Theor Chim Acta (1986) 70: 323-332

Theoretica
Chimica Acta

The evaluation of moments for polycyclic hydrocarbons

George G. Hall

Division of Molecular Engineering, Faculty of Engineering, Kyoto University, Kyoto 606, Japan

(Received June 3, revised July 22/Accepted August 26, 1986)

The evaluation of some moments of the energy in the Hückel theory of conjugated molecules is considered. It is shown that, for molecules consisting entirely of hexagons, the moments μ_4 and μ_6 can be expressed in terms of four graphical invariants. Partial results are given for other molecules. Since the total energy can be expressed as a series of moments the implications for the energy are discussed. In this discussion two other invariants play a major role. The conclusion is suggested that an analysis of moments in terms of graphical invariants should be of prime importance in understanding these molecules.

Key words: graph theory — Graphical invariants — Moments — Hückel theory

1. Introduction

In a recent paper in this journal [1] Jiang, Tang and Hoffman have discussed the moments of the orbital energies of acyclic chains, according to Hückel theory, and given formulae for them in terms of the number of fragment patterns which can be found in the molecule. They also relate their results to the total pi electron energy. Since the total energy can be expressed exactly as an infinite series of moments [2] good approximations can be obtained using a finite number of moments.

In this paper the evaluation of the first few moments will be considered for molecules consisting entirely of hexagonal rings (polyhex) and, by way of contrast, for a few other categories. Only four graphical invariants are required to express the first six moments exactly. The accuracy of the approximation to the total energy that is obtained from these moments is also examined.

In Hückel theory the interaction matrix A summarizes the relevant information about molecular interactions. It is adjusted to have zero diagonal elements and unit off-diagonal elements between nearest neighbours. Since hexagonal-ringed molecules are alternant the carbon atoms can be numbered so that the first group (odd numbers) have no nearest neighbours among themselves and similarly the second group (even numbers). The interactions are then entirely between the groups (odd–even). If the groups are listed together (all the odds followed by all the evens) the matrix A becomes [2]

$$A=\begin{pmatrix} O & B \\ B^T & O \end{pmatrix} \tag{1}$$

where B now has all the non-zero interactions. The first moment μ_1 is the sum of all the eigenvalues of A and, by the trace theorem, is also the trace of A so here $\mu_1=0$. It is easily seen that all the odd powers of A will have the same form as A and so a zero trace. The sum of the $(2n+1)$th powers of the eigenvalues is then

$$\mu_{2n+1}=\mathrm{Tr}\,(A^{2n+1})=0. \tag{2}$$

The traces of even powers can all be expressed in terms of B:

$$\mu_2=\mathrm{Tr}\,(A^2)=2\,\mathrm{Tr}\,(B^TB)=2\,\mathrm{Tr}\,(BB^T)$$

$$\mu_4=\mathrm{Tr}\,(A^4)=2\,\mathrm{Tr}\,(B^TB)^2\ldots$$

If n is the number of C atoms and N the number of CC bonds the first two moments are simply:

$$\mu_0=\mathrm{Tr}\,(I)=n \tag{3}$$

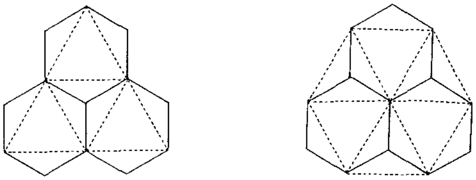
$$\mu_2=2\,(\text{number of unit elements in } B)=2N. \tag{4}$$

These results are well-known. In the next section the following moment μ_4 will be evaluated and μ_6 is found in Sect. 3. Some approximations for the total energy are described in Sect. 4 and the wider significance of the investigation is discussed in Sect. 5.

2. The evaluation of μ_4

The matrix B^TB is the second neighbour matrix and can easily be written down directly from the molecular graph [2]. Its diagonal elements are the degrees of the vertices, i.e. the number of C atoms bonded to the C in question. The off-diagonal elements are unity between second neighbours and zero otherwise. Since B involves only half as many atoms as A the number of elements is one quarter. The two second-neighbour graphs for a radical are shown in Fig. 1. The trace of the square of a symmetrical matrix, S, is the sum of the squares of every element in the matrix:

$$\mathrm{Tr}\,(S^2)=\sum_{ij} S_{ij}S_{ji}=\sum_{ij} S_{ij}^2. \tag{5}$$



B^TB BB^T

Fig. 1. The two second neighbour graphs for $C_{13}H_9$

Consequently the trace μ_4 can be deduced from this matrix $\boldsymbol{B}^T\boldsymbol{B}$ by squaring each element and adding. (It is readily proved that $\boldsymbol{B}\boldsymbol{B}^T$ gives the same result.) To do this it is necessary to examine all the different possible matrix elements in $\boldsymbol{B}^T\boldsymbol{B}$ and find how many of each there are. The largest element is 3 and this occurs only as the degree of a C at a junction between rings. Elements of 2 and 1 also occur. It is convenient, at this point, to take the hydrocarbon as a radical with u "unpaired electrons", i.e. an excess of u odd atoms over even atoms. The matrix $\boldsymbol{B}^T\boldsymbol{B}$ then refers to the $(n-u)/2$ even atoms. Since its trace is the sum of the degrees of the even vertices the number of vertices of degree 3 is $N-(n-u)$. The remaining C atoms have degree 2 and these are the only elements with 2 in the matrix. The elements which are 1 are the edges of the $\boldsymbol{B}^T\boldsymbol{B}$ graph, each counted twice, so the number of these is needed. The number of triangles in the graph is one for each hexagon ring, i.e. r, and one for each vertex of degree 3 among the odd atoms, i.e. $N-(n+u)$. Thus the number of triangles is $r+N-(n+u)$. This r is determined by n and N through the Euler relation for the molecular graph:

$$\text{rings} = \text{edges} - \text{points} + 1 \tag{6}$$

$$r = N - n + 1. \tag{7}$$

From the Euler relation (6) for this second-neighbour graph, which has $(n-u)/2$ vertices and $r+N-n-u$ rings, the number of its edges is $2N-3(n+u)/2$. The required trace is now

$$\begin{aligned}\mu_4 &= 2\,\mathrm{Tr}\,(\boldsymbol{B}^T\boldsymbol{B})^2 = 2(4(n-u)/2 + 5(N-n+u) + 2(2N-3(n+u)/2))\\ &= 18N - 12n.\end{aligned} \tag{8}$$

It is noteworthy that this is independent of u so that the formula applies equally to all polyhex whether molecules or radicals.

The result can be extended to arbitrary hydrocarbons. Two additional topological invariants are required to describe the relevant features of the hydrocarbon. The first is e, the number of "end" C atoms with only one C neighbour. The second is s, the number of "squares" or four-membered rings. The diagonal elements of $\boldsymbol{A}^2$ are the degrees of the different C atoms. The number with degree 1 is e, by definition, and the numbers with degrees 2 and 3 will be denoted by f and g, respectively. Since the total number of atoms is n

$$e + f + g = n$$

and, since the N bonds have two ends,

$$e+2f+3g=2N.$$

These equations determine f and g as

$$f=3n-2N-2e$$

$$g=2N-2n+e.$$

The off-diagonal elements of $\boldsymbol{A}^2$ are the numbers of paths of length two connecting different atoms. With the exception of squares, there is at most one such path between any two atoms and its two edges will meet at a unique centre atom at an angle of less than π. Conversely, at each atom, for each subtended angle of less than π there will be a unique path of length two connecting its neighbours. Thus, the number of paths equals the sum of the numbers of subtended angles at each atom. For an atom of degree 3 there are 3 angles, for one of degree 2 only 1 and none for those of degree 1. Thus, allowing for the elements above and below the diagonal, the sum of the off-diagonal elements is $2(3g+f)$. Squares play an exceptional role because, for the atoms across a diagonal and for these alone, there are two independent paths, by the right and by the left. For μ_4 each element of $\boldsymbol{A}^2$ is squared and the results added. The diagonal elements contribute $(e+4f+9g)$. In the absence of squares the off-diagonal elements are 1 or 0 and their squares are the same so the total is

$$\mu_4=e+4f+9g+2(3g+f). \tag{9}$$

Across each square are two diagonals each with two paths and so matrix elements of 2. When squared the latter give 2^2+2^2 instead of $1+1+1+1$ as counted in (9) so the effect is to increase the moment to

$$\mu_4=e+4f+9g+2(3g+f+4s)=18N-12n+4e+8s. \tag{10}$$

This expression for the moment holds for all hydrocarbons. Since both e and s are zero for polyhex it is a generalization of (8).

3. The evaluation of μ_6

The trace of the sixth power of $\boldsymbol{A}$ can be obtained from the third neighbour matrix $\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{B}$ by squaring every element and summing. This matrix has a form similar to that of $\boldsymbol{B}$. The number of paths of length 3 between two atoms varies from 5 to 0 and these need to be considered separately. One new graphical invariant is needed for this discussion of polyhex. This is the "bay", which is a bond on the exterior of the molecule having C atoms of degree 3 at both ends. The smallest molecule with a bay is phenanthrene. The number of bays will be called b.

The elements having 5, 4 or 3 in $\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{B}$ correspond to the bonds in $\boldsymbol{B}$. An example is given in Fig. 2. Such paths consist of a single bond traversed in both directions followed by the connecting bond. Thus the number of paths is the sum of the degrees of the two atoms less one, since traversing the connecting bond three

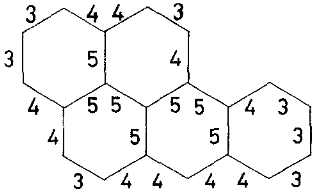


Fig. 2. The nearest neighbour elements of $\boldsymbol{BB}^T\boldsymbol{B}$. Each edge is the sum of the degrees of its vertices less 1

times should be counted only once. From this it is easily seen that all the internal CC bonds have elements of 5. The number of "internal" bonds, I, must be deduced. Each internal bond is shared between two rings so if I is added to N the shared bonds will be counted twice and each hexagon will have six edges, i.e.

$$I + N = 6r. \tag{11}$$

Furthermore the bay is an exterior bond with an element of 5. These bonds having 5 will be called "links", since they have rings at both ends, and their total number will be l where $l = I + b = 6r - N + b$.

An element of 3 occurs when both atoms of a CC bond have degree two. This may be called an exposed bond and can happen only at the "corners" of a molecule. It is readily verified that exposed bonds, for a hexagonal molecule without a bay, always number six. Each bay introduces another one. The number having 3 is then $6 + b$.

Elements with 4 occur on the remaining bonds. Since the bonds total N the number of these elements is

$$N - l - (6 + b) = 6r - 2l - 6.$$

Elements with 2 occur only across hexagons since there are two paths of length three across them. In $\boldsymbol{BB}^T\boldsymbol{B}$ there are three such paths for every ring so the total is $3r$.

Elements with 1 occur in two ways, see Fig. 3. Each bay has one across the mouth of the bay. The remainder are zig-zag paths which are best associated with their central bonds. It is quickly seen that each link, which had 5 in the previous discussion, has two of these paths and each 4 bond has one. The remaining bonds have none. The total number of the 1 elements is $b + 6(r - 1)$.

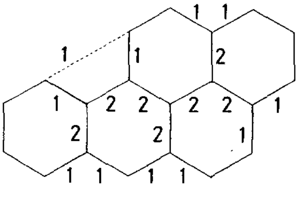


Fig. 3. The third neighbour zig-zag elements of $\boldsymbol{BB}^T\boldsymbol{B}$ listed according to their central bond. The bay element is dotted

The trace is now calculated by squaring each element and multiplying by the number of similar elements. The result is

$$\begin{aligned}\mu_6/2 &= 25l+16(6r-2l-6)+9(6+b)+4(3r)+b+6(r-1)\\ &= -7l+114r+10b-48=58N-48n+6r+3l.\end{aligned} \tag{12}$$

Because of the Euler relation between N, n and r, (10) and (12) can be put into various forms. The preferred form has no constant term so that the moment becomes a homogeneous linear combination of independent invariants and is zero for $n=0$. The formula (12) is valid for all polyhex, radicals as well as molecules.

The trace for the acyclic polyenes can be derived in a similar manner but more invariants are required. The edges with vertices of degree 1 and 3, whose number will be E, have elements of 3, and so need to be distinguished from the $(e-E)$ edges with 1 and 2 degree vertices which have 2. As well as the link, which has vertices of degree 3 at both ends and contributes 5 to the matrix, f, the number of "firths", i.e. bonds with vertices of degrees 2 and 3 and contributing 4, is needed. This can be deduced from the total number of degree 3 vertices which is (e−2) for an acyclic molecule and also $(E+f+2l)/3$ so that

$$f=3e-6-E-2l. \tag{13}$$

As before, the zig-zags are associated with their central bonds and there are $(N-e+l)$ of them. Each bay now has two other paths across its mouth and each firth has one. The trace is found by squaring and adding the elements as before:

$$\begin{aligned}\mu_6/2 &= 25l+16f+9(N-(e+f+l)+E)+4(e-E)+(N-e+l)+(f+2l)\\ &= 58N-48n+3l+18e-3E.\end{aligned} \tag{14}$$

Since, by using l as the new variable rather than b, this formula conforms to (12), for the hexagonal molecules, it can be simply generalized to

$$\mu_6/2=58N-48n+6r+3l+18e-3E. \tag{15}$$

This form now includes molecules which contain both hexagons and polyene side chains such as the polyphenyls. The generalization to four-membered rings does not conform so that they are excluded.

4. Approximations to the total energy

The total pi energy of a molecule is determined by its moments. It was proved earlier [2] that the energy could be expressed as a convergent infinite series of traces. Since the existence of such an expansion has been denied ([3], p. 53) it is worth emphasizing that the expansion will always converge. It could diverge for a radical because of the zero eigenvalue but this can be avoided by removing these eigenvalues. In the notation used above, $\boldsymbol{B}^T\boldsymbol{B}$ has $(n-u)/2$ dimensions and $\boldsymbol{B}\boldsymbol{B}^T$ has $(n+u)/2$ dimensions with u zero eigenvalues. By using $\boldsymbol{B}^T\boldsymbol{B}$ the exclusion of the zeros is automatically achieved for all benzenoid molecules. The matrix function $(\boldsymbol{B}^T\boldsymbol{B})^{1/2}$ then has as eigenvalues all the non-zero occupied orbital energies so that, with double occupation,

$$E_\pi = 2\,\mathrm{Tr}\,(\boldsymbol{B}^T\boldsymbol{B})^{1/2}.$$

A symmetrical matrix $\boldsymbol{C}$ is first defined by

$$\boldsymbol{B}^T\boldsymbol{B} = 9/2(\boldsymbol{I}-\boldsymbol{C}),\quad \boldsymbol{C} = \boldsymbol{I} - 2/9\,\boldsymbol{B}^T\boldsymbol{B} \tag{16}$$

so that $\boldsymbol{C}$ has eigenvalues strictly between -1 and 1. The total energy is, then, E_π where

$$E_\pi = 2\,\mathrm{Tr}\,(3/\sqrt{2}(\boldsymbol{I}-\boldsymbol{C})^{1/2}) = 3\sqrt{2}\,\mathrm{Tr}\,(\boldsymbol{I} - 1/2\boldsymbol{C} - 1/8\boldsymbol{C}^2 - 1/16\boldsymbol{C}^3 \cdots). \tag{17}$$

The convergence of the binomial series inside the interval -1 to 1 is well-known so the convergence of the matrix series (15) is assured. The traces of powers of $\boldsymbol{C}$ are readily expressed in terms of the traces of powers of $\boldsymbol{B}$. Although this series does converge it does so rather slowly so that other solutions are desirable. The partial sums of this series give a set of approximations to E_π which will now be evaluated for polyhex. The first term alone is $E_1 = 3/\sqrt{2}\,(n-u)$. The first two terms give

$$E_2 = 3\sqrt{2}/4(n-u) + \sqrt{2}/3N. \tag{18}$$

Since $\mathrm{Tr}\,(\boldsymbol{C}^2) = 11/54\,n - 1/2\,u + 8/81e$ the first three terms give

$$\begin{aligned} E_3 &= 97\sqrt{2}/144n - 9\sqrt{2}/16u + \sqrt{2}/3N - \sqrt{2}/27e \\ &= 0.95263n - 0.79550u + 0.47140N - 0.05238e. \end{aligned} \tag{19}$$

The trace of $\boldsymbol{C}^3$ is

$$67/486n - 1/2u + 22/729N - 16/243r - 8/243l - 16/81e + 8/243E.$$

The next approximation becomes

$$\begin{aligned} E_4 &= 6917\sqrt{2}/10368n - 15\sqrt{2}/32u + 637\sqrt{2}/1944N + \sqrt{2}/81r \\ &\quad + \sqrt{2}/162l - \sqrt{2}/162E + 0e \\ &= 0.91607n - 0.66291u + 0.46340N + 0.01746r + 0.00873l - 0.00873E. \end{aligned} \tag{20}$$

These approximations are the beginning of the convergent expansion but its convergence is fairly slow. The first term alone is a very bad approximation and the second is only a little better. E_3 begins to represent the shape of the energy more realistically and E_4 is an improvement at the expense of using three extra variables.

Another approach to the problem considers the approximation to $|x|$ which can be obtained using polynomials in x^2. A least squares fit of Legendre polynomials in the variable x over $0 < x < 1$ gives

$$x = 1/2 + 5/8P_2(x) - 3/16P_4(x) + 13/128P_6(x). \tag{21}$$

The largest difference between the two sides of this equation when only four terms are included is at $x = 0$ where it is 0.0854. If x is replaced by the matrix

$A/3$ then this polynomial gives an expression for E_π in terms of the traces up to the sixth power. The maximum error of each eigenvalue of A is 0.0854 so the error in E_π is bounded by $\pm 0.0854n$. The resulting expansion for E_π is

$$E_\pi = 0.93032n - 0.25634u + 0.35817N + 0.07242r + 0.03621l$$
$$-0.03621E - 0.02632e. \tag{22}$$

Although this does not agree well with (20) it has the same qualitative features.

A brief consideration of the higher moments shows that an increasing number of graphical invariants are needed to express them exactly but that those already introduced give the major contributions. This suggests an alternative approach to the problem of representing the energy. This is to fit an energy expression consisting of a linear combination of a limited number of invariants to the calculated energies of a large number of molecules. A least squares fitting of this kind was performed earlier [4] and gave, for the polyhex, using only n and N

$$E_\pi = 0.909n + 0.419N. \tag{23}$$

For the polyenes the three variables n, N, e are needed and an approximate fit gave [5]

$$E_\pi = n + N/4 - e/8. \tag{24}$$

For radicals u is also needed and the result was [5]

$$E_\pi = 0.9n + 0.43N - 0.17e - 0.44u. \tag{25}$$

These show rough agreement with the expansions above.

In these formulae the energy is expressed as a linear combination of invariants. This is suggested, but not proved, by the fact that the lower moments can be so expressed. Other types of formula have been used with some success. McClelland [6] has used the form

$$E_\pi = 0.92\surd(2nN) \tag{26}$$

and others have elaborated this. For references see Chap. 3 in Graovac et al. [3]. To express the linear dependence of the energy of isomers on the (algebraic) number of resonance structures K [4], the formula

$$E_\pi = 0.442n + 0.788N + 0.34K\ (0.632)^{N-n} \tag{27}$$

was found [7] by fitting a large number of polyhex molecules. A similar formula has been given recently by Gutman [8].

5. Discussion

The fact that the energy is largely determined by the two quantities n and N is important for these molecules. The quantity N can be replaced by h, the number of H atoms, using

$$h = 3n - 2N, \tag{28}$$

so that, for example, (23) becomes

$$E_{\pi}=1.5375n-0.210h. \tag{29}$$

This relates the energy directly to the formula C_nH_h and helps to explain the two-dimensional vector addition relation of molecular energies pointed out by Dias [9] and expressible as

$$E(C_{n+m}H_{h+j})=E(C_nH_h)+E(C_mH_j). \tag{30}$$

This paper shows that the study of the moments is a more controlled means of identifying and classifying the graphical invariants than a direct attack on the energy. The moment μ_{2r} counts the number of closed paths of length $2r$ on the graph and, as has been shown, this reduces to the sum of the squares of the number of open paths of length r. A graphical feature which influences this cannot involve more than r bonds. Thus for the moments up to μ_4, only n and N are essential with e and s in appropriate situations. The extra step, to μ_6, brings in r, E and l, whose definitions involve three bonds, but these describe only polyhex and polyenes. Preliminary evidence suggests that, even for polyhex, μ_8 requires three more invariants. In general the polyhex require fewer invariants than other molecules because of the tighter constraints on their structure starting with the fact that their C atoms cannot have degree one. The invariant u is of special interest. It is not required for any of the lower moments. It arises naturally when the energy is discussed in terms of $\boldsymbol{B}$ rather than $\boldsymbol{A}$ so that (3) is not used. Since there are $(n-u)/2$ linearly independent moments it measures a property of the set of moments as a whole and does not arise as a fragment pattern. Similarly, K is not required for the moments but, since $K=|\boldsymbol{B}|$ for a polyhex molecule, it is another useful graphical invariant for the whole molecule.

There is a connection between the study of the coefficients a_r of the characteristic polynomial and of the moments. These coefficients are defined by the identity

$$\sum_r a_r x^{n-r}=|x\boldsymbol{I}-\boldsymbol{A}| \tag{31}$$

and can all be expressed in terms of moments. Thus the a_4 coefficient, which is the sum of the eigenvalues of $\boldsymbol{A}$ taken four at a time and all different, is given, for an alternant, by

$$a_4=1/8(\mu_2)^2-1/4\mu_4. \tag{32}$$

This gives for a_4 using (4) and (10) the same expression as that obtained by Dias [10]. Similarly a_6 is given by

$$a_6=-1/48(\mu_2)^3+1/8\mu_2\mu_4-1/6\mu_6 \tag{33}$$

and this also agrees with that of Dias. In general the linear term in a_r is a multiple of μ_r so that the same graphical invariants are involved but since powers of other moments are also included it is less convenient to study the a_r than the separate moments.

When the molecule becomes large or even infinite its eigenvalues tend to form bands and new techniques are needed to discuss its electronic structure. It has been shown by Burdett et al. [11] that the energy moments remain important and, using a continued fraction technique, can give the density of states and other useful properties.

It has been convenient to base this discussion on Hückel theory but the analysis of the structure of a molecule in graph theoretical terms is of prime importance in any theory of these molecules so that its significance is much wider.

Acknowledgement. The author wishes to thank Professor J R Dias for communicating to him the results in references 9 and 10 in advance of publication and for his comments on this paper.

References

1. Jiang Y, Tang A, Hoffmann R (1984) Theor Chim Acta 66:183
2. Hall GG (1955) Proc Roy Soc 229:251
3. Graovac A, Gutman I, Trinajstic N (1977) Topological approach to the chemistry of conjugated molecules. Springer, Berlin Heidelberg New York
4. Hall GG (1973) Int J Math Ed Sc Tech 4:233
5. Hall GG (1976) In: Andrews JG, McLone RR (eds) Mathematical modelling. Butterworth, London
6. McClelland BJ (1971) J Chem Phys 54:640
7. Hall GG (1981) Bull Inst Math Applic 17:70
8. Gutman I (1984) Theor Chim Acta 65:23
9. Dias JR (1986) J Mol Str (Theochem) 137:9
10. Dias JR (1985) Theor Chim Acta 68:107
11. Burdett JK, Lee S, Sha WC (1984) Croat Chem Acta 57:1193