Allowed Boundary Sequences for Fused Polycyclic Patches and Related Algorithmic Problems

Michel Deza*

Department of Mathematics and Informatics, Ecole Normale Supérieure, 45 Rue d'Ulm, 75005 Paris, France

Patrick W. Fowler

School of Chemistry, University of Exeter, Stocker Road, Exeter EX4 4QD, U.K.

Viatcheslav Grishukhin

CEMI, Russian Academy of Sciences, Nakhimovskii pr. 47, 117418 Moscow, Russia

Received June 5, 2000

We consider sequences that encode boundary circuits of fused polycycles made up of polygonal faces with p sides, $p \leq 6$. We give a constructive algorithm for recognizing such sequences when $p = 5$ or 6. A simpler algorithm is given for planar hexagonal sequences. Hexagonal and pentagonal sequences of length at most 8 are tabulated, the former corresponding to planar benzenoid hydrocarbons C_xH_y with y up to 14.

1. INTRODUCTION

Many problems in mathematical chemistry reduce to the enumeration of polycyclic graphs having *fixed* configurations of fused p-faces and, in particular, to deciding the question of when such graphs exist. In this paper we give an algorithm for construction of some graphs of the polycyclic type. The usual approach is to construct fused polycycles with a given number of p-faces. Here we construct them with a given number of boundary vertexes of valency 3.

For this purpose, we consider one-connected *polygonal systems*, i.e., simple graphs consisting of polygons such that the intersection of any two polygons is either empty or consists of one edge. This implies that all interior vertexes have valency 3 and any boundary vertex has valency either 2 or 3. Such graphs (or maps) are of importance in organic, physical, and mathematical chemistry, where the vertexes usually represent carbon atoms, and boundary vertexes of valence 2 are either left unsaturated or completed to valence 3 by addition of a hydrogen atom.

We are interested especially in *hexagonal systems* (also called *benzenoid systems* or *benzenoid graphs*), which consist entirely of hexagons. To the boundary circuit C of length n is related a sequence $a(C) = a_1...a_k$ such that $n = k + \sum_1^k a_i$, where k is the number of vertexes of valency 3 and a_i is the number of vertexes of valency 2 between the ith and $(i + 1)$th vertexes of valency 3. The sequence $a(C)$ is *hexagonal* if C is the boundary of a hexagonal system.

If the hexagons of a hexagonal system can be realized as regular polygons, both equilateral and equiangular, then the hexagonal system is called a *polyhex*. Polyhexes are used to represent the molecular structure of polycyclic benzenoid aromatic hydrocarbons. If x and y are numbers of all vertexes and vertexes of degree 2 (respectively) of a polyhex, then the polyhex represents an actual or potential molecule of chemical formula C_xH_y. In the chemical realization as a molecule, actual hexagonal rings will generally depart from the equilateral, equiangular ideal to the extent allowed by the molecular point group symmetry and governed by the details of the energetics of interaction of their electrons and nuclei.

Enumeration of polyhexes is an active research topic in mathematics and mathematical chemistry (see the extensive surveys in refs 1−3) and has been attacked in many ways. It is usual to enumerate all polyhexes with a given number of hexagons. The algorithm described here enumerates boundaries with the chemical formula C_xH_y for a fixed number y. It can be applied not only to hexagonal sequences but also to pentagonal and square sequences, although in the last case an exact characterization renders its use unnecessary (see section 6). Polycycles made up of pentagons (polypentagons) have been considered before in chemistry and, in the special case of proper polypentagons (those that are subgraphs of the dodecahedron), listed exhaustively.[4] The present algorithm generates all pentagonal sequences, proper and improper.

It appears that the method of sequences described here may also be useful in any algorithm that enumerates polycyclic maps with mixed ring sizes. Similar sequences are used in a fast algorithm[5] for enumerating fullerenes (polyhedral cubic maps with 12 pentagonal faces and all others hexagonal). The task of filling the disk inside the circuit C by a planar graph which is a part of a fullerene map with C as the boundary circuit is called the *PentHex Puzzle*.[5] Several known codes that represent boundary circuits of planar maps are equivalent to our sequences (see ref 6 for a short review). For example, the sequence $a_1a_2...a_k$ is equivalent to a *zero−one code* $0^{a1}10^{a2}1...0^{ak}1$, where 0^a

* To whom correspondence should be addressed. Phone: +33 1 44322031. Fax: +33 1 44322180. E-mail: michel.deza@ens.fr.

10.1021/ci000060o CCC: $20.00
Published on Web 01/12/2001

is a sequence of a zeros[7] and if $a = 0$, the sequence 0^a is empty. The sequence $a_1a_2...a_k$ is also equivalent to the *boundary-edges code* $a_1 + 1, a_2 + 1, ..., a_k + 1$ introduced in ref 6.

Codes that concentrate on boundaries may have advantages in a chemical context, as they lend themselves to description of the benzenoids and related molecules in terms of their overall *shapes*. Shape is intimately associated with molecular electric and steric properties, such as quadrupole moment or van der Waals envelope, which influence intermolecular interactions and are implicated in structure–activity relationships of many kinds,[8] spanning such diverse phenomena as odor perception (see e.g. various papers in the conference volume MSOQ) and carcinogenicity (as in the early bay-region theory of the activity of benzenoids; see e.g. ref 10). The present approach, in which polycycles are generated according to the length of the code, has the property that examples of large polyhexes appear sooner in the generation sequence earlier than they would with methods working strictly according to the increasing total number of faces. This feature may be an advantage in some chemical contexts. Maps containing mixtures of ring sizes, including the fullerenes as noted above, but also fluoranthrenoids, indacenoids, and more general structures, are of interest in the chemistry of polycyclic hydrocarbons; the formal framework of Euler relations described in sections 2 and 3 has been extended to cover the case of structures with various different perimeters.[11]

The main aim of the present paper is to present an algorithm for *recognizing* hexagonal and pentagonal sequences; several intermediate tools for reduction of hexagonal and pentagonal sequences are presented. In the final section of the paper, a simpler algorithm is given that recognizes sequences related to planar polyhexes.

2. BOUNDARY CIRCUITS OF A MAP

Consider a simple cubic planar map M, i.e., a planar cubic graph embedded in the plane. A *face* of M is a circuit of M that is the boundary of a simply connected part of the plane not containing vertexes and edges of M. A p-gonal face (p-face, p-gon) is a face with p vertexes and p edges. We consider maps M_p having a number of mutually adjacent p-gonal faces, for fixed $p \leq 6$, and review some definitions, notation, and results needed in the later discussion.

Consider a simple (i.e. non-self-intersecting) circuit C of M_p. C is a *boundary circuit*. Since M_p is planar, C has an inner domain $D(C)$ and an outer domain $\bar{D}(C)$. Since M_p is cubic, each vertex of C is incident additionally with an edge not belonging to C.

Call a *tail* a vertex of C with half of the edge of M_p that is incident to this vertex of C but does not belong to C. Let C have k tails t_i, $1 \leq i \leq k$, lying in the inner domain $D(C)$. We can relate to C the sequence $a(C) = a_1a_2...a_k$, where a_i is the number of vertexes of C lying between t_i and t_{i+1}. (Tails incident on these a_i vertexes lie in the outer domain $\bar{D}(C)$.) The number k is the *length* $l(a)$ of the sequence $a(C)$. Sequences $a(C)$ are considered identical up to reversal and cyclic shifts; i.e., all the sequences $a^{q\pm} \equiv a_{q\pm1}a_{q\pm2}...a_{q\pm k}$, $q = 0, 1, ...$ (all indices with equal signs), are considered as identical, and indices in the sequences are to be taken modulo k. If C has n vertexes, then the sum $\sum_1^k a_i + k$ is equal to n. The difference $\sum_1^k a_i - k$ is the *leftness* of C.[12] (No orientation of C is implied by this term.)

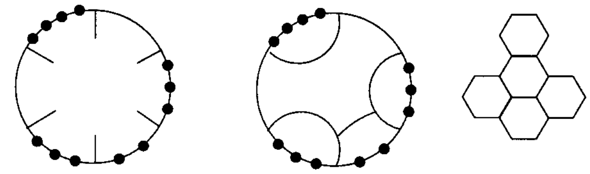

Figure 1. Circuit $C(a)$ (left), the map $M_p(a)$ (center), and the polyhex of chemical formula $C_{20}H_{12}$ (right) related to the sequence $a = 403230$. Note that all entries in the code lie between 0 and 4, so that the code can be written unambiguously without commas or spaces.

$$\sum_1^k a_i - k = \sum_1^{l(a)} a_i - l(a) = \lambda(a) \equiv \lambda(C) \qquad (1)$$

Denote by $M_p(a)$ a submap of M_p induced by the vertexes of $C \cup D(C)$. In what follows, we suppose that $M_p(a)$ is distinct on a p-gon. If a p-gon h of $M_p(a)$ is adjacent to C, then there is at least one edge of h not belonging to C. Hence we have

$$0 \leq a_i \leq p - 2, \quad \text{for all } i \qquad (2)$$

For $0 \leq r \leq p - 2$, let n_r be the number of i with $a_i = r$. Then

$$\lambda(a) = 3n_4 + 2n_3 + n_2 - n_0 \qquad (3)$$

where $n_4 = 0$ if $p < 6$, and $n_3 = 0$ if $p < 5$.

There are $(p - 1)^k$ distinct sequences a of length k satisfying (2). Denote by $D(a)$ the disk whose boundary is the circuit $C(a)$ related to the sequence. Call a sequence $a = a_1a_2...a_k$ *p-gonal* if it corresponds to the boundary circuit of a partition $M_p(a)$ of the disk $D(a)$ into p-gons (Figure 1). For $p = 5$ or $p = 6$, the sequence is *pentagonal* or *hexagonal*, respectively. Let $h(a)$ be the number of p-gons in this partition. Clearly, a p-gonal sequence satisfies (2), but most sequences satisfying (2) are not p-gonal. Note that $p - 2$, $p - 2 \equiv (p - 2)^2$ and $p - 3$, $p - 3$, $p - 3 \equiv (p - 3)^3$ are the shortest possible p-gonal sequences, and $h(a) = 2$ and 3, respectively. If a is p-gonal, then it may correspond to more than one distinct map $M_p(a)$. Hence, in what follows, when discussing a p-gonal sequence a, we always have in mind one possible map $M_p(a)$, rather than a unique map.

A condition on the leftness $\lambda(a)$ of a sequence a can be obtained from the Euler relation. Suppose that a planar map $M_p(a)$ has the boundary circuit C. The vertexes of C have valency 2 or 3. Let υ_2 and υ_3 be the numbers of all vertexes of C of valency 2 and 3, respectively. If a is the sequence corresponding to C, then its length $l(a) = k$ is equal to υ_3, $\sum_1^k a_i = \upsilon_2$, and $\lambda(a) = \upsilon_2 - \upsilon_3 = \lambda(C)$.

Lemma 1. *Let a be an arbitrary sequence. Suppose there is a map $M_p(a)$ corresponding to a. Then, for $p \leq 6$, the leftness $\lambda(a)$ satisfies the following inequality*

$$2\lambda(a) \leq 2p - (6 - p)k \qquad (4)$$

with equality only for $p = 6$.

Proof. We use the Euler relation $\upsilon - e + f = 1$, where υ, e, and f are the numbers of vertexes, edges, and faces of

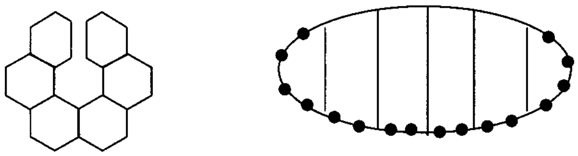

Figure 2. Nonplanar polyhex (right) and its map (left), related to the sequence $42^4 40^4$.

$M_p(a)$. We have $v = v_2 + v_3 + z$, $2e = 2v_2 + 3(v_3 + z)$, and $pf = v_2 + 2v_3 + 3z = v_2 - v_3 + 3(v_3 + z)$, where z is the number of interior vertexes of $M_p(a)$. Substituting these values in the Euler relation, we obtain

$$2(v_2 - v_3) + (6 - p)(v_3 + z) = 2p$$

Recall that $v_2 - v_3 = \lambda(a)$ and $v_3 = k$. For $p \leq 6$, we have $(6 - p)z \geq 0$, and this equality implies (4).

The minimal nonempty sequence $a(C) = p - 2, p - 2 \equiv (p - 2)^2$ corresponds to the boundary of two adjacent p-gons and hence has $v = 2p - 2$, $v_3 = 2$, $v_2 = 2p - 4$, $z = 0$, $e = 2p - 1$, $f = 2$, and $\lambda = 2p - 6$.

3. HEXAGONAL SYSTEMS

For hexagonal sequences (i.e. $p = 6$) we reformulate lemma 1 as follows.

Corollary 1. *A hexagonal sequence a has leftness*

$$\lambda(a) = 6 \tag{5}$$

Equations 1−6 or equivalent results can be found in various places in the literature, e.g. ref 3, p 29. A chemical formula can be associated with a hexagonal system P. Let a sequence a of length $k = l(a)$ correspond to the hexagonal system P. Let $h(a)$ be the number of hexagons in P. Let $M(a)$ be the graph corresponding to P. Let x and y be the numbers of all vertexes and vertexes of degree 2 of $M(a)$, respectively. In the notation of lemma 1, $x = v_2 + v_3 + z$ and $y = v_2$. All vertexes of degree 2 lie on the boundary of P and therefore belong to $C(a)$.

The parameters x and y are related to k and $h(a)$ as follows. By definition of a hexagonal sequence, $y = v_2 = \sum_{i=1}^{k} a_i$. Hence by corollary 1 and equality 1 we have

$$y = k + \lambda(a) = k + 6$$

By the Euler relation, we have $6h(a) = v_2 - v_3 + 3(v_3 + z) = \lambda(a) + 3(x - v_2) = 6 + 3(x - y)$. This implies $x = y + 2h(a) - 2$; i.e.

$$x = 2h(a) + k + 4$$

It follows that all hexagonal systems with the same numbers of vertexes of degree 2 and the same numbers of hexagons have the same chemical formula C_xH_y.

Recall that we represent hexagonal systems by planar maps. If a hexagonal system represents a polyhex, this polyhex may be either planar or not. A polyhex is *planar* if it can be embedded in the plane with no edges crossing; otherwise it is *nonplanar* (see Figure 2). In the final section of this paper we will give a simple polynomial algorithm for recognizing hexagonal sequences corresponding to planar polyhexes.

4. REDUCTIONS OF *P*-GONAL SEQUENCES

We want to know whether a given sequence a corresponds to the boundary circuit C of a map $M_p(a)$. A natural way to recognize whether a sequence a is p-gonal is to reduce it to a shorter sequence, so we describe below some reductions that preserve p-gonality.

Recall that there are $a_i + 1$ edges of C between the tails t_i and t_{i+1}. These $a_i + 1$ edges are common to C and a p-gon. Denote this p-gon by h_i, and call these common edges *main* edges of h_i. Denote by v_i the vertex of the tail t_i common to C. It is possible that a given p-gon h_i has other edges that are also common to C in addition to its main edges. This is the case when h_i coincides with a p-gon h_j, $j \neq i$, and, without loss of generality, $i < j$. Lemma 2 below describes one case when this coincidence of faces can be ruled out.

Lemma 2. *If a is a p-gonal sequence and* $a_i + a_j > p - 4$, *then the p-gons* h_i *and* h_j *do not coincide.*

Proof. If h_i and h_j coincide, then the common p-gon contains at least $(a_i + 1) + (a_j + 1) + 2 > p$ edges. This is a contradiction.

4.1. Reduction f_1: Cutting Out a *p*-gon. This first reduction corresponds to deletion of a p-gon from the map $M_p(a)$.

Suppose that h_i has only main edges common with C. Connect the tails t_i and t_{i+1} in an arc. This arc belongs to the boundary circuit of the p-gon h_i. If $a_i = p - 2$, then h_i has $p - 1$ edges common with C. The pth edge connects two vertexes of C; i.e., the pth edge is a gluing of the tails t_i and t_{i+1}. This edge is also an edge of two coinciding p-gons $h_{I-1} = h_{i+1}$. If $a_i = p - 3$, then h_i has $p - 2$ edges common with C, and two other edges of h_i are edges of h_{I-1} and h_{i+1}. Hence the p-gons h_{I-1} and h_{i+1} have a common edge that is incident to a vertex of h_i.

In general, suppose that h_i has no common edges with $C(a)$ apart from main edges. Connect the neighboring tails t_i and t_{i+1} to form the p-gon h_i. Consider the new circuit C' that is the boundary of $D(C)$ minus the p-gon h_i. Then we obtain the new sequence

$$a(C') = f_1(a_1 ... \hat{a}_i ... a_k).$$

Here the operation f_1 transforms $a(C)$ such that the triple $a_{I-1}a_ia_{i+1}$ (denoted as the triple abc) is changed in accordance with a value of b, $0 \leq b \leq p - 2$, as follows

$$f_1(a\hat{b}c) = \begin{cases} a + 1, 0^{p-3-b}, c + 1 & \text{if } b \leq p - 3 \\ a + c + 2 & \text{if } b = p - 2 \end{cases} \tag{6}$$

If $abc = a_{I-1}a_ia_{i+1}$, we say that the operation f_1 is *applied to the sequence a with respect to* a_i and *closes* the p-gon h_i. For example, applying f_1 to the sequence $(p - 3)^3$ with respect to the middle $p - 3$, we obtain the shortest sequence $p - 2$, $p - 2$ (Figure 3).

Recall that f_1 may be applied to close the p-gon h_i if the only edges of h_i common to C are the main edges. In this case we say that the application of f_1 is *feasible*.

Note that f_1 may increase, decrease, or leave unchanged the length of $a(C)$. If $c = p - 2$ or $c = p - 3$, then the

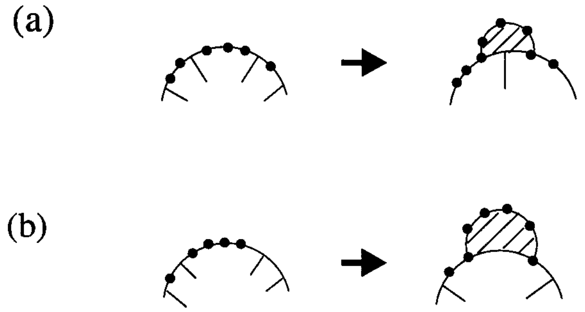

Figure 3. Reductions of type f_1 for $p = 6$. (a) A sequence ...231... is transformed to ...32... with exclusion of the shaded hexagon. This is the case $b = p - 3$ of eq 6. (b) A sequence ...140... is transformed to ...3... with exclusion of the shaded hexagon. This is the case $b = p - 4$ of eq 6.

length of a decreases. The following lemma asserts that f_1 preserves p-gonality: application of the operation f_1 with respect to $a_i = p - 2$ or $a_i = p - 3$ in a p-gonal sequence again produces a p-gonal sequence.

Lemma 3. *If a is a p-gonal sequence and $a_i = p - 2$ or $a_i = p - 3$, then the p-gon h_i has only main edges common with C.*

Proof. If h_i has an edge in common with C in addition to the $a_i + 1$ main edges, then it coincides with some p-gon h_j, $j \neq i$, with $a_j \geq 0$. Now lemma 2 implies lemma 3.

So, if a sequence has elements $p - 2$ or $p - 3$, we can reduce it to a shorter sequence. If $a_i \leq p - 4$ for all i, then the operations f_1 do not reduce the length of a. Call such a sequence *irreducible*. After application of the operations f_1 several times to a sequence that is not irreducible, we obtain either an irreducible sequence or the shortest p-gonal sequence $p - 2$, $p - 2$. Hence we need to be able to recognize p-gonality of irreducible sequences.

4.2. Reduction *f*: Cutting Out Multiple *p*-gons. Now we describe a reduction procedure for an irreducible sequence. Again we suppose that the only edges of h_i common with C are the main edges.

Let a have the sub-sequence

$$b = x, p - 4, (p - 5)^q, p - 4, y$$

of length $l(b) = q + 4$. (It may be that $l(b) = q + 3$, if x coincides with y, and then $b = a$.) If we apply f_1 to b with respect to the first $p - 4$, we obtain, according to (6), the following sequence

$$b' = x + 1, 0, p - 4, (p - 5)^{q-1}, p - 4, y.$$

Now again apply f_1 to b' with respect to the first $p - 4$, to give $b'' = x + 1, 1, 0, p - 4, (p - 5)^{q-2}, p - 4, y$. Similarly, applying f_1 $q - 2$ times to b'' with respect to the first $p - 4$, we obtain the sequence $x + 1, 1^q0, p - 3, y$. Now we apply f_1 with respect to $p - 3$ and obtain the sequence

$$f(b) = x + 1, 1^{q+1}, y + 1$$

of length $l(f(b)) = q + 3 = f(b) - 1$. We denote by f $q + 1$ consecutive applications of the operation f_1. If we apply separate the operation f_1 to any member of b of type $c = p - 4$ or $c = p - 5$, then according to (6) the length of b is either unchanged (if $c = p - 4$) or increased.

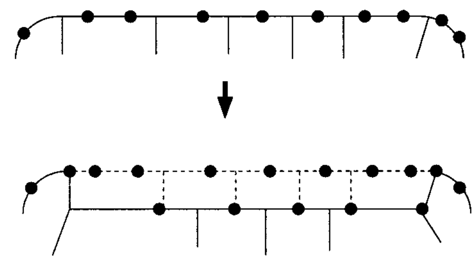

Figure 4. Reduction f for $p = 6$ with $q = 3$. A sequence ...$x21112y$... i.e., ...$x(p - 4)(p - 5)^q(p - 4)y$... is transformed to ...$(x + 1)1^4(y + 1)$..., with elimination of the dotted portion of the polyhex.

Let $p = 6$. An irreducible hexagonal sequence consists of 0's, 1's, and 2's. A sequence of type b has the form $x21^q2y$ (Figure 4).

Lemma 4. *An irreducible hexagonal sequence always has a sub-sequence $x21^q2y$ for some $q \geq 0$.*

Proof. Let a be a hexagonal irreducible sequence. According to (3) and (5), for $p = 6$, the leftness of a is equal to

$$\lambda(a) = n_2 - n_0 = 6$$

This implies that a has at least six pairs of consecutive twos without zeros between them. This proves the lemma.

Let $p = 5$. An irreducible pentagonal sequence consists of 0's and 1's. A sequence of type b has the form $x10^q1y$. The following is straightforward.

Lemma 5. *An irreducible pentagonal sequence containing at least two ones has a sub-sequence $x10^q1y$ for some $q \geq 0$.*

Now we give a sufficient condition for applicability of the operation f to a sub-sequence of type b of a p-gonal sequence a.

Consider a sequence a that corresponds to a partition $P(a)$ into p-gons of the disk $D(a)$ with the boundary circuit $C = C(a)$. Let $M_p(a)$ be the graph (map) of $P(a)$. Suppose that $M_p(a)$ is such that there are two vertexes v_i and v_j, $j \neq i \pm 1$, connected by an edge. Call this edge a *waist* of a. We say that a p-gonal sequence *a has no waist* if no map $M_p(a)$ has a waist. If a has no waist, any two vertexes of $C(a)$ are connected by a path of more than one edge not belonging to $C(a)$. Hence we have the following analogue of lemma 2.

Lemma 6. *If a is a p-gonal sequence without a waist and $a_i + a_j > p - 6$, then p-gons h_i and h_j are not coincident.*

An equivalent statement is as follows.

Corollary 2. *If a is a hexagonal sequence without a waist, then hexagons h_i and h_j, $j \neq i \pm 1$, coincide only if $a_i = a_j = 0$.*

If a is a pentagonal sequence without a waist, then no two pentagons h_i and h_j, $j \neq i \pm 1$, coincide.

This Corollary implies the following important property.

Lemma 7. *If a p-gonal sequence has no waist, then the operation f can be applied to a sub-sequence of type b.*

Proof. Recall that an application of f consists of $q + 1$ consecutive applications of the operation f_1. We have to verify that each application of f_1 is feasible. Obviously, the first application of f_1 is feasible. The next applications are feasible if the p-gon closed by f_1, say h_i, does not coincide with another p-gon, say h_j. Clearly, h_j may correspond only to a new element a_i with $j \neq i - 1$. But any new element a_j, $j \neq i - 1$, equals 1. By construction of f, $a_i = p - 4$. Hence

Table 1. Hexagonal Sequences of Length at Most 8

k	$n(k)$	$h(a)$	C_xH_y	hexagonal sequences				
2	1	2	$C_{10}H_8$	44				
3	1	3	$C_{13}H_9$	333				
4	3	3	$C_{14}H_{10}$	4240	4141			
		4	$C_{16}H_{10}$	3232				
5	2	4	$C_{17}H_{11}$	41330				
		5	$C_{19}H_{11}$	32231				
6	12	4	$C_{18}H_{12}$	404040	422400	421410	420420	411411
		5	$C_{20}H_{12}$	403230	330330	412320		
		6	$C_{22}H_{12}$	322230	321321	313131		
		7	$C_{24}H_{12}$	222222				
7	14	5	$C_{21}H_{13}$	4203310	4130410	4113301	4114030	4213300
				4131400				
		6	$C_{23}H_{13}$	4122310	3302320	4031320	4113220	
		7	$C_{25}H_{13}$	3221320	3213130	3212311		
		8	$C_{27}H_{13}$	3122221				
8	50	5	$C_{22}H_{14}$	42224000	42214100	42204200	42124010	42114110
				42104210	42040400	42024020	42014120	41214101
				41114111	41040401			
		6	$C_{24}H_{14}$	42123200	42032300	42023210	41303300	41230400
				41214020	41204120	41131310	41123201	41123120
				41040320	41033030	41032301	40304030	33113300
				33103310				
		7	$C_{26}H_{14}$	41222300	41213210	41032220	40312310	40303220
				40230320	33022310	32023202		
		8	$C_{28}H_{14}$	40222220	32212310	32203220	32130320	32123030
				32113211	32031311	32023121	31303130	
		9	$C_{30}H_{14}$	32122220	31222130	31221311	31213121	
		10	$C_{32}H_{14}$	22212221				

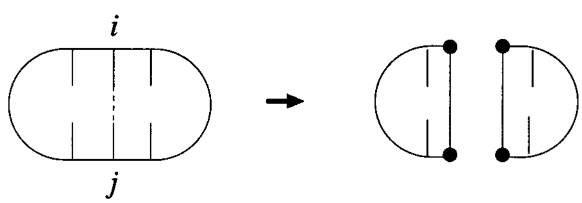

Figure 5. Reduction by cutting a sequence at a waist (ij). Cutting from the trivalent vertex between a_{I-1} and a_i to the trivalent vertex between a_{j-1} and a_j gives sequences ...$(a_{i-1} + a_j + 2)$... and ...$(a_i + a_{j-1} + 2)$....

$a_i + a_j = p - 3 > p - 4$. Now, lemma 2 asserts that h_i and h_j do not coincide. The same result holds for the final application of f_1, when $a_i = p - 3$. Hence all applications of f_1 are feasible. This proves the lemma.

4.3. Reduction of g_{ij}: Cutting the Sequence in Two. If there is an edge (ij) of $M_p(a)$ connecting the vertexes i and j, then the hexagons h_{i-1} and h_j coincide, as do the hexagons h_i and h_{j-1}.

The vertexes v_i and v_j partition the circuit C into connected components C' and C''. These components are related to two sub-sequences of a denoted by $a_ja'a_{i-1}$ and $a_ia''a_{j-1}$ (Figure 5). The edge (ij) partitions the disk $D(a)$ into two domains with boundaries $C^+ = C' \cup (ij)$ and $C^- = C'' \cup (ij)$. Denote the sequences corresponding to $C^\pm$ by $g_{ij}^{\pm}(a)$. Obviously

$$l(g_{ij}^{\pm})(a) < l(a).$$

Call the transformation of a into the two sequences $g_{ij}^{+}(a)$ and $g_{ij}^{-}(a)$ the *reduction* g_{ij}.

It is easy to check that

$$g_{ij}^{+}(a) = a', a_{i-1} + a_j + 2 \quad \text{and} \quad g_{ij}^{-}(a) = a'', a_{j-1} + a_i + 2$$

Suppose that a is irreducible. Then for sequences $g_{ij}^{\pm}(a)$ to satisfy (2), it is necessary that

$$a_{i-1} + a_j \le p - 4,\ a_{j-1} + a_i \le p - 4 \qquad (7)$$

For the sequences $g_{ij}^{\pm}(a)$ to be p-gonal, their leftness must satisfy (4). Clearly, if both sequences $g_{ij}^{\pm}(a)$ are p-gonal, then the sequence a is also p-gonal. If at least one of the sequences $g_{ij}^{\pm}(a)$ is not p-gonal, but a is p-gonal, then (ij) is not a waist of a.

5. ALGORITHM FOR RECOGNIZING *P*-SEQUENCES

Now we are ready to describe an algorithm **A** for recognizing p-gonal sequences. **A** uses reductions of types f_1, f, and g_{ij} to reduce a to a shorter sequence.

The reduction algorithm **A** is as follows. If a is not irreducible, apply the operation f_1 with respect to all $a_i = p - 2$ and all $a_i = p - 3$ until a is either $p - 2$, $p - 2$ (and hence p-gonal), or irreducible. Then for all pairs ij satisfying (7), apply reductions g_{ij}, and apply the algorithm **A** to the reduced sequences $g_{ij}^{\pm}(a)$. If these sequences are not p-gonal for all pairs (ij), then a has no waist and we can apply the reduction f. After that we apply **A** again.

Unfortunately, for $p = 5$, algorithm **A** does not recognize two classes of equivalence of pentagonal sequences of given length k. These are the class consisting of the unique sequence 0^k and the class containing the sequence 10^{k-1}. However, the pentagonality of these two sequences is easily recognized.

Algorithm **A** applied to hexagonal sequences thus gives a partial answer to the first question of Malkevitch[13] (i.e. "Characterize those binary sequences which can arise as the boundary of some patch in the regular tiling of the plane by hexagons").

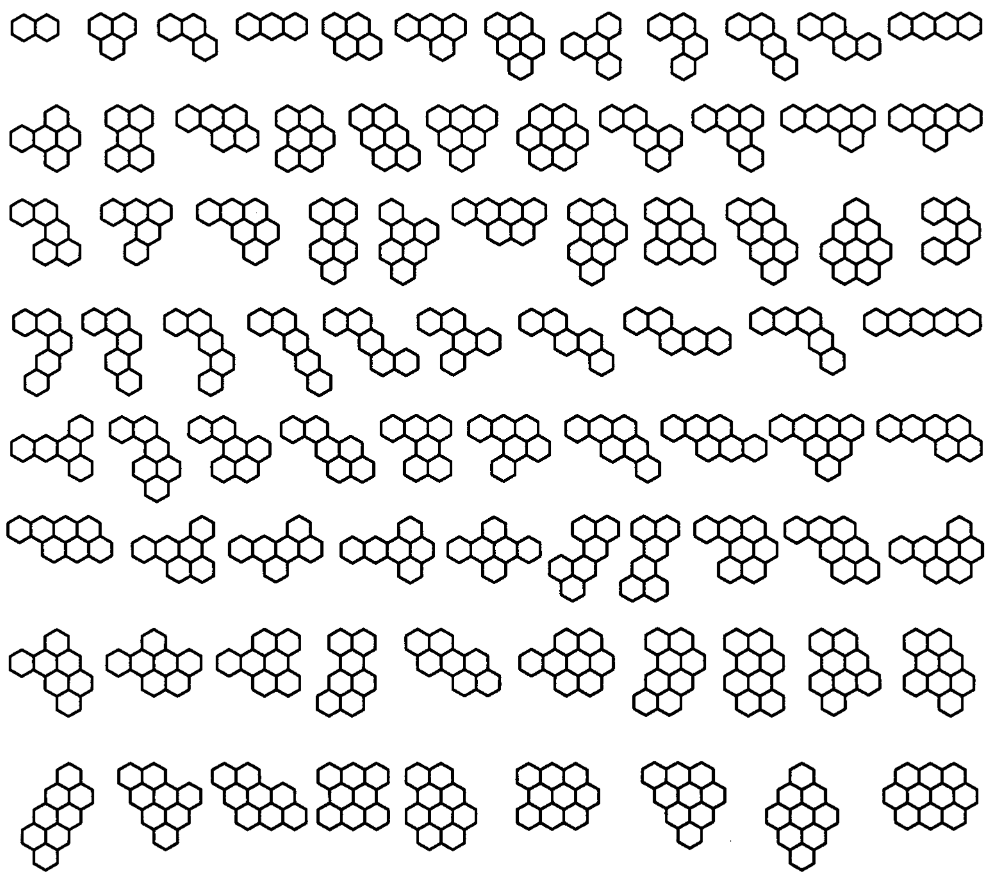

Figure 6. Polyhex realizations of the hexagonal codes of length $k \leq 8$. The diagrams follow the order of Table 1.

Table 2. Pentagonal Sequences of Length at Most 8

k	$n(k)$	$h(a)$	C_xH_y	pentagonal sequences		
2	1	2	C_8H_6	33		
3	1	3	$C_{10}H_6$	222		
4	2	3	$C_{11}H_7$	3130		
		4	$C_{12}H_6$	2121		
5	4	4	$C_{13}H_7$	30220		
		5	$C_{14}H_6$	21120		
		6	$C_{15}H_5$	11111		
		11	$C_{20}H_0$	00000		
6	9	4	$C_{14}H_8$	311300	310310	
		5	$C_{15}H_7$	301210		
		6	$C_{16}H_6$	202020	210210	
		7	$C_{17}H_5$	201110		
		8	$C_{18}H_4$	110110		
		9	$C_{19}H_3$	101010		
		10	$C_{20}H_2$	100100		
7	6	5	$C_{16}H_8$	3102200	3020300	
		6	$C_{17}H_7$	3011200		
		7	$C_{18}H_6$	2101200		
		8	$C_{19}H_5$	2010200		
		9	$C_{20}H_4$	2001100		
8	11	5	$C_{17}H_9$	31113000	31103100	31013010
		6	$C_{18}H_8$	31012100	30103010	22002200
		7	$C_{19}H_7$	30102100	30020200	
		8	$C_{20}H_6$	30011100	21002100	
		10	$C_{22}H_4$	20002000		

6. SHORT P-GONAL SEQUENCES

Using algorithm **A**, we can find all short p-gonal sequences, $p = 5$, 6, proceeding as follows.

For given length k, we generate in lexicographic order all sequences a of length k satisfying (2). In the class of sequences equivalent up to shifts and reversals of order, we take the lexicographically maximal sequence and check its p-gonality with a FORTRAN77 program. In Tables 1 and 2 we give lists of hexagonal and pentagonal sequences of length at most 8, along with the numbers $n(k)$ of sequences of length k. Note that, for $p = 5$, the Euler relation gives $h(a) = 6 + k - \sum_1^k a_i$ for the number of p-gons, and hence $x = k + 4 + h(a)$, $y = k + 6 - h(a)$ for the numbers of carbon and hydrogen atoms in the realization of the polycycle as a molecule of formula C_xH_y. The corresponding formula for $p = 6$ was given in section 3.

Table 3. Irreducible Hexagonal Sequences of Lengths 9−13

k	$n(k)$	$h(a)$	C_xH_y	irreducible sequences		
9	97	12	$C_{37}H_{15}$	221221221		
10	313	12	$C_{38}H_{16}$	2222022220		
		13	$C_{40}H_{16}$	2221212220	2221122211	
		14	$C_{42}H_{16}$	2212122121		
11	747	15	$C_{45}H_{17}$	22211221220		
		16	$C_{47}H_{17}$	22121212211		
12	2309	14	$C_{44}H_{18}$	222210222210		
		15	$C_{46}H_{18}$	222202122210		
		16	$C_{48}H_{18}$	222121122210	222120222120	222111222111
				222022202220		
		17	$C_{50}H_{18}$	222112122120	221220221220	222022121220
		18	$C_{52}H_{18}$	221212121220	221211221211	221122112211
		19	$C_{54}H_{18}$	212121212121		
13	6257	15	$C_{47}H_{19}$	2222111222200	2221122022210	
		16	$C_{49}H_{19}$	2222012212210		
		18	$C_{53}H_{19}$	2221202212210	2221112220220	2221112212201
		19	$C_{55}H_{19}$	2221121212210	2221112212120	2220221122120
				2212202122120		
		20	$C_{57}H_{19}$	2212121122120	2212112211220	2212112122111
		21	$C_{59}H_{19}$	2211212121211		

The isomer counts in Table 1 are in full agreement with previous tabulations.[1−3] Those in Table 2, for even C/H formulas, can be compared with the "Periodic Table" PCH5/PAH6;[14,15] in the previous work it was necessary to inter-

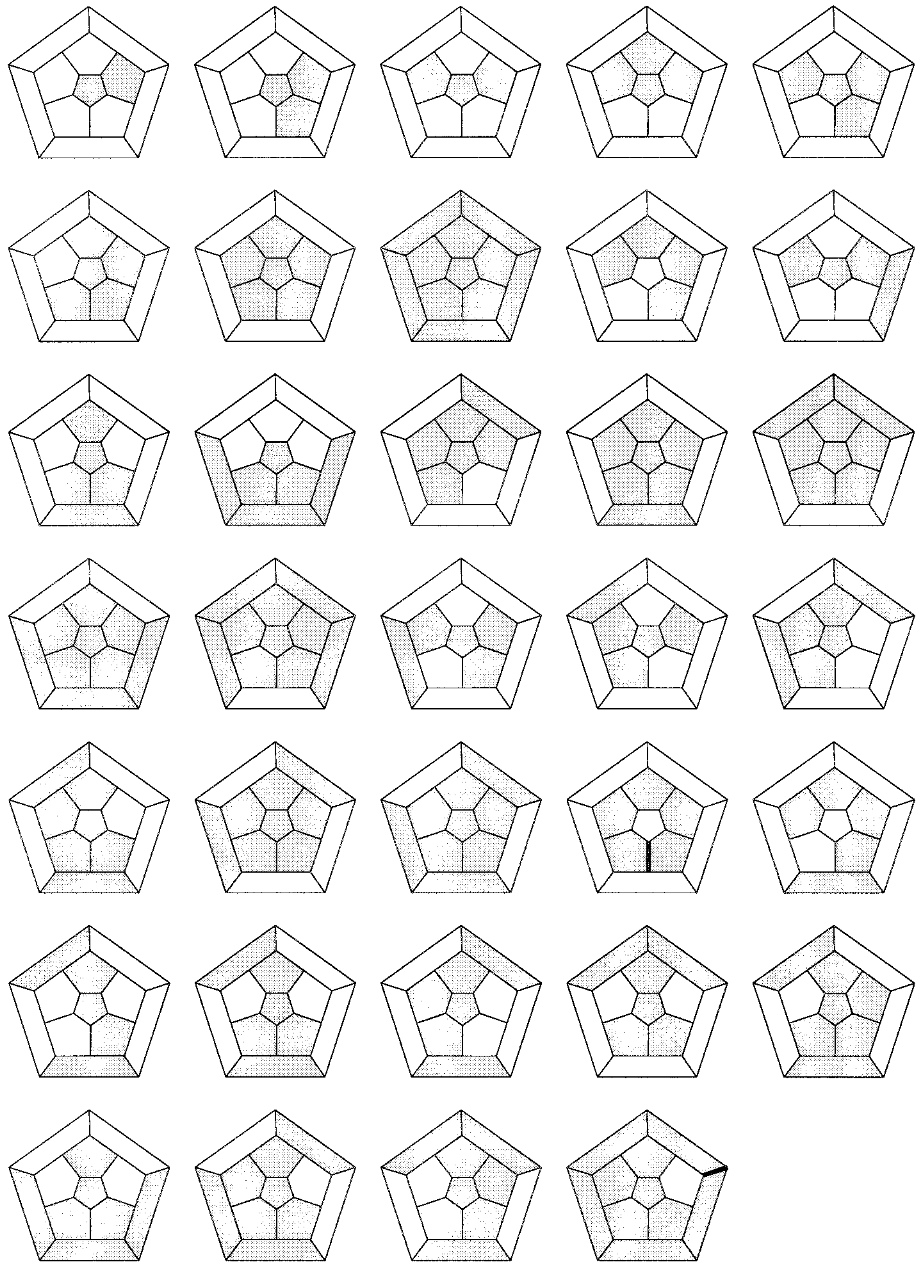

Figure 7. Illustrations of the pentagonal codes of length $k \leq 8$. Each polycycle is shown as a subset (the shaded faces) of the Schlegel diagram of the platonic dodecahedron. In two cases, a thick line is used to indicate a cut between two faces that are disjoint in the polycycle. The diagrams follow the order of Table 2.

polate to obtain counts for odd formulas; the present Table 2 does not have this limitation.

From a mathematical point of view, perhaps the most interesting hexagonal sequences are those which are irreducible hexagonal sequences. They have the maximal number of hexagons among all sequences of a given length. Note that any partition corresponding to a reducible hexagonal sequence can be obtained from a partition corresponding to an *irreducible* sequence by adding a hexagon that shares one or two edges with the original partition. In Table 3 we give the numbers $n(k)$ of all hexagonal sequences of lengths $k =$ 9, 10, 11, 12 and their corresponding irreducible sequences.

The accompanying Figures 6−9 show a filled polycyclic patch for each distinct code in Tables 1−3, given in the same lexicographic order as the codes. The hexagonal polycycles are shown in a "chemical" presentation, suitable for geometrically planar hydrocarbons. The pentagonal polyhexes, as far as the tabulation runs, can all be presented without overlap as subsets of the faces of the pentagonal dodecahedron and are shown in this way on Schlegel diagrams. Most longer codes, those corresponding to *improper* or *helical* polycycles, could not be displayed in this way without overlap of some faces. In fact there are 38 proper pentagonal polycycles, which together with the pentagon itself form the set of 39 5-gonal graphs that can be embedded on the surface of a regular dodecahedron without overlap of edges or faces (i.e. "proper polypentagons" in the terminology of ref 4). Figure 7 shows all but six of the proper pentagonal polycycles. The remaining six, which have codes of length 9, are the 6-pentagon $C_{19}H_9$ polycycles with codes 300300300, 301022001, 300230020, the 6-pentagon $C_{20}H_{10}$ 2110021100, and the 7-pentagon $C_{20}H_8$ polycycles with codes 310021100 and 220012100. The 34 codes of Table 2 include two improper pentagonal polycycles, 31113000 and 20002000,

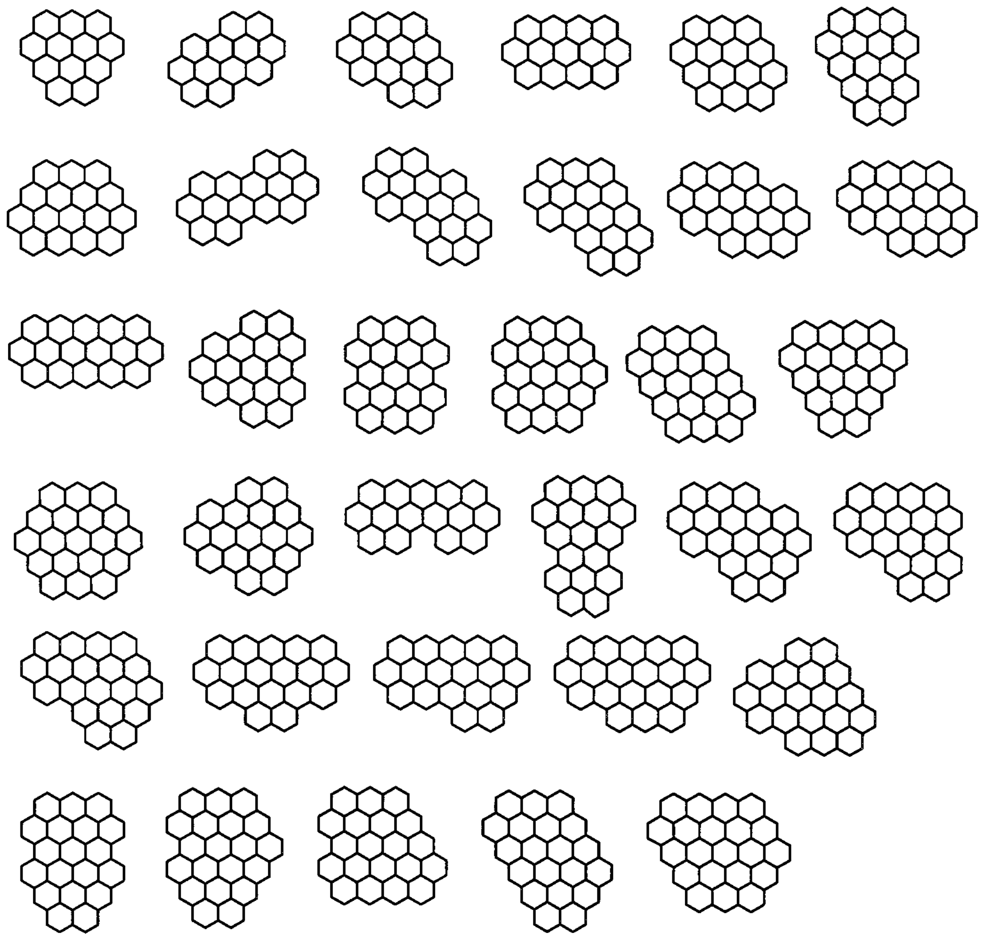

Figure 8. Polyhex realizations of the irreducible hexagonal codes of length $9 \le k \le 13$, following the ordering of Table 3.

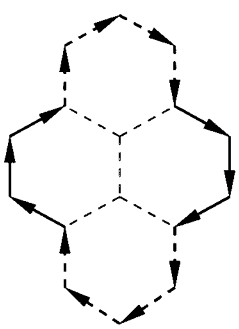

Figure 9. Hexagonal sequence and its corresponding *e*-sequence. The four strings E_j are indicated by alternately full and dashed sets of arrows running around the perimeter of the polyhex 3232.

shown in Figure 7 with overlapping edges marked by thick black lines.

In passing, we note that the case for $p = 4$ can be solved completely without the intervention of the algorithm. The polycycles with all faces 4-gonal are as follows: the cube (0^4), the cube minus a vertex (111), the cube minus an edge $10)^2)$, any finite path of h squares $(20^{h-1})^2$ for $h > 1$), and the two infinite paths $(0)^\infty$ and $2(0)^\infty$. Only the first three, plus the paths with $h \le 3$ are *proper* in that they are partial subgraphs of the cube.

7. SIMPLE ALGORITHM FOR HEXAGONAL SEQUENCES RELATED TO PLANAR POLYHEXES

In general, the algorithm **A** requires a time that is exponential in the length of the recognized sequence. However, there is a simple combinatorial algorithm (call it **B**) for recognition of sequences that correspond to planar, simply connected polyhexes. The complexity of **B** is $n(\log n)$, where n is the length of the sequence.

Let e_i, $1 \le i \le 6$, be the two-dimensional vectors:

$$e_1 = (2, 0),\ e_2 = (1, -1),\ e_3 = (-1, -1),\\ e_4 = -e_1,\ e_5 = -e_2,\ e_6 = -e_3$$

Note that $e_2 = e_1 + e_3$. These vectors generate a hexagonal lattice L, which consists of hexagons squeezed in the vertical direction. We take this form of the e_i's to allow algorithm B to work entirely in integer arithmetic. They are the shortest lattice vectors of the cubic map L. If we take the end of the vector e_i as origin, then e_i is adjacent in L to the vectors e_{i+1} and e_{I-1}. If we go from e_i to an adjacent vector, then we turn to right or left depending on whether we meet e_{i+1} or e_{I-1}, respectively.

Any path in L naturally generates an *e-sequence* of vectors e_i, and, in particular, any closed path generates such an *e*-sequence. This path determines an orientation of the corresponding closed curve. For any closed curve in L, we choose the clockwise orientation. For example, the boundary of a hexagon h of L generates the *e*-sequence $e_ie_{i+1}e_{i+2}...e_{i+5}$. As we go along the boundary of the hexagon, we always turn right.

From now on, we consider the vectors e_i for any integer $i > 0$, supposing that $e_i = e_k$, where $k \equiv i \pmod 6$, $1 \le k \le 6$, and, for simplicity, denote it by $e(i)$.

Consider a simply connected planar polyhex. Its boundary C is a plane circuit without self-intersection. Shifting and reversing operations on C generate the full set of equivalent hexagonal sequences, and from this set, we take the lexicographic maximal sequence $a = a(C)$, which determines uniquely the first vertex of C. The clockwise orientation of C generates a sequence $e(i_1)e(i_2)...e(i_n)$ corresponding to $a = a_1a_2...a_k$ with $n = \sum_{j=1}^{k} a_j + k$. Although the e-sequence determines uniquely the hexagonal sequence a, a determines the corresponding e-sequence up to a vector $e(i)$, $1 \le i \le 6$. We choose $e(i)$ such that the first vector of the e-sequence is $e(1)$.

The e-sequence corresponding to a hexagonal sequence consists of the strings

$$E_j = e(i_j)\ e(i_j{+}1)\ e(i_j{+}2)...e(i_j{+}a_j), \quad 1 \le i \le k$$

and is constructed as follows. Each string E_j corresponds to the hexagon h_j having an edge or edges in common with C. When we go along the boundary of h_j, we turn right, and when we come to the next hexagon h_{j+1}, we have to turn left. The left adjacent of $e(i_j + a_j)$ is the vector $e(i_j + a_j - 1)$. Hence the string E_{j+1} should begin with the vector $e(i_{j+1}) = e(i_j + a_j - 1)$. Clearly, $e(i_1) = e(1)$. Note that consecutive strings E_j and E_{j+1} intersect in two vectors $e(i_j + a_j - 1)$ and $e(i_j + a_j)$.

Recall that the boundary circuit C has n vertexes, say, v_i, $0 \le i \le n - 1$, where v_0 is the origin. Relabel the e-sequence of C as $e(l_1)e(l_2)...e(l_n)$. Then v_i is the end-vertex of the lattice vector

$$e(v_i) = \sum_{j=1}^{i} e(l_j)$$

Clearly, $e(v_0) = 0$. We set $\sum_{j=1}^{n} e(l_j) = e(v_n)$. Since C is closed, we have $e(v_n) = 0$. Since C does not intersect itself, $e(v_i) \neq e(v_j)$ for every pair i, j, $i \neq j$.

Now, algorithm **B** recognizing hexagonality of a sequence a is as follows. First, check whether a satisfies the conditions 2 and 5. If so, by the above rule, construct the e-sequence corresponding to a. Compute the n vectors $e(v_i)$, $1 \le i \le n$. If $e(v_n) \neq 0$, then a is not hexagonal. Otherwise, for every pair $1 \le i < j \le n - 1$, $i \neq j$, check whether the inequality $e(v_i) \neq e(v_j)$ holds. If so, a is hexagonal and corresponds to a planar polyhex.

The validity of this algorithm is clear as it explicitly constructs the related polyhex. The complexity of this algorithm is n (log n), where n is the number of vertexes v_i. Using a binary tree-search algorithm, we can find coinciding vectors $e(v_i)$ in n (log n) time. For construction of all the vectors $e(v_i)$, a time linear in n is required. Although this algorithm is to some extent “folklore” in that it is generally clear to those working in the field, it seems worth giving an explicit description for comparison with other ways of solving the problem.

This simple algorithm for recognizing polyhexes shows that a hexagonal sequence a determines uniquely the map $M(a)$ if it corresponds to a polyhex. This was also proved in ref 16. If, on the other hand, a hexagonal system cannot be embedded into the hexagonal lattice of regular hexagons, then its hexagonal sequence does not determine the map $M(a)$ uniquely. Examples and details are given in ref 17.

Finally, a remark about nonplanar cases can be made. It appears that all nonplanar polyhexes discussed in the literature can be embedded in the infinite hexagonal lattice if several edges are allowed to occupy the same position but on different levels and no two hexagons in the same position have a common edge. We believe that if we delete the condition on pairs, then algorithm **B** will in addition recognize hexagonal sequences corresponding to nonplanar polyhexes.

ACKNOWLEDGMENT

We thank the University of Exeter Research Committee and the European Union Framework IV program (TMR Network Contract FMRX CT97 0192 BIOFULLERENES) for financial support of this research.

REFERENCES AND NOTES

(1) Brunvoll, J.; Cyvin, B. N.; Cyvin, S. J. Enumeration of benzenoid systems and other polyhexes. *Top. Curr. Chem.* **1992**, *162*, 65−180.
(2) Knop, J. V.; Müller, W. R.; Szymanski, K.; Trinastić, N. *Computer Generation of Certain Classes of Molecules*; SKTH/Kemija u industriji: Zagreb, 1985.
(3) Dias, J. R. *Handbook of Polycyclic Hydrocarbons: Part A: Benzenoid Hydrocarbons, Part B: Polycyclic Isomers and Heteroatom Analogues of Benzenoid Hydrocarbons*; Physical Sciences Data Series; Elsevier: Amsterdam, 1988.
(4) Cyvin, S. J.; Cyvin, B. N.; Brunvoll, J.; Brensdal, E.; Zhang Fuji; Guo Xiaofeng; Tošić, R. Theory of polypentagons. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 466−474.
(5) Brinkmann, G.; Dress, A. W. M.; Perrey, S. W.; Stoye, J. Two applications of the Divide & Conquer principle in the molecular sciences. *Math. Program.* **1997**, *79*, 71−97.
(6) Hansen, P.; Lebatteux, C.; Zheng, M. The boundary-edges code for polyhexes. *J. Mol. Struc. (THEOCHEM)* **1996**, *363*, 237−247.
(7) Herndon, W. C.; Bruce, A. J. Perimeter code for benzenoid aromatic hydrocarbons. In *Graph Theory and Topology in Chemistry*; King, R. B., Rouvray, D. H., Eds.; *Studies in physical and theoretical chemistry* **1987**, *51*, 491−513.
(8) Hansch, C.; Leo, A. *Exploring QSAR. Fundamentals and Applications in Chemistry and Biology*; ACS Professional Reference Book; American Chemical Society: Washington, DC, 1995.
(9) *Molecular structure and organoleptic quality*; SCI Monograph No. 1; Society of Chemical industry: London, 1957.
(10) Pullman, B., Ts’O, P. O. P., Gelboin, H., Eds. *Carcinogenesis: fundamental mechanisms and environmental effects (Proceedings of the 13th Jerusalem Symposium on Quantum Chemistry and Biochemistry)*; D. Reidel Publishing Co.: Dordrecht, The Netherlands, 1980.
(11) Dias, J. R. Notes on Constant-isomer series. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 117−127.
(12) Grünbaum, B.; Zaks, J. The existence of certain planar maps. *Discrete Math.* **1974**, *10*, 93−115.
(13) Malkevitch, J. Geometrical and combinatorial questions about fullerenes. In *Discrete Mathematical Chemistry*; DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. 51; Hansen, P., Fowler, P. W., Zheng, M., Eds.; American Mathematical Society: Providence RI, 2000; pp 261−266.
(14) Dias, J. R. Deciphering the information content of polycyclic conjugated hydrocarbon formulas from benzenoids to fullerenes. *Tetrahedron* **1993**, *49*, 9207−9220.
(15) Dias, J. R. Graph theoretical invariants and elementary subgraphs of polyhex and polypent/polyhex systems of chemical relevance. *Discuss. Appl. Math.* **1996**, *67*, 79−114.
(16) Caporossi, G.; Hansen, P. Enumeration of polyhex hydrocarbons up to h = 21. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 610−619.
(17) Deza, M.; Grishukhin, V. P. Hexagonal sequences. In *Proceedings of the Conference on General Algebra and Discrete Mathematics, Potsdam 1998*; Denecke, K., Vogel, H.-J., Eds.; Shaker: Aachen, 1999; pp 47−68.

CI000060O