Numerical Bounds for the Perfect Matching Vectors of a Polyhex

Pierre Hansen* and Maolin Zheng

GERAD, École des Hautes Études Commerciales, 5255 Avenue Decelles, Montréal, Canada H3T 1V6

Received May 18, 1993®

Consider a polyhex H which admits a perfect matching M. Harary, Klein and Zivkovic define the perfect matching vector of M as (n_3, n_2, n_1, n_0) where n_j is the number of hexagons of H containing exactly j edges of M (j = 3, 2, 1, 0). Sharp lower and upper bounds on n_3 , n_2 , n_1 , and n_0 for a given H and all M are obtained by mixed-integer programming. Computational experience with polyhexes having over 100 hexagons is reported.

1. INTRODUCTION

Consider regular hexagons with sides of unit length. A polyhex is obtained by merging these hexagons in such a way that two hexagons are either disjoint or have exactly one edge in common and the resulting graph is connected. Polyhexes play an important role in theoretical chemistry (e.g., Clar, 1-3 Gutman and Cyvin, 4,5 Dias, 6,7 Trinajstić, 8 and Rouvray and Balaban^{9,10}) due to the hexagonal shape of benzene, mergings of which form benzenoids. Vertices of polyhexes represent carbon atoms and edges single or double bonds between them. Polyhexes which are geometrically planar and simply connected (i.e., without interior faces which are not hexagons) are benzenoid systems. 4,5 Planar multiply connected polyhexes are coronoid systems (or multiple coronoids if they have several internal faces which are not hexagons). Note that we do not require here that polyhexes be planar; this allows us in particular to consider helicenes, in which several hexagons may be superposed. A polyhex without internal vertices (i.e., vertices belonging to three hexagons) is catacondensed. Otherwise, it is pericondensed.

A natural question to ask given a polyhex H is "Does Hcorrespond to a molecule?" Empirical evidence^{4,5} suggests that a necessary condition for this to be the case is that H possess a perfect matching M (or Kekulé structure: in other words a set of disjoint edges which cover all vertices). Many specialized graph theoretic algorithms have been devised to find if a benzenoid system H has a Kekulé structure or not (e.g., Sachs, 11 John and Sachs, 12 Sheng, 13,14 He and He15 and Hansen and Zheng^{16,17}). Recently, one such algorithm with a complexity linear in the number N of vertices of H (or in the size of the input) has been proposed.¹⁷ Some of these methods extend to the cases of coronoids or general polyhexes. Moreover, it is also possible to use graph theoretic algorithms of operations research, as, e.g., the perfect matching algorithm for bipartite graphs of Hopcroft and Karp. 18 This algorithm has a time complexity in $O(N^{3/2})$ for polyhexes, as their number of edges is proportional to their number of vertices. Finally, one could consider this problem as one of mathematical programming, i.e., optimization of an objective function subject to constraints, and use general methods from that field (note that while mathematical programming algorithms, designed to solve problems called mathematical programs, are implemented by computer programs, the term programming does not refer here to computers). The relevant tool is then linear programming (which addresses problems with linear objective function and constraints) in 0-1 variables. Let N(i) denote the set of neighbors of vertex i of H. Associate a variable x_{ij} with the edge of H joining vertices i and j, for all adjacent i

and j. Then H has a perfect matching if and only if the following set of equations:

$$\sum_{i \in N(i)} x_{ij} = 1 \qquad \forall i \in H \tag{1}$$

has a solution in 0-1 variables. Solving (1) by a general method is much less efficient than using one of the algorithms listed above. However, the system (1) may be useful if one wishes to discriminate between Kekulé structures, according to some criterion. One such criterion is the number of mutually resonant hexagons induced by the perfect matching (or, in other words, the number of disjoint hexagons containing three double bonds). It was proposed in ref 19 to call the maximum number of mutually resonant disjoint hexagons induced by a perfect matching in a polyhex H its Clar number. Finding the Clar number of catacondensed polyhexes is discussed in refs 19 and 20; the Clar number of pericondensed benzenoid systems can be efficiently found by mixed-integer programming²¹ (and, in fact, apparently by linear programming alone, as for all known examples the solution of the linear relaxation, obtained by replacing the constraints $x_{ij} \in \{0,1\}$ by $x_{ij} \in$ [0,1], is optimal).

In this paper we consider another way to discriminate among the perfect matchings of H. Harary et al. 22 define the perfect matching vector (n_3, n_2, n_1, n_0) of a perfect matching M of H to be such that n_i is equal to the number of hexagons of H containing exactly j edges of M (it appears that n_3 , n_2 , n_1 , and no were already mentioned by Sahini²³ and Sahini and Savin²⁴). Then they observe that "natural invariants of H are the maximum value of n_i , the minimum value of n_i , and the average value of n_j , j = 3, 2, 1, 0". Clearly these maximum, minimum, and average values of the n_i can be found by enumeration of all perfect matchings of H and examination of all hexagons for each of them. Such an approach is rapidly prohibitive when the size of the polyhex H (i.e., its number of hexagons) augments. The number of Kekulé structures of a benzenoid system is discussed in detail in ref 5. Very recently, catacondensed polyhexes with a given number of hexagons and a maximum number of Kekulé structures have been determined (Balaban, 25 Balaban et al., 26 Hansen and Zheng 27). This number has 23 digits when there are 100 hexagons. A more efficient way to bound the n_i for j = 3, 2, 1, 0 is to use mixed-integer programming.^{21,28} Mathematical programs to do so are given in the next section. Computational results are reported in the last section.

2. MATHEMATICAL PROGRAMS

To state the mixed-integer programs (i.e., programs with some integer and some continuous variables) for finding the

[•] Abstract published in Advance ACS Abstracts, February 15, 1994.

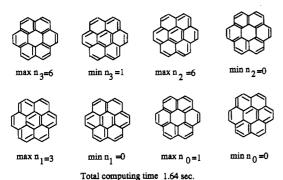


Figure 1. Illustration for the eight mathematical programs.

maximum and minimum of n_j for j = 3, 2, 1, 0, we introduce variables y_s and z_s associated with each hexagon s of H as well as 0-1 variables $x_{s1}, x_{s2}, ..., x_{s6}$ associated with the consecutive edges of s, beginning at the highest vertex (all hexagons are assumed to have the second and fourth edges vertical). Variables associated with edges belonging to two neighboring hexagons s and s' are of course the same (e.g., if s and s' are at the same level and s' to the right of s, $x_{s2} = x_{s'5}$; if s' is above and to the left of s, $x_{s6} = x_{s3}$). For ease of exposition we use the notation for a hexagon s only. Fries²⁹ has argued that among all Kekulé structures those with maximum n_3 are chemically the most significant. We therefore propose to call the maximum of n_3 the Fries number of H (note that n_3 is called the conjugated six-circuit count of the corresponding matching or of all perfect matchings of $H^{30,31}$). Let M denote the set of all perfect matchings of H.

Problem 1. Fries number: $MAX n_3$

maximize
$$z = \sum_{s} (y_s + z_s)$$

subject to (1) and

$$y_{s} \leq x_{s1} \qquad z_{s} \leq x_{s2}$$

$$y_{s} \leq x_{s3} \qquad z_{s} \leq x_{s4} \qquad \forall s$$

$$y_{s} \leq x_{s5} \qquad z_{s} \leq x_{s6} \qquad (2)$$

where $y_s, z_s \ge 0, x_{s1}, ..., x_{s6} \in \{0, 1\}$. The hexagon s will contain three double bonds if and only if either y_s or z_s is equal to 1; both cannot be equal to 1 at the same time. Constraints (1) (with adequate correspondence for the x variables) impose that the x equal to 1 define a perfect matching M of H; if hexagon s contains three edges of M, they are either the first, third, and fifth ones (i.e., in proper position³²) or the second, fourth, and sixth ones (i.e., in improper position³²). In the first case the continuous variable y_s , which has a positive coefficient in the objective function, will take the value 1; in the second case variable z_s will take the value 1. If less than three edges of s belong to M at least one of x_{s1} , x_{s3} , x_{s5} and at least one of x_{s2} , x_{s4} , x_{s6} will be equal to 0 and both y_s and z_s will be forced down to 0 by constraints (2).

Problem 2. MIN n_3

$$\begin{array}{l}
\text{MIN } h_3 \\
\text{M \in M} \\
\text{minimize } h - \sum_s y_s
\end{array}$$

subject to (1) and

$$y_s \le 3 - x_{s1} - x_{s2} - x_{s3} - x_{s4} - x_{s5} - x_{s6}$$
 $\forall s$

$$y_s, x_{s1}, ..., x_{s6} \in \{0,1\} \quad \forall s$$
 where h denotes the number of hexagons of H.

The hexagon s will contain less than three edges of M if and only if y_s is equal to 1. Indeed, constraints (3) force y_s down to 0 if s contains three edges of M and are redundant otherwise. The objective function forces y_s to its largest value. Maximizing $n_2 + n_1 + n_0$ is equivalent to minimizing n_3 .

Problem 3. MAX n₂

maximize
$$z = \sum_{s} y_{s}$$

subject to (1), (3), and

$$y_{s} \leq x_{s1} + x_{s2} + x_{s3} + x_{s4} + x_{s5}$$

$$y_{s} \leq x_{s2} + x_{s3} + x_{s4} + x_{s5} + x_{s6}$$

$$y_{s} \leq x_{s3} + x_{s4} + x_{s5} + x_{s6} + x_{s1}$$

$$y_{s} \leq x_{s4} + x_{s5} + x_{s6} + x_{s1} + x_{s2}$$

$$y_{s} \leq x_{s5} + x_{s6} + x_{s1} + x_{s2} + x_{s3}$$

$$y_{s} \leq x_{s6} + x_{s1} + x_{s2} + x_{s3} + x_{s4}$$
(4)

where $0 \le y_s \le 1$; x_{s1} , x_{s2} , ..., $x_{s6} \in \{0,1\} \ \forall s$. The variables y_s are equal to 1 if and only if hexagon s contains two edges of M. Indeed, if two of $x_{s1}, ..., x_{s6}$ are equal to 1, the right-hand sides of (3) are equal to 1 and the right-hand sides of (4) to 2 except in two cases in which they are equal to 1. If s contains three edges of M (3) forces y_s down to 0; if s contains one edge of M one of the constraints (4) forces y_s down to 0 and if s contains no edge of M all constraints (4) do so.

Problem 4. MIN n_2

$$\begin{array}{ll}
M \in \mathcal{M} \\
\text{minimize } z = h - \sum_{s} y_{s}
\end{array}$$

subject to (1) and

$$y_{s} \le 2 - x_{s1} - x_{s2} - x_{s4} - x_{s5} + x_{s3} + x_{s6}$$

$$y_{s} \le 2 - x_{s2} - x_{s3} - x_{s5} - x_{s6} + x_{s1} + x_{s4} \quad \forall s$$

$$y_{s} \le 2 - x_{s3} - x_{s4} - x_{s6} - x_{s1} + x_{s2} + x_{s5}$$
(5)

where $0 \le y_s \le 1$; $x_{s1}, x_{s2}, ..., x_{s6} \in \{0,1\} \ \forall s$. The hexagon s will contain two edges of M if and only if $y_s = 0$. Indeed the right-hand sides of (5) are greater than or equal to 1, unless s contains exactly two edges of M; then for one of them two variables x with negative coefficient are equal to 1 and none with positive coefficient. This right-hand side being equal to 0, it forces y_s down to 0.

Problem 5. MAX n_1

maximize
$$z = \sum_{s} y_{s}$$

subject to (1), (3), and

(3)

$$y_{s} \le 2 - x_{s1} - x_{s2} - x_{s4} - x_{s5}$$

$$y_{s} \le 2 - x_{s2} - x_{s3} - x_{s5} - x_{s6} \quad \forall s$$

$$y_{s} \le 2 - x_{s3} - x_{s4} - x_{s6} - x_{s1}$$
(6)

$$y_s \le x_{s1} + x_{s2} + x_{s3} + x_{s4} + x_{s5} + x_{s6} \quad \forall s$$
 (7)

where $y_s \ge 0$; x_{s1} , x_{s2} , ..., $x_{s6} \in \{0,1\} \ \forall s$.

The hexagon s will contain one edge of M if and only if y_s = 1. Indeed, if exactly one of the $x_{s1}, ..., x_{s6}$ is equal to 1, the

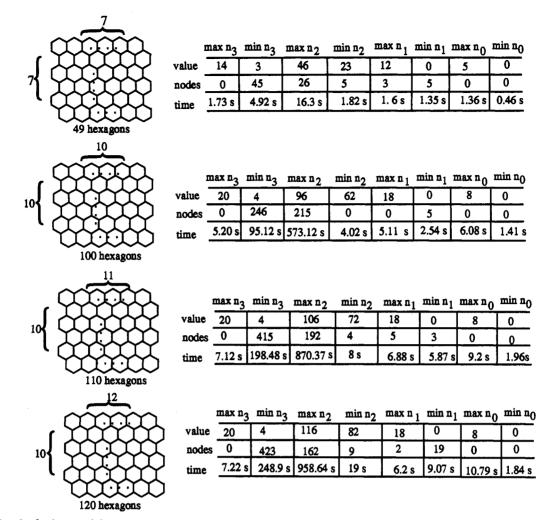


Figure 2. Results for large polyhexes.

right-hand sides of constraints (3) are equal to 2, those of constraints (6) to 1 or 2 and that of constraint (7) to 1. If s contains three edges of M, (3) forces y_s down to 0; if s contains two edges of M (which cannot be adjacent), one of the constraints (6) forces y_s down to 0. Indeed, the sets of variables in these constraints correspond to the three sets of four edges adjacent to opposite vertices of s. Any pair of edges of s belonging to M must be in one such set. Finally, if s contains no edge of M, constraint (7) forces y_s down to

Problem 6. MIN
$$n_1$$

minimize
$$\sum_{s} y_{s}$$

subject to (1) and

$$y_{s} \ge x_{s1} - x_{s3} - x_{s4} - x_{s5}$$

$$y_{s} \ge x_{s2} - x_{s4} - x_{s5} - x_{s6}$$

$$y_{s} \ge x_{s3} - x_{s5} - x_{s6} - x_{s1} \quad \forall s$$

$$y_{s} \ge x_{s4} - x_{s6} - x_{s1} - x_{s2}$$

$$y_{s} \ge x_{s5} - x_{s1} - x_{s2} - x_{s3}$$

$$y_{s} \ge x_{s6} - x_{s2} - x_{s3} - x_{s4}$$
(8)

where $y_s \ge 0$; x_{s1} , x_{s2} , ..., $x_{s6} \in \{0,1\} \ \forall s$.

The hexagon s will contain one edge of M if and only if y_s is equal to 1. Indeed, if s contains exactly one edge of M, the right-hand side of one of the constraints (8) will be equal to 1 and that of the others will be equal to 0 or -1; ν_s will be forced to take the value 1. If s contains none or more than one edge of M, the right-hand sides of the constraints (8) will be nonpositive and y_3 will take the value 0.

Problem 7. MAX n_0 $M \in \mathcal{M}$ maximize $\sum y_s$

subject to (1) and

$$y_{s} \le 1 - x_{s1} - x_{s2} \qquad y_{s} \le 1 - x_{s4} - x_{s5}$$

$$y_{s} \le 1 - x_{s2} - x_{s3} \qquad y_{s} \le 1 - x_{s5} - x_{s6} \qquad \forall s$$

$$y_{s} \le 1 - x_{s3} - x_{s4} \qquad y_{s} \le 1 - x_{s6} - x_{s1}$$
(9)

where $y_s \ge 0$; x_{s1} , x_{s2} , ..., $x_{s6} \in \{0,1\} \ \forall s$. The hexagon s will contain no edge of M if and only if y_s is equal to 1. Indeed, if s contains no edge of M the righthand sides of constraints (9) are all equal to 1 and y_s will take the value 1; if s contains at least one edge of M the right-hand side of at least two constraints (9) will be equal to 0 forcing y_s down to 0.

Problem 8. MINno minimize $h - \sum y_s$

subject to (1), (7), and

$$0 \le y_s \le 1; x_{s1}, x_{s2}, ..., x_{s6} \in \{0,1\} \ \forall s$$

The hexagon s will contain no edge of M if and only if $y_s =$ 0. Indeed minimizing n_0 amounts to maximizing $n_1 + n_2 +$ n_3 . As mentioned above, constraints (7) force y_s down to 0 if and only if s contains no edge of M. Otherwise, due to the upper bounds, y_s will take the value 1.

Observe that the mathematical programs given for problems 1-8 are not unique: other ones exist which are logically equivalent, i.e., lead to the same optimal solution. Nor are these programs the simplest possible ones. For instance the constraints (9) in problem 7 could be replaced by

$$y_s \le 1 - x_{s1}$$
 $y_s \le 1 - x_{s4}$
 $y_s \le 1 - x_{s2}$ $y_s \le 1 - x_{s5}$ $\forall s$
 $y_s \le 1 - x_{s3}$ $y_s \le 1 - x_{s6}$ (9')

The right-hand sides of (9') contain one x variable, corresponding to an edge of s, instead of two, corresponding to adjacent edges. If the x variables take values 0 or 1 both (9) and (9') force y_s down to 0 as soon as one x variable is equal to 1. The reason for preferring (9) to (9') is the following: mixed-integer programs are usually solved by branch-and-bound techniques, and the efficiency of the solution process is highly dependent on problem formulation. In this process, the integrality constraints on the x variables are first relaxed and replaced by

$$0 \le x_{sj} \le 1 \ \forall s, \quad j = 1, 2, ..., 6$$

This leads to a linear program, the linear relaxation of the given mixed-integer program. If all variables take integer values in the solution of this linear program, the optimal solution has been found. Otherwise, branching takes place: a variable taking a fractional value is selected and branched upon by fixing it at 0 on the one hand and at 1 on the other. The procedure is then iterated on each subproblem. Moreover, if the optimal value of the linear relaxation of the current subproblem is more than that of the best integer solution yet found (when minimizing) or if this subproblem is infeasible, it can be discarded. Optimal values of subproblems depend crucially on how tight is the linear relaxation, i.e., how close is the solution in continuous variables of the linear relaxation to the solution in integers. Constraints (9) are tighter than constraints (9'): in particular they imply that any hexagon of H with a vertex of degree 2 must have $y_s = 0$ (due to the matching constraints (1)), whereas constraints (9') do not.

3. COMPUTATIONAL EXPERIENCE

We first illustrate on one small polyhex the results of the eight mathematical programs presented in the last section (see Figure 1). For this polyhex, computing time using the CPLEX code (on a SUN SPARC station) is very moderate.

Efficiency of the proposed approach is then evaluated by solving all eight problems for members of a parametric family of polyhexes with an increasing number of hexagons. The resulting mixed-integer programs are quite large: they have several hundred constraints and variables. The CPU times and number of nodes in the branch-and-bound tree are given in Figure 2. It appears that large problems, with over 100 hexagons can be solved in reasonable computing time. For these examples, the difficulty of resolution (reflected in the number of nodes of the branch and bound tree) varies greatly

from problem to problem. Finding max n_3 (the Fries number), max n_0 , and min n_0 seems to be easy, whereas finding min n_3 and max n_2 takes much more time.

ACKNOWLEDGMENT

P.H. has been supported by FCAR (Fonds pour la Formation de Chercheurs et l'Aide à la recherche) Grant 92EQ1048, NSERC (Natural Science and Engineering Research Council of Canada) Grant to H.E.C. and Grant GP105574, and AFOSR Grants 89-0512B and F49620-93-1-0041 to Rutgers University. M.Z. has been supported by the International Fellowships Program of NSERC.

REFERENCES AND NOTES

- (1) Clar, E.; Zander, M. 1:12-2:3-10:11-Tribenzoperylene. J. Chem. Soc.
- Clar, E. Polycyclic Hydrocarbons, Academic Press: New York, 1964; Vols. I, II.
- Clar, E. The Aromatic Sextet; Wiley: London, 1972.
- Gutman, I.; Cyvin, S. J. Introduction to the Theory of Benzenoid Hydrocarbons; Springer: Berlin, 1989.
- (5) Cyvin, S. J.; Gutman, I. Kekulė Structures in Benzenoid Hydrocarbons; Springer: Berlin, 1988.
- (6) Dias, J. R. Handbook of Polycyclic Hydrocarbons, Part A: Benzenoid hydrocarbons; Elsevier: Amsterdam, 1987.
- (7) Dias, J. R. Handbook of Polycyclic Hydrocarbons, Part B: Benzenoid hydrocarbons; Elsevier: Amsterdam, 1988.
- (8) Trinajstić, N. Chemical Graph Theory, 2nd ed.; CRC Press: Boca Raton, FL. 1992.
- Balaban, A. T., Ed. Chemical Applications of Graph Theory; Academic Press: London, 1976.
- (10) Rouvray, D. H.; Balaban, A. T. In Applications of Graph Theory; Wilson, R. J., Beineke, L. W., Eds.; Academic Press: London, 1979; pp 177-
- (11) Sachs, H. Perfect Matchings in Benzenoid Systems. Combinatorica 1984, 4, 89-99.
 (12) John, P.; Sachs, H. Wegesysteme und Linearfaktoren in hexagonalen
- und quadratischen Systemen. In Graphen in Forschung und Unterricht; Bodendiek, R., et al., Eds.; Franzbecken Verlag: Bad Salzdetfurth, 1985;
- (13) Sheng, R. Identification of The Kekulé Structures of A Hexagonal System. Chem. Phys. Lett. 1987, 142, 196-199.
- (14) Sheng, R. Rapid Ways to Recognize Kekuléan Benzenoid Systems. Top. Curr. Chem. 1990, 153, 211-226.
- (15) He, W.; He, W. On Kekulé Structure and P-V Path Method. Stud. Phys. Theor. Chem. 1987, 51, 476-483.
 (16) Hansen, P.; Zheng, M. A Revised Peeling Algorithm for Determining
- If a Benzenoid System is Kekuléan. J. Mol. Struct. (THEOCHEM). 1991, 235, 293-309.
- (17) Hansen, P.; Zheng, M. A Linear Algorithm For Perfect Matching in Hexagonal Systems. Dis. Math. 1993, 122, 179-196.
- (18) Hopcroft, J. E.; Karp, R. M. An^{3/2} Algorithm for Maximum Matchings in Bipartite Graphs. SIAM J. Comput. 1973, 2, 225-231.
- (19) Hansen, P.; Zheng, M. Upper Bounds for The Clar Number of Benzenoid Hydrocarbons. J. Chem. Soc., Faraday Trans. 1992, 88, 1621–1625.
- (20) Hansen, P.; Zheng, M. The Clar number of a Benzenoid System and Linear Programming. J. Math. Chem., to appear.

 (21) Nemhauser, G. L.; Wolsey, L. A. Integer and Combinatorial Optimi-
- zation; Wiley: New York, 1988.
 (22) Harary, F.; Klein, D. J.; Zivkovic, T. P. Graphical Properties of Polyhexes: Perfect Matching Vector and Forcing. J. Math. Chem. 1991, 6, 295-306.
- (23) Sahini, V. E. Sur la Stabilité des Hydrocarbones Aromatiques Condensés. J. Chim. Phys. Phys.-Chim. Biol. 1961, 59, 177.
- (24) Sahini, V. E. On the Stability of Aromatic Hydrocarbon. On the Linearity of the Delocalization Energy Relationships. Rev. Roum. Chim. 1977,
- (25) Balaban, A. T. Benzenoid Catafusenes: Perfect Matchings, Isomerization, Automerization. Pure Appl. Chem. 1993, 65, 1.
- (26) Balaban, A. T.; Liu, X.; Cyvin, S. J.; Klein, D. J. Benzenoids with Maximum Kekulé Structure Counts for Given Numbers of Hezagons.
- J. Chem. Inf. Comput. Sci. 1993, 33, 429-436.
 Hansen, P.; Zheng, M. The Maximum Number of Kekulé Structures of Cata-condensed Polyhexes. Z. Naturforsch. 1993, 48a, 1031-1038.
 Schrijver, A. Theory of Linear and Integer Programming; Wiley: New
- York, 1986.
- (29) Fries, K. Über Bicyclische Verbindungen und ihren Vergleich mit dem Naphtalin. Ann. Chem. 1927, 454, 121-324.

 (30) Randić, M. Conjugated Circuits and Resonance Energies of Benzenoid
- Hydrocarbons. Chem. Phys. Lett. 1976, 38, 68-70.
- (31) Randić, M. Aromaticity and Conjugation. J. Am. Chem. Soc. 1977, 99,
- (32) Hosoya, H. Clar's Aromatic Sextet and Sextet Polynomial. Top. Curr. Chem. 1990, 153, 255-272.