , an observation already reported in ref 12. This signals that the mean square radius is not a critical property on which one should test molecular descriptors, at least in the case of octane isomers.

We have included in Table 3 information on all regressions tested despite the fact that most of the regressions are found not to be good. The reason for their inclusion is to see how they improve a regression when they are combined with other descriptors. On the other hand, even such limited regressions allow one to classify the molecular properties studied. Consider F ratios and the grouping of the molecular properties according to the F values found. We will use the following range of the values:

$$F < 10; 10 < F < 25; 25 < F$$

Then the properties are grouped as follows:

CT, T_c , H_v , H_c , bp, and DF associated with $F < 10$

H_v , density, V_c , and DH associated with $10 < F < 25$

S , R^2 , CP, and P_c associated with $25 < F < 100$

The F values are those found in regressions using the first eigenvalue x_1 as the descriptor. Essentially the same classification follows if one uses WW, W , the ratio WW/ W , or the JJ index. The exception is, as already mentioned, the second eigenvalue, which exhibits a significantly different behavior. The above grouping separates properties into those that are satisfactorily described by a single (simple) descriptor from those that need additional descriptors if a better regression is desired.

A close comparison of the regression coefficients for the four descriptors shows overall a considerable parallelism. The corresponding regression coefficients are similar, whether the values are above 0.800, whether they are of an intermediate value for R around 0.500, or whether R is very low (below 0.200). This can be understood if one compares the correlation matrix for the descriptors:

	WW	λ_1	JJ	x_1
WW	1.000	0.992	-0.983	-0.966
λ_1		1.000	-0.993	-0.957
JJ			1.000	0.963
x_1				1.000

Regressions Based on Two Descriptors. In Table 4 we have summarized the results for the regressions using two descriptors. For several of the considered properties we get regressions with R over 0.900. As we see, the combination of Wiener number W and hyper-Wiener number (the third column in Table 4) does not significantly improve the regressions). All other combinations lead to at least one regression which is best among the half-dozen considered (with $R > 0.900$). The combination of WW and x_1 gives the standard error $S = 2.81$ for the boiling points in octanes. This is slightly better than the standard error based on the connectivity indices $^1\chi$ and $^2\chi$.¹² All two-parameter regressions except one involving WW and x_1 give better R and S values than were derived from the two connectivity indices $^1\chi$ and $^2\chi$.¹² All regressions for R^2 have regression coefficients above 0.900, and are better than the results based on the connectivity indices (with $R = 0.886$). On the other hand, for the properties for which the connectivity indices $^1\chi$ and $^2\chi$ are very good descriptors (like heats of vaporization), the present descriptors show very limited improvement. Exceptions are the properties which are described well by path-based descriptors.

We may conclude that the novel invariants are suitable descriptors in multiple regression analysis for properties that

Table 5

isomer	BP	CT	CP	S	DH	R ²	dens	H _v	H _t	DF	H _c	V _c	T _c	P _c
n-octane	125.665	296.2	24.64	111.67	9.915	2.0449	0.7025	73.19	-49.82	4.14	1222.77	0.492	568.76	24.54
2M	117.647	288	24.8	109.84	9.484	1.8913	0.6980	70.30	-51.50	3.06	1221.08	0.408	595.57	24.52
3M	118.925	292	25.6	111.26	9.521	1.7984	0.7058	71.30	-50.82	3.29	1221.76	0.464	565.60	25.13
4M	117.709	290	25.6	109.32	9.483	1.7673	0.7046	70.91	-50.69	4.00	1221.89	0.476	561.67	25.09
3E	118.534	292	25.74	109.43	9.476	1.7673	0.7136	71.70	-50.40	3.59	1222.19	0.455	565.42	25.74
22MM	106.84	279	25.6	103.42	8.915	1.6744	0.6953	67.70	-53.71	2.56	1218.88	0.478	549.80	24.76
23MM	115.607	293	26.6	108.02	9.272	1.6464	0.7121	70.20	-51.13	4.23	1221.45	0.468	563.42	25.94
24MM	109.429	282	25.8	106.98	9.029	1.6142	0.7004	68.50	-52.44	2.80	1220.15	0.472	553.45	25.23
25MM	109.103	279	25	105.72	9.051	1.6449	0.6935	68.60	-53.21	2.50	1219.37	0.482	549.99	24.54
33MM	111.969	290.84	27.2	104.74	8.973	1.7377	0.7100	68.50	-52.61	3.17	1219.97	0.443	561.95	26.19
34MM	117.725	298	27.4	106.59	9.316	1.523	0.7200	70.20	-50.91	4.97	1221.68	0.466	568.78	26.57
2M3E	115.45	295	27.4	106.06	9.209	1.5525	0.7193	69.70	-50.48	5.08	1222.11	0.443	567.02	26.65
33ME	118.259	305	28.9	101.48	9.081	1.5214	0.7274	69.30	-51.38	4.76	1221.20	0.455	576.51	27.71
223MMM	109.841	294	28.2	101.31	8.826	1.4306	0.7161	67.30	-52.61	4.09	1219.98	0.436	563.43	26.94
224MMM	99.238	271.15	25.5	104.09	8.402	1.401	0.6919	64.87	-53.57	3.13	1219.01	0.468	543.89	25.34
233MMM	114.76	303	29	102.06	8.897	1.4931	0.7262	68.10	-51.73	4.52	1220.86	0.455	573.49	27.83
234MMM	113.467	295	27.6	102.39	9.014	1.3698	0.7191	68.37	-51.97	4.32	1220.61	0.461	566.34	26.94
2233MMMM	106.47	270.8	24.5	93.06	8.41	1.4612	0.8242	66.20	-53.99	4.88	1218.51	0.461	567.80	28.30

are global, i.e., the properties that are not bond additive. However, for bond additive properties the connectivity indices give better regressions. In a way, therefore, most of the new descriptors complement the well-established bond additive descriptors, such as $^1\chi$ and $^2\chi$.

HIGHER JJ INDICES

We have seen from Table 3 and Table 4 that the JJ index is a promising descriptor in many of the structure-property correlations, particularly when two or more descriptors are used. With JJ being a useful descriptor, particularly when combined with other descriptors (of different structural origin even if derived from the Wiener matrix), a question is raised: Can we generalize the JJ index and define indices that are structurally closely related to the JJ index, in analogy with the connectivity index $^1\chi$ and higher connectivity indices $^m\chi$?

We will show not only that this is possible, but that the proposed construction is quite general, hence, it can be extended to other molecular descriptors. Thus beside "higher" JJ indices we can, for example, also construct "higher" J indices, indices that would be structurally closely related to Balaban's J index. The generalized procedure is defined for any matrix. If one uses the adjacency matrix, the process leads to higher connectivity indices (or weighted path indices). If it is applied to the distance matrix, higher J indices follow. The same procedure can be applied to other recently proposed graph matrices, such as the Hosoya matrix¹⁴ and the restricted random walk matrix,¹⁵ as well as to the matrices already introduced in the literature: the extended distance matrix of Tratch, Stankevich, and Zefirov¹⁶ and the graph matrices of Kier and Hall.¹⁷

Let us illustrate the approach again on 2,3-dimethylhexane (the Wiener matrix of which was illustrated in the Introduction). For rows 1–8 the row sums R_i are, respectively,

$$R_1 = 20 \quad R_2 = 50 \quad R_3 = 62 \quad R_4 = 41$$

$$R_5 = 51 \quad R_6 = 24 \quad R_7 = 20 \quad R_8 = 18$$

To obtain the JJ index, or as we can now say, the first-order 1JJ index, for each bond of the graph we construct $1/(R_i R_j)^{1/2}$ so we have:

bond	contribution	numerical
1–2	$1/[(20)(50)]^{1/2}$	0.031 623
2–3	$1/[(50)(62)]^{1/2}$	0.017 961
2–7	$1/[(20)(50)]^{1/2}$	0.031 623
3–4	$1/[(62)(51)]^{1/2}$	0.017 784
3–8	$1/[(62)(18)]^{1/2}$	0.029 934
4–5	$1/[(51)(41)]^{1/2}$	0.021 869
5–6	$1/[(41)(24)]^{1/2}$	0.031 879

When summed, this gives the JJ index for 2,3-dimethylhexane $^1JJ = 0.182\ 673$.

To obtain "higher" JJ indices we consider, instead of bonds, i.e., paths of length 1, higher paths. Thus for 2JJ we consider all paths of length 2 and then apply the reciprocal square root algorithm: $1/(R_i R_j R_k)^{1/2}$ as illustrated below:

path	contribution	numerical
1–2–3	$1/[(20)(50)(62)]^{1/2}$	0.004 016 1
1–2–7	$1/[(20)(50)(20)]^{1/2}$	0.007 071 1
2–3–4	$1/[(50)(62)(51)]^{1/2}$	0.002 515 0
2–3–8	$1/[(50)(62)(18)]^{1/2}$	0.004 233 3
3–4–5	$1/[(62)(51)(41)]^{1/2}$	0.002 777 3
3–2–7	$1/[(62)(50)(20)]^{1/2}$	0.004 016 1
4–5–6	$1/[(51)(41)(24)]^{1/2}$	0.004 463 9
4–3–8	$1/[(51)(62)(18)]^{1/2}$	0.004 191 6

By summing the above contributions, we obtain $^2JJ = 0.033\ 284\ 9$. In similar manner we get 3JJ by considering all paths of length 3 and making the corresponding contributions by multiplying the row sum values for the four vertices involved in each such path. In columns 4 and 5 of Table 2 we list the higher JJ indices for octane isomers.

We mentioned that the construction of "higher" order descriptors is possible for any topological index which can be related to a matrix of a graph. Since Balaban's J index is derived from the distance matrix, one can generate similar "higher" order descriptors of J . We will illustrate the construction of the 2J index. We can now refer to the original Balaban index J as the first-order index 1J . Again we will consider 2,3-dimethylhexane. The row sums for the distance matrix now are

$$R_1 = 20 \quad R_2 = 14 \quad R_3 = 12 \quad R_4 = 14$$

$$R_5 = 18 \quad R_6 = 24 \quad R_7 = 20 \quad R_8 = 18$$

The construction of 2J follows from consideration of paths of length 2:

path	contribution	numerical
1-2-3	$1/[(20)(14)(12)]^{1/2}$	0.017 252
1-2-7	$1/[(20)(14)(20)]^{1/2}$	0.013 363
2-3-4	$1/[(14)(12)(14)]^{1/2}$	0.020 620
2-3-8	$1/[(14)(12)(18)]^{1/2}$	0.018 185
3-4-5	$1/[(12)(14)(18)]^{1/2}$	0.018 185
3-2-7	$1/[(12)(14)(20)]^{1/2}$	0.017 252
4-5-6	$1/[(14)(18)(24)]^{1/2}$	0.012 859
4-3-8	$1/[(14)(12)(18)]^{1/2}$	0.018 185

This gives ${}^2J = 0.136\ 217$. In a similar way other higher J indices can be obtained.

Indices which are structurally closely related, such as χ and ${}^m\chi$, J and mJ , or JJ and mJJ , may be good candidates for use as basis descriptors.^{18,19} Advantages of basis descriptors have been recently outlined.²⁰ The most important is the possibility that such basis descriptors can be orthogonalized^{21,22} and thus offer a direct and unambiguous way to interpret the results of multiple regression analysis.

ACKNOWLEDGMENT

We would like to thank the Upjohn Pharmaceutical Company, Kalamazoo, MI, and Dr. G. M. Maggiora in particular, for partial support of this research. We also thank Professor A. T. Balaban for useful comments on the manuscript while in preparation.

APPENDIX

The numerical values for the selected physicochemical properties of alkenes are listed in Table 5. The sources of the data are as follows:

Boiling points (bp), critical temperature (CT), critical pressure CP: Needham, D. E.; Wei, I.-C.; Seybold, P. G. *J. Am. Chem. Soc.* **1988**, *110*, 4186.

Entropy (S): Scott, D. W. *J. Chem. Phys.* **1974**, *60*, 3144.

Heat of vaporization (H_v), heat of formation (H_f), density: Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1966.

Heat of vaporization (DH): Garbalena, M.; Herndon, W. C. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 37.

Critical volume V_c : Balaban, A. T.; Feroiu, V. *Rep. Mol. Theory* **1990**, *1*, 133.

Quadratic mean radius (R^2): Altenburg, K. Z. *Phys. Chem. (Leipzig)* **1980**, *261*, 389.

REFERENCES AND NOTES

- (1) Randić, M. Novel molecular descriptor for structure-property studies. *Chem. Phys. Lett.* **1993**, *211*, 478-483.
- (2) Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, *69*, 17-20.
- (3) Randić, M.; Guo, Xiaofeng; Oxley, T.; Krishnapriyan, H. Wiener Matrix: Source of novel graph invariants. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 709-716.
- (4) Zaretskij, K. A. Postroenie dereva ponaboru rasstojanij mezdu visjacimni versinami. *Usp. Mat. Nauk* **1965**, *20*, 90-92.
- (5) Lovasz, L.; Pelikan, J. *Period. Math. Hung.* **1973**, *3*, 175.
- (6) Randić, M.; Kleiner, A. F.; DeAlba, L. M. Distance/Distance matrices. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, in this issue.
- (7) Randić, M.; Wilkins, C. L. On graph theoretical ordering of structures. *Chem. Phys. Lett.* **1979**, *63*, 332-336. Randić, M.; Wilkins, C. L. Graph theoretical ordering of structures as a basis for systematic searches for regularities in molecular data. *J. Phys. Chem.* **1979**, *83*, 1525-1540.
- (8) Randić, M. On the characterization of the molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609-6615.
- (9) Balaban, A. T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **1992**, *89*, 399-404.
- (10) Balaban, A. T.; Quintas, L. V. The smallest graphs, trees, and 4-trees with degenerate topological index J . *MATCH* **1983**, *14*, 213-233.
- (11) Randić, M.; Guo, Xiaofeng. To be published.
- (12) Randić, M. Comparative regression analysis. Regressions based on a single descriptor. *Croat. Chem. Acta* **1993**, *66*, 289-312.
- (13) Randić, M.; Seybold, P. G. Molecular shape as a critical factor in structure-property-activity studies. *SAR QSAR Environ. Res.* **1993**, *1*, 77-85.
- (14) Randić, M. Hosoya matrix—a source of new molecular descriptors. *Croat. Chem. Acta*, in press.
- (15) Randić, M. Restricted random walk on a graph as a source of novel molecular descriptors. *Theor. Chem. Acta*, in press.
- (16) Tratch, S. S.; Stankevich, M. I.; Zefirov, N. S. Combinatorial models and algorithms in chemistry. The expanded Wiener number—A novel topological index. *J. Comput. Chem.* **1990**, *11*, 899-908.
- (17) See: Hall, L. H. Computational aspects of molecular connectivity and its role in structure-property modeling. *Computational Graph Theory*; Nova Publishers: New York, 1990; pp 202-233.
- (18) Randić, M. Representation of molecular graphs by basic graphs. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 57-69.
- (19) Randić, M. Use of basis descriptors in structure-property studies. *J. Mol. Struct. (Theochem)*, in press. Randić, M.; Trinajstić, N. Viewpoint 4-comparative structure-property studies: the connectivity basis. *J. Mol. Struct. (Theochem)* **1993**, *284*, 209-221.
- (20) Randić, M. Similarity based on extended basis descriptors. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 686-692.
- (21) Randić, M. Orthogonal molecular descriptors. *New J. Chem.* **1991**, *15*, 517-525.
- (22) Randić, M. Resolution of ambiguities in structure-property studies by use of orthogonal descriptors. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311-320.