Title
COMPUTER-ASSISTED ENUMERATION OF NMR SIGNALS

Author
Balasubramanian, K.

Publication Date
1981-12-01

LBL-13693
Preprint

Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

Materials & Molecular Research Division

Submitted to the Journal of Magnetic Resonance

COMPUTER-ASSISTED ENUMERATION OF NMR SIGNALS

K. Balasubramanian

December 1981

Prepared for the U.S. Department of Energy under Contract W-7405-ENG-48

DISCLAIMER

Computer-assisted Enumeration of NMR Signals

K. Balasubramanian

Department of Chemistry
and
Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

Abstract

The NMR signals of both rigid and non-rigid molecules are enumerated using computer-assisted combinatorial techniques. These techniques essentially require the knowledge of the symmetry group of the molecule (cycle indices) from which NMR signals (such as ^{13}C NMR, proton NMR, etc.) are enumerated. The program can be used not only for structure elucidation but also for studying dynamic NMR processes and phenomena.

This work was supported by the Director, Office of Energy Research, Office of Basic Energy Sciences, Chemical Sciences Division of the U.S. Department of Energy under Contract Number W-7405-ENG-48.

1. Introduction

An important problem in magnetic resonance is the prediction of the number of NMR signals of a compound theoretically. It is not surprising that the number of NMR signals of a compound is intimately related to the topology of molecules and the associated symmetry. Randić and co-workers [1,2] have recently shown that the chemical shift of nuclei is related to their topology. Consequently, these authors have demonstrated that the chemical shifts can be predicted by a topological correlation of atomic environments. In this paper we consider another important topological aspect, namely, the prediction of the number of NMR signals of a compound using its symmetry.

Chemical applications of non-numerical computational techniques is on the increase in recent years [3,8]. Applications of group theory and graph theory are becoming quite important in several areas [9-21]. In this paper we shall use a theorem of Pólya [22] for enumerating the NMR signals and subsequently computerizing the procedure to have an automated algorithm which enumerates the NMR signals of any molecule. The theory behind the algorithm was developed by the present author [11]. For the enumeration of the NMR signals of non-rigid molecules we shall use the generalized wreath product groups discussed in detail elsewhere [11,13]. The computer-assisted procedures developed here should be of immense use in organic structure elucidation, understanding dynamic NMR processes and phenomena. This could become a part of artificial intelligence packages of interest to several workers [23,24].

2. Theoretical Formulation

Let D be the set of the nuclei of the same kind (such as H, ^{13}C etc.) in the molecule. For example, if ^{13}C NMR of naphthalene is under consideration, then D would be the set of 10 Carbon nuclei present in the molecule. Let R be a set containing just 2 elements. Let G be the point group or the permutation-inversion group of the molecule. Since G is the set of all permutational and composite permutation-inversion operations any $g \varepsilon G$ induces permutations on elements in D since D is just the set of nuclei of the same kind in the molecule. Consider the set F of all maps from D to R. The action of G on D can in turn be transferred to F by the following recipe. Every $g \varepsilon G$ acts on F as defined by the following formulae.

$$gf(d) = f(g^{-1}d) \text{ for every } d \varepsilon D.$$

Two maps f_i and $f_j \varepsilon F$ are equivalent if

$$f_i(d) = f_j(gd) \text{ for every } d \varepsilon D.$$

Maps in F that are equivalent can be grouped into the same equivalence class. Thus the group G partitions F into equivalence classes. Let us restrict ourselves to the maps in F which have the following structure. Let the elements of R be denoted by α_1 and α_2. Label the elements of D as $d_1, d_2, \ldots . d_n$ with $n = |D|$. Then consider a subset F_w of F with every $f_i \varepsilon F_w$ defined as follows.

$$F_i(d_j) = \begin{cases} \alpha_1 & \text{if } i \neq j,\ d_j \varepsilon D \\ \alpha_2 & \text{if } i = j \end{cases}$$

It can be seen that two nuclei d_i and d_j are magnetically equivalent if f_i is equivalent to f_j. Thus the equivalence classes of F_w are the magnetic equivalence classes of nuclei in the Set D. Hence the number of equivalence classes gives the number of NMR signals of the molecule.

If one introduces the concept of weight of an element r in R as a formal symbol w(r) used to book-keep the number of times any $r \epsilon R$ occurs in a function, then the weight of any function $f \epsilon F$ can be defined as the product of the weight of its images. Symbolically, the weight of f, W(f), is

$$W(f) = \prod_{d \epsilon D} w(f(d)).$$

Since any $f_i \epsilon F_w$ takes all $d_j \epsilon D$ to α_1 except when i = j, the weight of any f_i in F_w, $w = \alpha_1^{n-1} \alpha_2^1$ if α_1 is the weight associated with α_1 and α_2 is the weight assigned to α_2.

Pólya [22] proved a theorem which generates the equivalence classes of F from a group structure known as cycle index which we shall now define. Suppose any $g \epsilon G$ when it acts on D generates b_1 cycles of length 1, b_2 cycles of length 2 etc. Equivalently the cycle type of g acting on D is $(b_1, b_2, \ldots)$. Then we can associate with g a cycle representation $x_1^{b_1} x_2^{b_2} \ldots$. The cycle index of G, P_G, is defined as

$$P_G(x_1, x_2, \ldots) = \frac{1}{|G|} \sum_{g \epsilon G} x_1^{b_1} x_2^{b_2} \ldots .$$

Pólya showed that in generating function (G.F.) for the equivalence classes of F is given by the following substitution in the cycle index.

$$G.F. = P_G(x_k \rightarrow \sum_{r\varepsilon R} (w(r))^k).$$

The coefficient of $w = \alpha_1^{n-1} \alpha_2$ in G.F. gives the number of magnetic equivalence classes of nuclei in D or the NMR signals of nuclei in D. Thus in order to enumerate the NMR signals we need to evaluate the generating function.

3. Computer Enumeration

In this section we outline computational techniques which essentially generate the generating function from the cycle indices and consequently enumerate the NMR signals of any compound.

A. Algorithms

Any NMR generating function is of the form

$$G.F. = \frac{1}{|G|} \sum_{g\varepsilon G} C(g)$$

where

$$C(g) = (\alpha_1+\alpha_2)^{n_{1g}} (\alpha_1^2+\alpha_2^2)^{n_{2g}} \ldots (\alpha_1^i+\alpha_2^i)^{n_{ig}}$$

if the cycle type of g is $(n_{1g}, n_{2g}, \ldots . n_{ig})$. Thus we need procedures to generate each term $(\alpha_2^i+\alpha_2^i)^{n_{ig}}$ in the above product, to multiply each such term to form C(g) and sum over all g to evaluate the total generating functions.

In order to generate the terms in the binomial expansion $(\alpha_1^i+\alpha_2^i)^{nig}$ we use the procedure of generating the compositions of an integer into 2 parts. The composition (n_1, n_2) of an integer n into 2 parts is defined by

$$n_1 + n_2 = n \qquad n_1 \geq 0, \; n_2 \geq 0.$$

For example, all the compositions of 4 into 2 parts are shown below.

$$4 + 0$$

$$0 + 4$$

$$3 + 1$$

$$1 + 3$$

$$2 + 2$$

Each composition corresponds to a term in the generating function. For example, the composition 3 + 1 corresponds to the term $\alpha_1^3\alpha_2$ in the expansion $(\alpha_1+\alpha_2)^4$. Similarly, the composition 2 + 2 would correspond to $\alpha_1^2\alpha_2^2$ in the binomial expansion. We use the procedure given in Nijenhuis and Wilf (25) to generate the compositions of an integer into 2 parts. The procedure which generates compositions of an integer into a given number of parts in a lexicographic order is called NEXCOM. NEXCOM generates each composition $n_1 + n_2$ (or the term in the binomial expansion) as a vector (n_1, n_2). An algorithm called VEC "multiplies" each vector generated by NEXCOM with other vectors in the product C(g). It also finds the appropriate combinatorial coefficient for each such term in the product C(g). The main program assembles all the vectors and sums over all g in the group G.

B. Computer Programs, Input and Output Description.

The algorithms outlined in Section 3A were programmed into a computer. These set of subroutines and main program require the minimal input concerning the cycle indices. The program is more general in that it can handle the other irreducible representations of the point group of the molecule even though, for

the present purpose we need to give only the cycle index of the totally symmetric A1 representation. The input description for the program NMR which generates the NMR signals is shown in Table 1. Let us expound further here on the variables listed in Table 1 especially cards succeeding the card 3. Consider ^{19}F NMR of the rigid PF_5 molecule. The point group of this molecule is D_{3h} and its cycle index is shown below.

$$P_{D_{3h}} = \frac{1}{12}\left[x_1^5 + 2x_1^2x_3 + 3x_1x_2^2 + 4x_1^3x_2 + 2x_2x_3\right]$$

NCI would be 5 and the array ICOCI(I) = 1, 2, 3, 4, 2. The five terms should be given in the same order in which the coefficients were fed. For example, the term $x_1x_2^2$ will have the following card.

2 1 2 1 2 [NPRO = 2, NCI(J) = 1, 2,Iexp(I,J) = 1,2].

The input for all the five terms are shown below.

1	5	1			(x_1^5)
2	3	1	1	3	$(x_1^3x_3)$
2	1	2	1	2	$(x_1x_2^2)$
2	3	1	1	2	$(x_1^3x_2)$
2	1	1	2	3	(x_2x_3)

The program first reads the input and checks for internal consistency in the input. For example, any term $x_1^{b_1}x_2^{b_2}..x_n^{b_n}$ in the cycle index should satisfy the following condition:

$$\sum_i ibi = NNUCLI$$

If not, the program prints out an error message. The term just printed out by the program contains this error. The other error detected by the program is using the condition that

$$|G|\ P_G(x_k \to \sum_{r \in R} (w(r))^k)$$

Should be divisible by $|G|$. If it is not divisible by $|G|$, then it prints out an error message. The error is either in the set of coefficients or in the terms in the cycle index that is otherwise not detectible by the earlier criterion. The program also prints out the final result, namely, the number of NMR signals of the compound under consideration.

4. Examples

In this section we consider two examples of molecules whose ^{13}C and proton NMR signals will be enumerated. First, consider the polycyclic pericondensed benzenoid hydrocarbon shown in Fig. 1. Using our program NMR we will enumerate both the ^{13}C and proton NMR signals of this molecule. This molecule has 96 carbon nuclei. The point group can be seen to be D_{6h}. The cycle index for these carbon nuclei are shown below.

$$P_G = \frac{1}{24} [2x_1^{96} + 4x_6^{16} + 4x_3^{32} + 8x_2^{48} + 6x_1^8 x_2^{44}]$$

$$= \frac{1}{12} [x_1^{96} + 2x_6^{16} + 2x_3^{32} + 4x_2^{48} + 3x_1^8 x_2^{44}].$$

The input for this molecule is shown in Table 2. The program prints out the terms in the cycle index and the number of NMR signals of the compound. The output for this input is shown in Table 3. Now we shall consider the proton NMR of the same

molecule. The cycle index for the transformation of protons is shown below.

$$P_G = \frac{1}{12} [x_1^{24} + 7x_2^{12} + 2x_6^4 + 2x_3^8]$$

The input for the proton NMR of this molecule is shown in Table 4 and the output is in Table 5. One can immediately infer that this molecule gives rise to 10 ^{13}C resonances and 2 proton resonances. As another non-trivial example we consider the chiral macrocycle containing enforced cavities reported by Helgenson, Mazaleyrat and Cram (26) recently. It is shown in Fig. 2. This chiral molecule possesses only a four-fold axis of rotation and its point group can be seen to be C_4. The cycle index for the ^{13}C NMR is shown below.

$$P_G = \frac{1}{4} [x_1^{88} + 2x_4^{22} + x_2^{44}]$$

The input and output are shown in Table 6 and 7, respectively. From Table 7 we infer the low resolution ^{13}C NMR of this molecule would contain 22 signals. This molecule has 8 methyl rotors exhibiting torsion at room temperature and is thus an example of a non-rigid molecule. The symmetry of such a system can in general be described by generalized wreath product groups. The symmetry group of this non-rigid molecule is $C_4[C_3]$, where C_3 is the torsional group for each methyl protons. The cycle index of generalized wreath products can be obtained using the method described elsewhere. The cycle index for protons is shown below.

$$P_G = \frac{1}{4} [x_1^{40}Z_1^8 + 2x_4^{10}Z_4^2 + x_2^{20}Z_2^4]$$

where $Z_i = \frac{1}{3}(x_i^3 + 2x_{3i})$, the subscripts on the x variables are the products. The input for the proton NMR of this molecule can be constructed in the usual manner. The output is shown in Table 8. When the torsional permutations can be differentiated by a dynamic NMR at feasible experimental conditions, the group becomes C_4 and the cycle index is shown below

$$P_G = \frac{1}{4}[x_1^{64} + 2x_4^{16} + 3x_2^{32}].$$

The output for this case is in Table 9. Thus as one can infer from Tables 8 and 9, the molecule exhibits 16 proton resonances with a dynamic NMR which coalesce to 12 signals at high temperature because of torsional tunneling. Such dynamic effects can be efficiently studied with our present computer program.

ACKNOWLEDGMENT

This work was supported by the Director, Office of Energy Research, Office of Basic Energy Sciences, Chemical Sciences Division of the U. S. Department of Energy under Contract Number W-7405-ENG-48.

References

1. M. Randić and B. C. Gerstein, J. Mag. Resonance, 43, 207 (1981).
2. M. Randić, Chem. Phys. Lett. 58, 180 (1978).
3. K. Balasubramanian, J. J. Kaufman, W. S. Koski, and A. T. Balaban, J. Computational Chem. 1, 149 (1980).
4. M. Randić, Computers and Chemistry, 3, 5 (1979).
5. M. Randić, J. Am. Chem. Soc. 97, 6609 (1975).
6. M. Randić, J. Computational Chem. 1, 386 (1980).
7. K. Balasubramanian, ibid (in press).
8. K. Balasubramanian, ibid (in press).
9. A. T. Balaban (Editor) "Chemical Applications of Graph Theory" Academic, N.Y. 1976.
10. D. H. Rouvray, Chem. Soc. Rev. 3, 3551 (1974).
11. K. Balasubramanian, Theor. Chim. Acta. 51, 37 (1979).
12. K. Balasubramanian, Ibid., 53, 129 (1979).
13. K. Balasubramanian, J. Chem. Phys. 72, 665 (1980).
14. K. Balasubramanian, ibid., 73, 3321 (1980).
15. K. Balasubramanian, ibid., 74, 6824 (1981).
16. K. Balasubramanian, Theor. Chim. Acta. 59, 47 (1981).
17. K. Balasubramanian, J. Chem. Phys. (in press).
18. K. Balasubramanian, Theor. Chim. Acta. 59, 237 (1981).
19. M. Randić, Chem. Phys. Lett. 55, 547 (1978).
20. M. Randić, J. Chem. Imp. Computer , 18, 101 (1978).
21. H. Dolhaine, Computers & Chemistry, 5, 41 (1981).
22. G. Pólya, Acta. Math. 38, 145 (1937).
23. C. A. Shelley and M. E. Munk, Anal. Chem. 50, 1522 (1978).

24. L. M. Masinter, N. S. Sridharan, J. Lederberg and D. H. Smith, J. Am. Chem. Soc. 9 , 7702, 7714 (1974).

25. A. Nijenhuis and H. S. Wilf, "Combinatorial Algorithimal", Academic, N.Y. 1975.

26. R. C. Helgenson, J-P. Mazaleyrat and D. J. Cram, J. Am. Chem. Soc. 103, 3929 (1981).

Table 1. The Input Description for the Program NMR

Card	Variable	Format	Description of the Variable
1	Title		Alpha numeric title
2	NGCI	16I5	Number of cycle indices (always 1)
	NSUB		Number of substituents (always 2)
	NNUCLI		Number of nuclei whose NMR is being enumerated.
	MOD G		Order of the point group of the molecule.
3	Sym		Label of the irreducible representation (always A1)
4	NCI	16I5	Number of terms in the cycle index
5	ICOCI(I), I=1, NCI	16I5	Coefficients of each cycle type in the cycle index
	For each cycle type (term) in the cycle index give a card described as Card 6.		
	For	I=1,	NCI
6	NPRO,N(I,J), J=1, NPRO, IExp(I,J), J=1,NPRO	16I5	NPRO = number of terms in the cycle type $x_1^{b_1} x_2^{b_2} \ldots$. N(I,J): the superfixes of each term in the cycle type. IExp(I,J): the suffixes of each term in the cycle type.

Table 2. The Input for ^{13}C NMR of the Molecule in Figure 1

Card	
1	^{13}C NMR signals of the pericondensed benzenoid compound in Figure 1
2	1 2 96 12
3	A1
4	5
5	1 2 2 4 3
6	1 96 1
7	1 16 6
8	1 32 3
9	1 48 2
10	2 8 44 1 2

Table 3

```
 C13 NMR SGNALS OF THE PERICONDENCED BENZENOID COMPOUND IN FIG.1
   2
A1
NPRO,N(I),I=1,NPRO      1    96
 IEXPS     1
NPRO,N(I),I=1,NPRO      1    16
 IEXPS     6
NPRO,N(I),I=1,NPRO      1    32
 IEXPS     3
NPRO,N(I),I=1,NPRO      1    48
 IEXPS     2
NPRO,N(I),I=1,NPRO      2     8    44
 IEXPS     1     2
    ENUMERATICN OF NMR SIGNALS USING THE GENERATING FUNCTION TECHNIQUES
              NUMBER OF NMR SIGNALS          10
```

Table 3

Table 4. The Input for Proton NMR of the Molecule in Figure 1

Card				
1	Proton NMR of the pericondensed benzenoid compound in Figure 1			
2	1	2	24	12
3	Al			
4	4			
5				
6	1	7	2	2
7	1	24	1	
8	1	12	2	
9	1	4	6	
10	1	8	3	

```
  2
A1
NPRO,N(I),I=1,NPRO     1    24
 IEXPS     1
NPRO,N(I),I=1,NPRO     1    12
 IEXPS     2
NPRO,N(I),I=1,NPRO     1     6
 IEXPS     4
NPRO,N(I),I=1,NPRO     1     8
 IEXPS     3
    ENUMERATION OF NMR SIGNALS USING THE GENERATING FUNCTION TECHNIQUES
              NUMBER OF NMR SIGNALS          2
```

Table 6. The Input for the ^{13}C NMR of the Chiral Macrocycle in Figure 2

Card	
1	^{13}C NMR signals of the chiral macrocycle in Figure 2
2	1 2 88 4
3	A1
4	3
5	1 2 1
6	1 88 1
7	1 22 4
8	1 44 2

Table 7

```
 C13 NMR SIGNALS OF THE CHIRAL MACROCYCLE IN FIG.2
   2
A1
NPRO,N(I),I=1,NPRO     1    88
 IEXPS     1
NPRO,N(I),I=1,NPRO     1    22
 IEXPS     4
NPRO,N(I),I=1,NPRO     1    44
 IEXPS     2
    ENUMERATION OF NMR SIGNALS USING THE GENERATING FUNCTION TECHNIQUES
              NUMBER OF NMR SIGNALS          22
```

Table 8


```
PROTON NMR OF THE NON-RIGID CHIRAL MACROCYCLE IN FIG.2
   2
A1
NPRO,N(I),I=1,NPRO     1    64
 IEXPS     1
NPRO,N(I),I=1,NPRO     2    61     1
 IEXPS     1     3
NPRO,N(I),I=1,NPRO     2    58     2
 IEXPS     1     3
NPRO,N(I),I=1,NPRO     2    55     3
 IEXPS     1     3
NPRO,N(I),I=1,NPRO     2    52     4
 IEXPS     1     3
NPRO,N(I),I=1,NPRO     2    49     5
 IEXPS     1     3
NPRO,N(I),I=1,NPRO     2    46     6
 IEXPS     1     3
NPRO,N(I),I=1,NPRO     2    43     7
 IEXPS     1     3
NPRO,N(I),I=1,NPRO     2    40     8
 IEXPS     1     3
    ENUMERATION OF NMR SIGNALS USING THE GENERATING FUNCTION TECHNIQUES
              NUMBER OF NMR SIGNALS          12
```

```
 PROTON NMR OF THE RIGID MOLECULE IN FIG.2
   2
A1
NPRO,N(I),I=1,NPRO      1    64
 IEXPS     1
NPRO,N(I),I=1,NPRO      1    16
 IEXPS     4
NPRO,N(I),I=1,NPRO      1    32
 IEXPS     2
    ENUMERATION OF NMR SIGNALS USING THE GENERATING FUNCTION TECHNIQUES
              NUMBER OF NMR SIGNALS          16.
```

Figure Captions

Figure 1. A pericondensed benzenoid hydrocarbon. For the computer-assisted enumeration of ^{13}C NMR of this molecule, see Tables 2 and 3. For proton NMR, see Tables 4 and 5.

Figure 2. Achiral macrocycle. For the computer-assisted enumeration of NMR signals (^{13}C and proton) of this molecule see Tables 6 to 9.

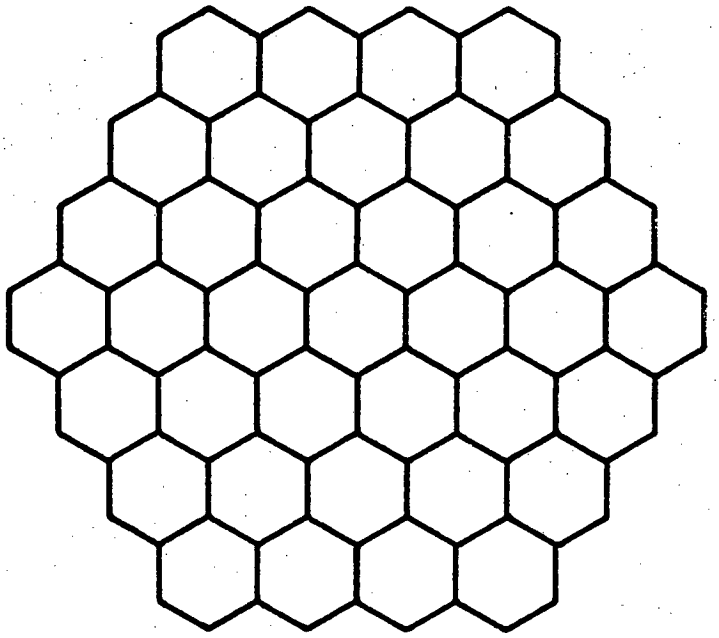

Fig. 1

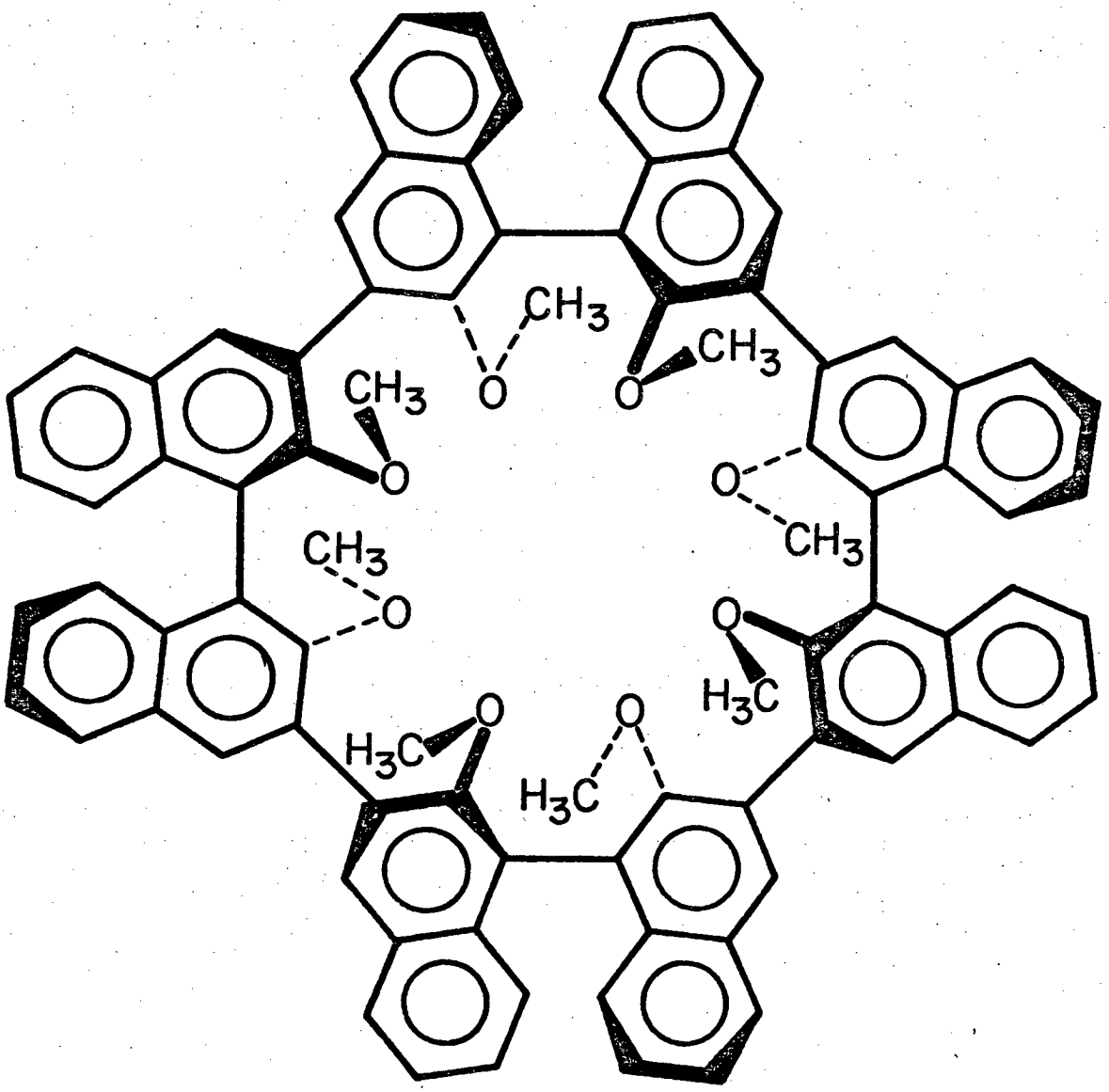

Fig. 2

TECHNICAL INFORMATION DEPARTMENT
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720