

APPROXIMATING THE TOTAL  $\pi$ -ELECTRON ENERGY OF BENZENOID HYDROCARBONS:A RECORD ACCURATE FORMULA OF  $(n,m)$ -TYPE

Ivan GUTMAN

*Faculty of Science, University of Kragujevac, P.O.Box 60,  
YU-34000 Kragujevac, Yugoslavia*

(Received: September 1992)

**Abstract:** A novel  $(n,m)$ -type lower bound and an approximate topological formula for the total  $\pi$ -electron energy ( $E$ ) of benzenoid hydrocarbons are put forward, improving some recent results by Lemi Türker;  $n$  = number of carbon atoms,  $m$  = number of carbon-carbon bonds. The new approximation for  $E$  is shown to be more accurate than any of the 48 previously considered formulas of  $(n,m)$ -type.

## INTRODUCTION

The total  $\pi$ -electron energy ( $E$ ) of a benzenoid hydrocarbon depends on a variety of structural features and topological (i.e. graph-theoretical) invariants of the respective molecule, but the far most important are the number of carbon atoms ( $n$ ) and the number of carbon-carbon bonds ( $m$ ). This fact is nowadays firmly established and supported by numerous theoretical arguments and extensive computer work [1]. As a matter of fact, about 99.5% of  $E$  is determined by the invariants  $n$  and  $m$ .

A plethora of mathematical expressions, depending solely on  $n$  and  $m$ , has been proposed for the approximate calculation of the total  $\pi$ -electron energy. If  $E^*$  is such a function of  $n$  and  $m$  then eqs. (1) and (2):

$$E = a_1 E^* \quad (1)$$

$$E = a_2 E^* + b_2 \quad (2)$$

are called  $(n,m)$ -type approximate formulas for total  $\pi$ -electron energy. The constants  $a_1$ ,  $a_2$  and  $b_2$  are usually determined by means of least-squares fitting, using the exact  $E$ -values of a set of pertinently selected benzenoid systems. In [1] 24 different expressions  $E^*$ , previously introduced in the chemical literature, have been collected. The accuracy of the resulting 48 formulas of  $(n,m)$ -type was compared using a data base that consists of the  $E$ -values of 104 Kekuléan benzenoid hydrocarbons with three or more condensed six-membered rings, from the book of Zahradník and Pancir [2]. Throughout this work we employ the same data base and thus the results reported here are directly comparable with those from [1].

The most accurate among the known  $(n,m)$ -type formulas for  $E$  have an average error of 0.33% eq. (1) and 0.30% eq. (2) and a correlation coefficient of 0.9998 [1]. In this paper we report some novel  $(n,m)$ -type expressions whose precision is slightly better than of any approximation examined so far. Our results are closely related to and based on certain recent findings by Lemi Türker [3,4].

# TÜRKER'S UPPER AND LOWER BOUNDS FOR TOTAL $\pi$ -ELECTRON ENERGY

In a paper [3] published few years ago Türker deduced the upper bound (3) for the total  $\pi$ -electron energy of an arbitrary alternant hydrocarbon:

$$E \leq E_U = 2 \left[ m + 2 \left[ \binom{\nu}{2} a_4 \right]^{1/2} \right]^{1/2} \quad (3)$$

where  $\nu = n/2$  and where  $a_4$  is the fourth coefficient of the characteristic polynomial. Recently Türker arrived at a lower bound of analogous form [4]:

$$E \geq E_L = 2 \left[ m + 2 [a_4]^{1/2} \right]^{1/2} \quad (4)$$

In eq. (3) as well as throughout this paper it is assumed that  $n$  is an even number. The case of odd  $n$  (which is chemically much less relevant) can be treated in a fully analogous manner, bearing in mind that then the graph eigenvalue  $x_{(n+1)/2}$  (see below) is necessarily equal to zero.

We provide here an elementary derivation of (3) and (4), which differs from the reasoning used in [3,4]. Its slight modification (described in the subsequent section) will result in improvements of (4).

We start with the relation (5) which holds for all alternant hydrocarbons [1]:

$$E = 2 \sum_{i=1}^{\nu} x_i \quad (5)$$

where  $x_1, x_2, \dots, x_n$  are the eigenvalues of the respective molecular graph ordered so that  $x_i \leq x_j$  for  $i > j$ . From (5) it immediately follows

$$\frac{1}{4} E^2 = \left( \sum_{i=1}^{\nu} x_i \right)^2 = \sum_{i=1}^{\nu} (x_i)^2 + 2 \sum_{i < j} x_i x_j$$

i.e.

$$\frac{1}{4} E^2 = m + 2 \sum_{i < j} x_i x_j \quad (6)$$

because of the well known identity

$$\sum_{i=1}^{\nu} (x_i)^2 = m.$$

Since  $\left(\frac{\nu}{2}\right)^{-1} \sum_{i < j} x_i x_j$  is the arithmetic mean of the products  $x_i x_j$

one has

$$\begin{aligned} \left(\frac{\nu}{2}\right)^{-1} \sum_{i < j} x_i x_j &\leq \left[ \left(\frac{\nu}{2}\right)^{-1} \sum_{i < j} (x_i)^2 (x_j)^2 \right]^{1/2} \\ &= \left[ \left(\frac{\nu}{2}\right)^{-1} a_4 \right]^{1/2}. \end{aligned} \quad (7)$$

Substituting (7) back into (6) we arrive at

$$\frac{1}{4} E^2 \leq m + 2 \left(\frac{\nu}{2}\right) \left[ \left(\frac{\nu}{2}\right)^{-1} a_4 \right]^{1/2}$$

from which (3) follows straightforwardly.

In order to deduce (4) start with the identity

$$\begin{aligned} \sum_{i < j} x_i x_j &= \left[ \sum_{i < j} (x_i)^2 (x_j)^2 + 2 \sum_{i, j, k, l} x_i x_j x_k x_l \right]^{1/2} \\ &= \left[ a_4 + 2 \sum_{i, j, k, l} x_i x_j x_k x_l \right]^{1/2}. \end{aligned} \quad (8)$$

Bearing in mind that the summands  $x_i x_j x_k x_l$  are non-negative numbers,

it is clear that

$$\sum_{i < j} x_i x_j \geq [a_4]^{1/2}$$

which substituted into (6) renders (4).

AN IMPROVEMENT OF TÜRKER'S LOWER BOUND

Instead of completely neglecting the term  $\sum_{i,j,k,l} x_i x_j x_k x_l$  in eq. (8) which leads to Türker's lower bound (3) we try to decrease the right-hand side of (8) in a less severe manner. Observe first that  $\sum_{i,j,k,l} x_i x_j x_k x_l$  consists of  $\nu(\nu-1)(\nu^2-\nu-2)/8$  summands. Consequently,

$$\begin{aligned} & \left[ \nu(\nu-1)(\nu^2-\nu-2)/8 \right]^{-1} \sum_{i,j,k,l} x_i x_j x_k x_l \\ & \geq \left[ \prod_{i,j,k,l} x_i x_j x_k x_l \right] \left[ \nu(\nu-1)(\nu^2-\nu-2)/8 \right]^{-1} \end{aligned} \quad (9)$$

where we use the fact that the geometric mean of non-negative numbers cannot exceed their arithmetic mean. Every eigenvalue in the product  $\prod_{i,j,k,l} x_i x_j x_k x_l$  occurs  $(\nu-1)(\nu^2-\nu-2)/2$  times. Hence the right-hand side of (9) is equal to

$$\begin{aligned} & \left[ \prod_{i=1}^{\nu} x_i \right]^{[(\nu-1)(\nu^2-\nu-2)/2]/[\nu(\nu-1)(\nu^2-\nu-2)/8]} \\ & = \left[ \prod_{i=1}^{\nu} x_i \right]^{4/\nu} = |\det A|^{2/\nu} \end{aligned} \quad (10)$$

where  $\det A = \prod_{i=1}^n x_i$  is the determinant of the adjacency matrix. Combining (8), (9) and (10) we obtain

$$\sum_{i < j} x_i x_j \geq \left[ a_4 + \frac{1}{4} \nu(\nu-1)(\nu^2-\nu-2) |\det A|^{2/\nu} \right]^{1/2}$$

which substituted back into (6) yields

$$E \geq E_{La} = 2 \left[ m + 2 \left[ a_4 + \frac{1}{4} \nu(\nu-1)(\nu^2-\nu-2) |\det A|^{2/\nu} \right]^{1/2} \right]. \quad (11)$$

If we restrict the consideration to the (chemically most interesting) case when  $\det A \neq 0$ , then because of  $|\det A| \geq 1$  we can simplify (11) as

$$E \geq E_{Lb} = 2 \left[ m + 2 \left[ a_4 + \frac{1}{4} \nu(\nu-1)(\nu^2-\nu-2) \right]^{1/2} \right]^{1/2}. \quad (12)$$

The inequalities (11) and (12) are, evidently, improvements of the Türker's lower bound (4).

For benzenoid hydrocarbons the estimates  $E_U$ ,  $E_L$ ,  $E_{La}$  and  $E_{Lb}$  can be further simplified by using the relations [5]

$$a_4 = \frac{1}{2} (m^2 - 9m + 6n)$$

$$|\det A| = K^2$$

where  $K$  is the Kekulé structure count. Thus we obtain:

$$E_U = \left[ 4m + \left[ 4n(n-2)(m^2 - 9m + 6n) \right]^{1/2} \right]^{1/2} \quad (13)$$

$$E_L = \left[ 4m + \left[ 32(m^2 - 9m + 6n) \right]^{1/2} \right]^{1/2} \quad (14)$$

$$E_{La} = \left[ 4m + \left[ 32(m^2 - 9m + 6n) + n(n-2)(n^2 - 2n - 8) K^{8/n} \right]^{1/2} \right]^{1/2}$$

$$E_{Lb} = \left[ 4m + \left[ 32(m^2 - 9m + 6n) + n(n-2)(n^2 - 2n - 8) \right]^{1/2} \right]^{1/2}. \quad (15)$$

Observe that for benzenoid hydrocarbons  $E_U$ ,  $E_L$  and  $E_{Lb}$  are expressions of  $(n,m)$ -type. The function  $E_{La}$  is applicable only to Kekuléan benzenoid species.

# TÜRKER'S APPROXIMATE FORMULA FOR TOTAL $\pi$ -ELECTRON ENERGY

In [4] Türker proposed an approximate expression for total  $\pi$ -electron energy (of alternant hydrocarbons) of the form

$$E_T = \alpha E_L + (1 - \alpha) E_U \quad (16)$$

For benzenoids  $E_T$  is of (n,m)-type.

Türker himself determined the value of the coefficient  $\alpha$  by means of theoretical arguments [4] and found that  $\alpha = 1/8 = 0.125$ . We optimized  $\alpha$  numerically so as to gain a minimal average relative error of eq. (1) and found a remarkably close value of  $\alpha = 0.1197$ .

In full analogy to Türker's approximation (16) we may consider

$$E_{Tb} = \alpha E_{Lb} + (1 - \alpha) E_U \quad (17)$$

which for Kekuléan benzenoid hydrocarbons is an (n,m)-type expression. Numerical optimization gave  $\alpha = 0.306$ .

## TESTING THE NOVEL (n,m)-TYPE APPROXIMATE FORMULAS

The approximate formulas (1) and (2) were tested on our standard data base [1] for the following five novel (n,m)-type expressions  $E^*$ :  $E_L$ ,  $E_{Lb}$ ,  $E_T(\alpha = 1/8)$ ,  $E_T(\alpha = 0.1197)$  and  $E_{Tb}(\alpha = 0.306)$ . The results obtained are collected in Table 1. For completeness also the results for the (previously examined [1]) expression  $E_U$  are included in Table 1.

An inspection of Table 1 reveals that both Türker's lower bound (14) and its present improvement (15) are fully inapplicable for purposes of approximating the total  $\pi$ -electron energy. On the other hand, the linear combination of the upper and lower bounds, especially (16), significantly increases the precision of the respective approximate to-

Table 1a. Coefficients in the approximate (n,m)-formulas (1) & (2)

Equation for E <sup>*</sup>	a <sub>1</sub>	a <sub>2</sub>	b <sub>2</sub>
(13)	0.919	0.899	0.90
(14)	2.224	3.949	-30.23
(15)	1.333	1.434	-3.01
(16), $\alpha=1/8$	0.993	0.995	-0.09
(16), $\alpha=0.1197$	0.989	0.990	-0.04
(17), $\alpha=0.306$	1.016	1.015	0.05

Table 1b. Results of numerical testing of approximate formulas (1) & (2)

Equation for E*	Eq. (1)		Eq. (2)		correlation coefficient
	mean error (%)	max. error observ. (%)	mean error (%)	max. error observ. (%)	
(13)	0.54	2.3	0.30	1.0	0.9998
(14)	7.69	37.8	1.34	12.4	0.996
(15)	1.41	8.2	0.48	2.1	0.9994
(16), $\alpha=1/8$	0.30 <sup>a</sup>	1.0	0.30 <sup>c</sup>	1.0	0.9998
(16), $\alpha=0.1197$	0.30 <sup>b</sup>	1.0	0.30 <sup>d</sup>	1.0	0.9998
(17), $\alpha=0.306$	0.31	1.2	0.31	1.2	0.9998

Unrounded mean errors of (16): a = 0.2971 c = 0.2996  
b = 0.2960 d = 0.2995

polological formula. As a matter of fact, eq. (16) provides the (n,m)-type expression for E that has the highest accuracy ever observed. Although there is no significant difference between Türker's approximation ( $\alpha = 1/8$ ) and ours ( $\alpha = 0.1197$ ), the latter has a slightly smaller mean error and is thus record accurate.

It is also noteworthy that, contrary to (13), (14) and (15), the expressions (16) and (17) render a very small (near-zero)  $b_2$ -value. This indicates that there may be a deeper reason for the success of (16) and (17), which we don't fully understand at the present moment.



REFERENCES

- [1] For review and an exhaustive bibliography on research of the total  $\pi$ -electron energy of benzenoid hydrocarbons and, in particular, its (n,m)-dependence see: I.Gutman, Topics Curr.Chem. **162**, 29 (1992).
- [2] R.Zahradnik and J.Pancir, HMO Energy Characteristics, Plenum Press, New York, 1970.
- [3] L.Türker, Match **16**, 83 (1984).
- [4] L.Türker, Match **28**, 261 (1992).
- [5] I.Gutman, Topics Curr.Chem. **162**, 1 (1992).