

Computer Generation of Nuclear Equivalence Classes Based on the Three-Dimensional Molecular Structure

K. Balasubramanian

Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-1604

Received December 16, 1994[®]

Euclidian distance matrices of molecular structures are used to partition the nuclei into equivalence classes. Graph-theoretical algorithms originally developed for the adjacency matrices are applied to the Euclidian distance matrices to generate the nuclear equivalence classes. Application to NMR spectroscopy is considered.

1. INTRODUCTION

The development of computational techniques for the vertex partitioning of graphs and for the machine generation of automorphism groups has received considerable attention in recent years.^{1–14} The related problems pertinent to the molecular structure seek to generate nuclear partitioning under the three-dimensional molecular symmetry. In fact, it is envisaged that faster algorithms for the machine perception of molecular symmetry would require, as a first step, algorithms for the generation of nuclear equivalence classes. Compared to the graph automorphism and isomorphism problems the nuclear partitioning problem has received much less attention.

The nuclear partitioning problem has several applications. The ¹³C and proton NMR depends on the nuclear equivalence classes.¹⁷ The number of equivalence classes yields the number of NMR signals, while the ratio of the number of elements in the various equivalence classes yields the intensity ratio in the case of proton NMR. Therefore an ideal starting point for such applications is the development of fast machine algorithms for the generation of nuclear equivalence classes. The hyperfine pattern in the ESR spectra of radicals is also dependent on the nuclear equivalence classes.^{15,16}

The generation of symmetry-adapted linear combination¹⁸ (SALC) of atomic orbitals of large polyatomic systems requires, as a first step, the nuclear equivalence classes since the atomic orbitals belonging to nuclei that are in different equivalence classes would not mix in a SALC. Since fullerenes containing large numbers of nuclei exhibit several nuclear equivalence classes, computational techniques could be especially useful in the machine generation of SALCs for larger fullerenes and thus in quantum chemical computations that make use of SALCs.

The objective of this study is to develop algorithms suitable for machine manipulation to generate nuclear equivalence classes. The algorithms developed up to now for this purpose are based on graph connectivity and not on the three-dimensional geometry of the molecule. We use the information contained in the Euclidian distance matrices for this purpose. These matrices have been used as molecular structural invariants elsewhere.¹⁹ Subsequent to the author's proposal to use the polynomials and other invariants derived from Euclidian matrices, Nikolic et al.²¹ have shown that Euclidian matrices and their invariants can successfully

discriminate conformational isomers. The Euclidian distance matrices can be readily generated from inputs such as the **Z**-matrix used in the Gaussian package of codes. The algorithms developed here are applied to several molecular structures.

Computational techniques for the vertex partitioning of graphs are based on the adjacency matrices. While these techniques are quite useful in graph theoretical problems, they do not always provide correct nuclear partitions for molecular structures as the latter depends on the three-dimensional molecular geometry. All graph theoretical techniques developed up to now use the adjacency matrices directly or indirectly for the purpose of vertex partitioning. Some of the techniques fail to differentiate isospectral points yielding incorrect vertex partitions. The adjacency matrices contain information only on the connectivity or neighborhood of the vertices and not the bond angles or dihedral angles. The nuclear partitions of molecular structures would depend not only on the neighborhood information but also on the bond angles and dihedral angles. Therefore, it is necessary to develop more general procedures that incorporate this three-dimensional molecular feature, especially for applications dealing with proton NMR, ESR, and so on.

2. COMPUTATIONAL ALGORITHMS

A. Preliminaries and Definitions Related to Graphs.

The adjacency matrix of a graph which depends only on the connectivity of the vertices of the graph is defined as

$$A_{ij} = \begin{cases} 1 & \text{if the vertices } i \text{ and } j \text{ are connected by an edge} \\ 0 & \text{otherwise} \end{cases}$$

A permutation of the vertices of the graph can be described by the permutation matrix **P** defined as

$$P_{ij} = \begin{cases} 1 & \text{if the vertex } i \text{ is permuted to } j \text{ by the permutation} \\ 0 & \text{otherwise} \end{cases}$$

The adjacency matrices **A** and **B** are said to be isomorphic if

$$P^T A P = B$$

where **P^T** is the transpose of the permutation matrix. A permutation of the vertices of the graph is said to be an

[®] Abstract published in *Advance ACS Abstracts*, March 1, 1995.

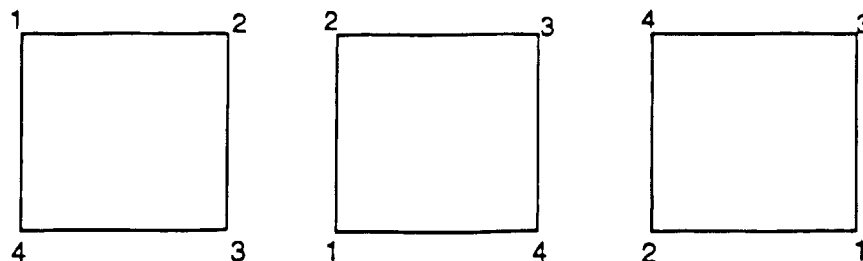


Figure 1. (A, left) The graph of cyclobutadiene, (B, middle) the effect of permutation (1 4 3 2) on the vertices of the graph in A, and (C, right) the effect of permutation (1 3 2 4) on the graph is A.

automorphic permutation if it satisfies

$$\mathbf{P}^T \mathbf{A} \mathbf{P} = \mathbf{A}$$

In general there are $n!$ possible permutation matrices for a graph which yield different isomorphisms. However, all of them may *not* satisfy the automorphic relation. All those permutations which satisfy the automorphism condition belong to a group called the automorphisms group.

Consider the cyclobutadiene graph shown in Figure 1A. The adjacency matrix of this graph with the labeling in Figure 1A is

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

The permutation (1 4 3 2) is an automorphic permutation of the graph in Figure 1A, but the permutation (1 3 2 4) is not an automorphic permutation and is merely an isomorphic permutation. Graphically the effects of these two permutations are shown in Figure 1 (parts B and C, respectively). Note that no edges are made or broken by the permutation (1 4 3 2), and thus the permutation (1 4 3 2) preserves the original connectivity. On the other hand, the permutation (1 3 2 4) does not preserve the original connectivity.

Mathematically the effect of these permutations is described by the matrix multiplications shown below for the (1 4 3 2) permutation.

$$\begin{aligned} \mathbf{P}^T(1\ 4\ 3\ 2)\mathbf{A}\mathbf{P}(1\ 4\ 3\ 2) &= \\ &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} = \mathbf{A} \end{aligned}$$

Note that the result of multiplying \mathbf{P}^T to the left of \mathbf{A} and \mathbf{P} to the right of \mathbf{A} for the above permutation gives the matrix \mathbf{A} back and thus the permutation (1 4 3 2) is an automorphic permutation.

Matrix multiplications for the permutation (1 3 2 4) are shown below:

$$\begin{aligned} \mathbf{P}^T(1\ 3\ 2\ 4)\mathbf{A}\mathbf{P}(1\ 3\ 2\ 4) &= \\ &= \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \neq \mathbf{A} \end{aligned}$$

The matrix thus obtained after the matrix multiplications is not equal to the original adjacency matrix \mathbf{A} and thus the permutation (1 3 2 4) is *not* an automorphic permutation.

Two vertices of a graph are said to be equivalent if there exists an automorphic permutation that transforms one vertex into other. For example, the permutation (1 4 3 2) which belongs to the automorphism group of the graph in Figure 1 transforms 1 to 4, 4 to 3, 3 to 2, and 2 to 1, and thus it is readily seen that all four vertices of the graph in Figure 1 are equivalent and they belong to a single equivalence class. Consider a chain graph containing four vertices such that labels 1–4 are assigned successively from left to right: the labeling is for the purpose of identifying equivalence classes only as the cardinalities of equivalence classes do not depend on labeling. The automorphism group which contains only two permutations $\{(1)(2)(3)(4), (14)(23)\}$ generates two different equivalence classes characterized by $\{1,4\}$, $\{2,3\}$.

Although the automorphism partitioning of vertices is defined above using the automorphism group, the generation of the automorphism vertex partitioning *does not* require the full automorphism group. Several techniques have been described elsewhere^{4–13} for the generation of automorphism partitionings without the knowledge of the automorphism groups. A recent comprehensive article⁴ describes different techniques such as the Morgan algorithm, the principal eigenvector algorithm, the canonical labeling scheme, the extended Morgan algorithm, the ring perception algorithm, and so on. It was noted that no single algorithm may be computationally suited for all graphs. Graphs which contain a large number of vertices that span a single vertex equivalence class are often considered to be difficult although the outcome is trivial.

We shall describe procedures appropriate for nuclear partitionings in the ensuing sections. First let us consider preliminaries pertinent to molecular structural characterization.

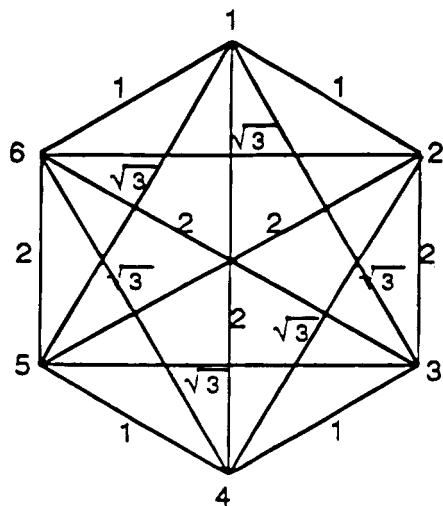


Figure 2. The Euclidian graph of benzene.

B. Euclidian Graphs. The three-dimensional Euclidian distance between two nuclei is given by

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$$

where (x_i, y_i, z_i) and (x_j, y_j, z_j) are the coordinates of the nuclei i and j , respectively. An Euclidian graph of a molecule is defined by the weighted adjacency matrix

$$D_{ij} = \begin{cases} d_{ij} & \text{if } i \neq j \text{ where } d_{ij} \text{ is the Euclidian distance} \\ & \text{between the nuclei } i \text{ and } j \\ 0 & \text{if } i = j \end{cases}$$

The Euclidian distance matrix thus defined has several advantages compared to the conventional Cartesian coordinates or internal coordinates. The Cartesian coordinates are relative in that they depend on the coordinate axis and thus vary through rotational, translational, and reflection operations. The Euclidian matrix, on the other hand, is invariant to such operations. The internal coordinates such as bond lengths, bond angles, and dihedral angles depend on the atoms that are used to define them. For example, the dihedral angles require four atoms not in one plane. Computer representation of such internal coordinates would require a neural network, in general, since bond distances depend on two atoms, bond angles on three atoms, while the dihedral angles on four atoms. The Euclidian matrix provides a simple matrix representation that can be used readily for matrix manipulations, eigenvector computations, and so on.

A permutation of the vertices of the Euclidian graph is an automorphic permutation of the Euclidian representation and a member of the point group if and only if the permutation matrix P satisfies

$$P^T D P = D$$

where P^T is the transpose of the permutation matrix. All such permutations of the nuclei which preserve the connectivity of the Euclidian graph of the molecular structure form a group called the Euclidian distance group by the author.¹⁴ The Euclidian distance group was shown to be isomorphic to the permutation-inversion group of Longuet-Higgins²⁰ when improper rotations in the point group generate new permutations not already contained in the

permutational subgroup of proper rotational operations. If the improper rotational operations (including mirror planes, inversion operation, etc.) generate the same permutations as the ones generated by proper rotational subgroup then the Euclidian distance group is shown to contain the unique permutations thus comprising only half the number of elements in the point group of the molecule.

The Euclidian graph can be illustrated with benzene whose Euclidian graph is shown in Figure 2, by setting all C—C bond distance to unities. The Euclidian distance matrix of benzene is shown below:

$$D = \begin{bmatrix} 0 & 1 & \sqrt{3} & 2 & \sqrt{3} & 1 \\ 1 & 0 & 1 & \sqrt{3} & 2 & \sqrt{3} \\ \sqrt{3} & 1 & 0 & 1 & \sqrt{3} & 2 \\ 2 & \sqrt{3} & 1 & 0 & 1 & \sqrt{3} \\ \sqrt{3} & 2 & \sqrt{3} & 1 & 0 & 1 \\ 1 & \sqrt{3} & 2 & \sqrt{3} & 1 & 0 \end{bmatrix}$$

The automorphic permutations of the Euclidian graph of benzene are shown below.

$$G = \{(1)(2)(3)(4)(5)(6), (1\ 5)(2\ 4)(3)(6), \\ (1\ 6\ 5\ 4\ 3\ 2), (1\ 6)(2\ 5)(3\ 4), (1)(4)(2\ 6)(3\ 5), \\ (1\ 5\ 3)(2\ 6\ 4), (1\ 4)(2\ 5)(3\ 6), (1\ 2)(3\ 6)(5\ 4), \\ (1\ 3\ 5)(2\ 4\ 6), (1\ 3)(2)(5)(4\ 6), (1\ 2\ 3\ 4\ 5\ 6), \\ (1\ 4)(2\ 3)(5\ 6)\}$$

The resulting group of 12 permutations span the D_6 group comprising proper rotations of the benzene molecule. The improper axes are apparently missing since these operations *do not* generate any new nuclear permutations for benzene. For example, the σ_h plane of the benzene molecule generates the same nuclear permutation as the identity operation.

The nuclear partitions are equivalence classes of nuclei under the automorphic permutations that preserve the Euclidian connectivity. That is, two nuclei are equivalent if there exists an automorphic permutation in the Euclidian distance group such that one nucleus is transformed to the other by this permutation. All equivalent nuclei are grouped together in one equivalence class. Two nuclei in the same equivalence class are chemically, geometrically, structurally, and magnetically equivalent. Thus properties that are based on individual nucleus such as chemical shifts and symmetry adaptation would be the same for the nuclei in an equivalence class.

The generation of the nuclear partitions is thus fundamentally important. All techniques for this purpose described in the next section would be based on the Euclidian distance matrices and not on the adjacency matrices. The graph theoretical techniques developed up to now if extendible to weighted graphs are used in the next section.

C. Nuclear Partitioning Algorithms. The earliest technique for graph-vertex partitioning problem is due to Morgan,⁷ although originally developed for labeling chemical structures. This algorithm has been extended further in several forms. The canonical labeling procedure developed by Randić^{12,13} is useful for the generation of vertex partitioning of graphs. The algorithm however is not suitable for graphs containing a large number of vertices and graphs which exhibit fewer vertex-equivalence classes. Graphs

which contain vertices of the same degree (regular graphs) and transitive graphs often pose problems.

The principal eigenvector algorithm is based on the eigenvalues and eigenvectors of the adjacency matrix of the graph. Suppose $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ is the set of n eigenvalues of the adjacency matrix A , the principal eigenvalue is defined as

$$\lambda = \max_i \{|\lambda_i|\}$$

It is the maximal absolute eigenvalue. The principal eigenvector U then satisfies

$$AU = \lambda U$$

Suppose the principal eigenvector is of the form (u_1, u_2, \dots, u_n) , two vertices i and j of the graph are said to be equivalent if

$$|u_i| = |u_j|$$

Consequently, based on the components of the principal eigenvector, one could divide the vertices into equivalence classes. However the algorithm fails for ordinary graphs if the vertices i and j are isospectral. That is, although isospectral vertices may not be equivalent, they would come out to be equivalent under the principal eigenvector algorithm.

Liu et al.¹⁰ have developed a more general version of the principal eigenvector algorithm in that instead of associating each vertex i with the component u_i , a row vector L_i is assigned to the vertex i . There are three types of vector generator algorithms described by Liu et al. which work satisfactorily.

The current nuclear partitioning algorithm employs a combination of the above techniques. Unlike ordinary graphs, no difficulties were encountered since Euclidian distance matrices are far more discriminating and contain structurally discriminating three-dimensional Euclidian distances. We found the following simple procedure based on Euclidian vertex degrees to be satisfactory for several molecular structures that we investigated with the exception of one polycyclic aromatic that we shall discuss subsequently. We shall also consider vectorial algorithms and more general techniques based on higher-order Euclidian matrices.

The Euclidian degree of the nucleus i in the molecule is defined as

$$e_i = \sum_{j \neq i} d_{ij}$$

where d_{ij} 's are Euclidian distances, and the sum is over all j except $j = i$. This corresponds to the sum of the Euclidian distance in the i th row of the Euclidian matrix. The resulting values e_1, e_2, \dots, e_n are called the Euclidian degrees of the nuclei (1, 2, ..., n , respectively). As a starting point, the nuclei can be divided into partitions simply based on the Euclidian degrees e_1, e_2, \dots, e_n of the nuclei. That is, two equivalent nuclei i and j should have the same Euclidian degrees e_i and e_j although the converse in general is not true as we found a counter example. Consequently, the Euclidian degree vector was found to be a good and useful starting point for generating nuclear partitions.

The ordinary degrees of vertices can be rarely used for the vertex automorphism partitionings. This is because there is too much "accidental degeneracy" in the vertex degrees of graphs, the worst case being regular graphs which would

also lead to a single set of vertex partitioning if the vertex degrees were used for the purpose of partitioning. Among the numerous molecular structures that we considered, we found only one structure which exhibited such accidental degeneracies in the Euclidian degrees. Since the Euclidian distances were defined for all the nuclei relative to the chosen nucleus, these distances provide a more complete basis for the nuclear partitions. To the contrary, the adjacency matrices are generated based on the nearest neighbors only. However powers of adjacency matrices can be used for vertex partitioning which are more discriminating.¹⁰

While the Euclidian degrees may just be a starting point for a more rigorous approach, the following procedure based on higher powers of Euclidian matrices is adopted in our algorithm. Let us define the k th order Euclidian degree index as

$$e_i^{(k)} = \frac{1}{s^k} \sum_j D_{ij}^k$$

where D^k is the k th power of the Euclidian distance matrix and s is a scaling factor introduced so that the k th order Euclidian degrees do not grow out of bound. The k th order Euclidian degrees thus introduced reduce to ordinary or first order Euclidian degrees when $k = 1$ with a unit scaling factor. In practice, we found that a scaling factor of 10 introduced both in the Euclidian distances and in the above definition keep the Euclidian degrees of multiple orders as manageable numbers.

The process of computing the equivalence class structures can be iterated for any order. That is, one can compute the equivalence class structures starting from the first order Euclidian degrees. Subsequently, the equivalence class structures are computed using the second order Euclidian degrees. The two equivalence class structures are compared. If they are the same, then convergence is reached, and thus the correct nuclear partitioning is generated. If convergence is not reached at the second iteration, one proceeds to obtain the third-order Euclidian degrees and the nuclear equivalence classes based on the third-order Euclidian degrees. If the nuclear equivalence classes obtained from the third-order degrees is the same as the second-order, then convergence is reached. Otherwise, the process is iterated until convergence is achieved. This process is mathematically formulated as follows.

Let $S_1^{(k)}, S_2^{(k)}, \dots, S_m^{(k)}$ be the disjoint set of nuclear equivalence classes generated in the k th iteration. That is

$$S = S_1^{(k)} \cup S_2^{(k)} \cup \dots \cup S_m^{(k)}$$

$$S_i^{(k)} \cap S_j^{(k)} = \phi$$

where S is the set of nuclei and ϕ is the null set. The algorithm for the k th order Euclidian degree partitions is shown on the next page.

A practical question on any iterative algorithm is related to its convergence. Such iterative algorithms are quite common in quantum chemistry. The Hartree-Fock self-consistent field (SCF), diagonalization of large matrices, etc. are based on iterative algorithms seeking convergence. For all practical problems convergence is usually achieved.

The Euclidian degrees were found to yield converged and correct nuclear partitions in almost two iterations for some

```

k ← 1
ei1 ← ∑j dij

compute S1(1) ∪ S2(1) ... ∪ Sm(1),
based on {e1(1), e2(1) ... en(1)}

* k ← k + 1
eik ← ∑j dijk

compute S1(k) ∪ S2(k) ... ∪ Smk(k)
if mk · eq · mk-1 and
Sjk · eq · Sj(k-1) then
convergence is reached; exit
else go to *

```

50 structures that we tested. Only for one structure a third iteration was required as we discuss. However, even for this structure convergence was reached immediately at the third iteration. We have not encountered a single structure that required beyond third-order Euclidian degrees for the computation of converged equivalence classes.

We have developed a computer code which computed Euclidian degrees to any desirable order until convergence. For very high orders the scaling factor *s* will have to be increased, and quadruple precision arithmetic had to be invoked. The author has developed such a code which makes use of quadruple precision and computes Euclidian degrees to order 50. The main concern for higher-order computations is accuracy. Modular arithmetic in conjunction with available quadruple precision can be used to compute the Euclidian degrees in octuple and higher precision. The quadruple precision itself yields 31 digit accuracy.

The principal eigenvector of the Euclidian matrices could also be used to generate the nuclear partitions. In this case we invoke the Givens-Housolder method to extract the eigenvalues and eigenvectors of the Euclidian matrices. Suppose $\lambda_1, \lambda_2, \dots, \lambda_n$ are the Euclidian eigenvalues then we have

$$DX_i = \lambda_i X_i$$

The components of the eigenvector X_i , whose absolute eigenvalue λ_i is maximal, could be used to partition the nuclei. More sophisticated algorithms described for ordinary graphs in ref 10 could also be used for the Euclidian graph.

We consider several molecular structures in the next section. It was found that the Euclidian degrees of the vertices were sufficient to generate the nuclear partitions. Among the molecules we considered, we found only a single structure for which the Euclidian degrees failed to yield the correct nuclear partitioning.

3. APPLICATIONS OF MOLECULAR STRUCTURES

We first illustrate the utility of Euclidian degree algorithm for the nuclear partitioning. Consider the anthracene mol-

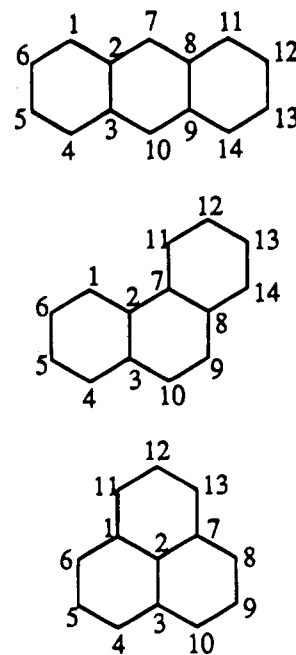


Figure 3. Structures of polycyclic aromatics containing three hexagons.

ecule as an example (Figure 3). All Euclidian distance matrices considered here were defined in terms of unit distances for the bonded atoms for simplicity. The C—C distances common to two rings could be slightly different from other C—C distances. This subtle difference is however not pertinent for the molecular symmetry or equivalency classes of polycyclic aromatics. Hence all the Euclidian distance matrices were defined in terms of unit distances between the bonded carbon atoms. The vertex labeling is arbitrary and does not influence the equivalence class structure.

The Euclidian vertex degrees for anthracene based on the labeling in Figure 3 is

{33.893 231, 26.643 607, 26.643 607, 33.893 231,
39.608 635, 39.608 635, 27.511 208, 26.643 607,
26.643 607, 27.511 208, 33.893 231, 39.608 635,
39.608 635, 33.893 231}

Based on these Euclidian degrees we obtain the following vertex equivalence classes:

{1, 4, 11, 14}, {2, 3, 8, 9}, {5, 6, 12, 13} {7, 10}

This is exactly what one expects to obtain using the complete molecular symmetry of anthracene. However we have accomplished the same without the use of molecular symmetry which is important for machine algorithm.

Let us consider phenanthrene and the peri condensed three-ring radical shown in Figure 3. For the phenanthrene molecule, the Euclidian degrees of the nuclei according to the vertex labels in Figure 3 are

{29.557 308, 23.951 757, 26.643 607, 33.893 321,
37.950 208, 36.120 980, 23.951 757, 26.643 607,
30.203 059, 30.203 059, 29.557 308, 36.120 980,
37.950 209, 33.893 231}

Base one these Euclidian degrees the nuclei are partitioned

Table 3. Computation of Higher-Order Euclidian Degrees for the Third Structure in Table 2

Iteration Number 1				
Euclidian degrees arranged				
9.408 351 32	9.271 658 88	9.197 354 55	8.909 946 72	8.361 727 07
8.361 727 07	7.882 497 71	7.524 240 14	6.850 824 93	6.850 824 92
6.786 263 83	6.567 675 94	6.563 488 91	6.344 901 02	6.255 011 01
6.170 184 18	6.004 341 48	5.900 999 10	5.712 971 43	5.609 629 05
5.307 273 71	5.307 273 71			
equivalence class information				
no. of equivalence classes: 19				
equivalence class partition: 2,2,2,1,1,1,1,1,1,1,1,1,1,1,1,1				
Iteration Number 2				
second order Euclidian degrees arranged				
6.612 368 35	6.582 402 11	6.470 180 75	6.354 075 16	5.984 159 03
5.943 491 62	5.621 355 84	5.456 409 69	5.003 654 72	4.991 172 76
4.971 412 13	4.842 147 58	4.782 980 44	4.700 180 41	4.687 487 70
4.637 060 24	4.495 728 09	4.397 138 27	4.317 363 79	4.246 645 62
4.052 184 46	4.045 118 13			
equivalence class information				
no. of equivalence classes: 22				
equivalence class partition: 1,1				
Iteration Number 3				
third order Euclidian degrees arranged				
4.836 907 64	4.788 335 90	4.732 809 13	4.618 857 36	4.349 294 24
4.340 525 95	4.104 597 75	3.957 397 20	3.627 135 29	3.620 529 08
3.616 876 53	3.515 210 57	3.482 371 91	3.397 426 50	3.392 018 81
3.349 295 14	3.252 103 85	3.189 171 24	3.121 614 42	3.064 388 26
2.923 491 97	2.922 134 24			
equivalence class information				
no. of equivalence classes: 22				
equivalence class partition: 1,1				
convergence reached after third iteration				

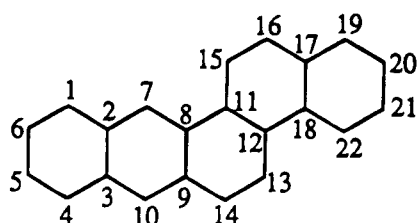


Figure 4. A cata-condensed polycyclic aromatic compound containing four hexagons which exhibits accidental degenerate Euclidian degrees. The Euclidian degrees of pairs of nuclei (4, 19), (3, 17), and (8, 11) are the same. See text for further discussion on second order Euclidian degrees which partition the nuclei correctly into equivalencies classes.

class structure does not depend on the labeling of the nuclei, no labels are given.

Table 2 shows all possible cata-condensed and two of the peri-condensed polycyclic aromatic compounds containing five hexagonal rings together with the Euclidian degrees of the vertices and the equivalence class information. All the Euclidian degrees were generated by assuming that all carbon-carbon bond distances are unities. That is, all hexagons were assumed to be regular. Among the structures shown in Table 2 with the exception of one structure discussed below, the Euclidian degrees were found to discriminate the nuclei properly into equivalence classes. The number of equivalence classes thus generated is also the number of ^{13}C NMR signals of the corresponding polycyclic aromatic compounds.

The third structure shown in Table 2 exhibits accidental degeneracy with respect to the ordinary Euclidian degrees. Figure 4 shows this structure with labelings for further discussion. As seen from Table 2, three sets of vertices each containing two members exhibit accidental Euclidian degree

degeneracy. This degeneracy has nothing to do with the numerical precision since the coordinates were read up to eight decimal points, and double precision arithmetic was used. The nuclei labeled 4 and 19, 3 and 17, and 8 and 11 (each pairs) exhibited the same Euclidian degrees. Yet we know that these nuclei are not equivalent. That is, the sum of the Euclidian distances from 4 to all other vertices is the same as 19 to all other vertices. This accidental degeneracy is, however, easily removed leading to convergence at the third iteration in our algorithm.

Table 3 illustrates the computation of higher-order Euclidian degrees for the third structure in Table 2. A scaling factor of 10 was used in all computations in Table 3. As seen from Table 3, although the first order Euclidian degrees yield incorrect equivalence class structure information, second and third-order Euclidian degrees yield identical nuclear partitions resulting in convergence at the third iteration. This is quite encouraging since most of the numerical procedures do not converge this rapidly.

The technique can be applied to nonplanar molecule without any additional difficulty. Let us illustrate it with ferrocene in the staggered form. The C-C distances as well as the metal to ring atom distances were taken to be unities for the purpose of computing the equivalence classes. Of course, these are not realistic bond distances but for the purpose of generating the equivalence classes these distances suffice. The Euclidian degrees of all the carbon atoms come out to be 16.853 369 in the model and thus generating a single equivalence class containing all 10 carbon atoms. Application of this method to the eclipsed conformation Ferrocene also lead to a single equivalence class containing all ten carbon atoms.

Table 4. Some Nonplanar Compounds, Their First and Second Order Euclidian Degrees, and the Equivalence Class Structures^a

molecule/structure	Euclidian degrees		equivalence classes
	first order	second order	
ethane (eclipsed)	0.846 410(6)	0.071 641(6)	6
ethane (staggered)	0.852 860(6)	0.072 737(6)	6
CH ₄ (tetrahedron)	0.848 528(4)	0.072 000(4)	4
PF ₅ (trigonal bipyramid)	0.629 253(3), 0.624 264(2)	0.039 455(3), 0.039 182(2)	3, 2
cubane (cube)	1.794 938(8)	0.322 180(8)	8
icosahedron	1.499 2283(12)	0.224 769(12)	12
octahedron	0.765 685(6)	0.058 627(6)	6
square bipyramid	0.843 398(4), 1.021 11(2)	0.077 540(4), 0.091 452(2)	4, 2
distorted square bipyramid	1.051 068(2), 0.960 669(2), 0.880 711(2)	0.100 802(2), 0.093 259(2), 0.086 267(2)	2, 2, 2
square pyramid	0.624 264(4), 0.565 685	0.038 142(4), 0.035 314	4, 1
edge-capped tetrahedron	1.180 191(2), 1.021 733(2), 1.009 735(1)	0.124 668(2), 0.113 150(2), 0.113 679	2, 2, 1
prismane	1.009 513(6)	0.101 912(6)	6

^a A scaling factor $s = 10.0$ was incorporated into both Euclidian distances and in the computation of the k th order Euclidian degrees.

Table 4 shows 13 nonplanar molecules or structures to which our computer code for obtaining the nuclear equivalence classes was applied. In the case of ethane we considered both the eclipsed and staggered conformations. As anticipated the actual values of the Euclidian degrees are different for the eclipsed and staggered conformations. That is, the Euclidian degrees differentiate conformers. However, the equivalence class partitions are the same for both conformers. As seen from Table 4, all 13 three-dimensional structures exhibit convergence with respect to equivalence class partitions at first and second order Euclidian degrees. That is, the equivalence classes predicted at first and second iterations are identical leading to convergence at the second iteration itself.

As seen from Table 4, the code based on Euclidian degrees predicts a single equivalence class for regular structures such as tetrahedron, cube, octahedron, icosahedron, dodecahedron, and so on. It is also interesting that the code converges at the second iteration for other structures such as square bipyramid, trigonal bipyramid, edge-capped tetrahedron, and so on.

In conclusion, the code developed here based on powers of Euclidian matrices for computing nuclear equivalence classes converges rapidly. Among 35 structures considered here and another 15 cases not shown here all but one structure converged within two iterations yielding correct nuclear equivalence classes. For the third polycyclic aromatic in Table 2 also shown in Table 3, the code based on higher-order Euclidian degrees required three iterations for convergence. Considering that iterative numerical procedures such as SCF, CI diagonalizations, etc. require several

iterations for convergence, indeed the Euclidian approach for generating the equivalence classes is very promising. The traditional graph theoretical methods would require several more iterations even for the planar polycyclic aromatics and may not always generate the correct equivalence classes which reflect three-dimensional molecular symmetry. In fact, the vertex degrees alone rarely suffice to generate equivalence classes, while as exemplified here first-order Euclidian degrees suffice for all but one structure. The developed code would find applications in proton-NMR, ESR, and in computing spin-lattice relaxation constants.

REFERENCES AND NOTES

- (1) Balasubramanian, K. Applications of Combinatorics and Graph Theory to Spectroscopy and Quantum Chemistry. *Chem. Rev.* **1985**, 85, 599–618.
- (2) Balasubramanian, K. Computational Techniques for the Automorphism Groups of Weighted Graphs. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 1146–1150.
- (3) Balasubramanian, K. Computational Techniques for the Automorphism Groups of Graphs. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 621.
- (4) Razinger, M.; Balasubramanian, K.; Munk, M. E. Graph Automorphism Perception Algorithms in Computer-Enhanced Structure Elucidation. *J. Chem. Inf. Comput. Sci.* **1993**, 33, 197–201.
- (5) Balaban, A. T.; Mekenyan, W.; Bonchev, D. Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities (HOC Procedures). I. Algorithms for Finding Graph Orbits and Canonical Numbering of Atoms. *J. Comput. Chem.* **1985**, 6, 538–551.
- (6) Balasubramanian, K. Symmetry Groups of chemical Graphs. *Int. J. Quantum Chem.* **1982**, 21, 411–418.
- (7) Morgan, H. L. Generation of Unique Machine Description of chemical Structures. A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, 5, 107–112.
- (8) Razinger, M. Extended Connectivity in Chemical Graphs. *Theor. Chim. Acta* **1982**, 61, 581–586.
- (9) Wipke, W. T.; Dyott, T. M. Stereochemically Unique Naming Algorithm. *J. Am. Chem. Soc.* **1974**, 96, 4834–4842.
- (10) Liu, X.; Balasubramanian, K.; Munk, M. E. Computational Techniques for Vertex Partitioning of Graphs. *J. Chem. Inf. Comput. Sci.* **1990**, 30, 263–269.
- (11) King, R. B. The Bonding Topology of Polyhedral Molecules. In *Chemical Applications of Topology and Graph Theory*; King, R. B., Ed.; Studies in Physical and Theoretical Chemistry; Elsevier: Amsterdam, 1983; Vol. 28, pp 108–122.
- (12) Randić, M. Symmetry Properties of Graphs of Interest in Chemistry II. Desargues-Levi Graph. *Int. J. Quantum Chem.* **1979**, 15, 663–682.
- (13) Randić, M.; Davis, M. I. Symmetry Properties of Chemical Graphs VI. Isomerizations of Octahedral Complexes. *Int. J. Quantum Chem.* **1984**, 26, 69–89.
- (14) Balasubramanian, K. Graph Theoretical Perception of Molecular Symmetry. *Chem. Phys. Lett.*, in press.
- (15) Balasubramanian, K. Combinatorial Methods in ESR Spectroscopy *J. Magn. Reson.* **1991**, 91, 45–56.
- (16) Balasubramanian, K. ESR Hyperfine Patterns for Radicals Derived from C₆₀ and C₇₀. *Chem. Phys. Lett.* **1993**, 210, 153.
- (17) Balasubramanian, K. Combinatorics of NMR and ESR Spectral Simulations. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 296–298.
- (18) Balasubramanian, K. The Combinatorics of Symmetry-Adaptation. *Theor. Chim. Acta* **1981**, 59, 47–54.
- (19) Balasubramanian, K. Geometry-Dependent Characteristic Polynomials of Molecular Structures. *Chem. Phys. Lett.* **1990**, 169, 224–230.
- (20) Longuet-Higgins, H. C. The Symmetry Group of a Non-Rigid Molecule. *Molecular Phys.* **1963**, 6, 445.
- (21) Nikolić, S.; Trinajstić, N.; Mihalčić, Z.; Carter, S. On the Geometric Distance Matrix and the Corresponding Structural Invariants of Molecular Systems. *Chem. Phys. Lett.* **1991**, 179, 21–28.

CI940136E