Pentagonal chains and annuli as models
for designing nanostructures from cages

Vladimir R. Rosenfeld[§][*], Andrey A. Dobrynin[★], Josep M. Oliva[†], Juanjo Rué[‡]

[§]*Department of Computer Science and Mathematics*
Ariel University, Ariel 40700, Israel;

[★]*Sobolev Institute of Mathematics*
Siberian Branch of Russian Academy of Sciences
4 Acad. Koptyug Ave, Novosibirsk 630090, Russia;

[†]*Instituto de Química-Física, CSIC, 28006 Madrid, Spain;*

[‡]*Freie Universität Berlin, Institut für Mathematik*
Arnimallee 3, Office 204, 14195 Berlin, Germany

Abstract

Carbon is the most versatile of chemical elements in combining with itself or other elements to form chains, rings, sheets, cages, and periodic 3D structures. One of the perspective trends for creating new molecules of nanotechnological interest deals with constructs which may be formed by chemically linking of cage molecules.

The growing interest to fullerene polyhedra and other molecules with pentagonal rings raises also a question about geometrically consistent in $\mathbb{E}^3$ nanoarchitectures which may be obtained by aggregating many such molecules. Simple examples are chains and rings assembled from pyramidal (car)borane subunits. Adequate geometrical models of such objects are a chain and an annulus built from regular pentagons wherein any two adjacent pentagons share an edge.

Among arising combinatorial problems may be both analytical and constructive enumeration of such chains and annuli drawn in plane with no two edges crossing each other. This may also employ several mathematical disciplines, such as geometry, (spectral) graph theory, semigroup theory, theory of fractals, and others.

We discuss some practical approaches for solving the mentioned mathematical problem.

Keywords: cage molecule, (car)borane, nanoarchitecture, pentagonal chains and annuli, constructive and analytical enumeration, quasicrystal.

[*]Corresponding author: Dr. Vladimir R. Rosenfeld
Email addresses: `rosenfev@tamug.edu`, `vladimir_rosenfeld@yahoo.com` (Vladimir R. Rosenfeld[§]), `dobr@math.nsc.ru` (Andrey A. Dobrynin[★]), `j.m.oliva@iqfr.csic.es` (Josep M. Oliva[†]), `jrue@zedat.fu-berlin.de` (Juanjo Rué[‡])

1. Introduction

The enumeration of chemical graphs (molecular graphs) and geometric molecular models satisfying given constraints [1–5] is one of the fundamental problems in chemoinformatics, because it leads to a variety of useful applications including structure determination and development of novel chemical compounds. One of the perspective trends for creating new molecules of nanotechnological interest deals with constructs which may be formed by chemically linking of cage molecules [3–5]. Among such constructs, chains and rings from pentagonal pyramidal clusters are rather simple instances, which have simplified "unfolded" representations by regular pentagons in $\mathbb{E}^2$ [5]. This allows to reduce the consideration of 3D molecules in question to the consideration of pentagonal chains and annuli drawn in 2D plane (see Figure 1).
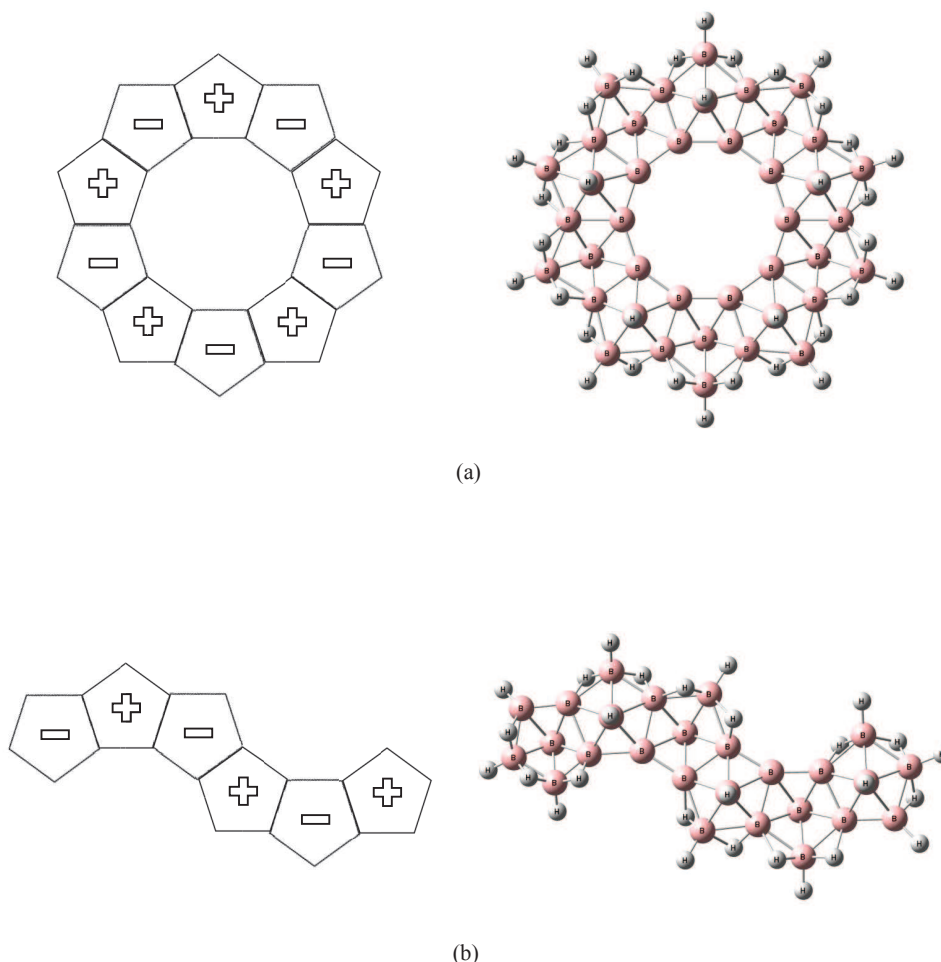

(a)

(b)

Figure 1: (a) Ten-pentagon annulus – left – and the corresponding cyclic borane molecule $B_{40}H_{40}$ – right – corresponding to an energy minimum, and therefore a potential real molecule. (b) One of the six 6-pentagon chain isomers – left – and the corresponding borane molecule – right – $B_{26}H_{30}$ corresponding to an energy minimum. The geometry of the borane molecules shown in Figure 1 is obtained using an optimizing algorithm which minimizes the energy derived from Schrödinger's equation, as the nuclei move within an energy hypersurface defined by their coordinates, using the B3LYP/6-31G* model chemistry [6]. The "+" and "−" signs represent boron atoms above and below the pentagon in the plane corresponding to the page/screen. Boron and hydrogen atoms are represented by bigger and smaller spheres, respectively. See [5] for a more comprehensive explanation of these structures.

Among the elements from the Periodic Table, aside from carbon, boron is possibly the most versatile element for forming rich architectural constructs in combination with hydrogen (boranes), carbon (carboranes), and most elements of the Periodic Table [7, 8]. One of the most conspicuous aspects of boron is the transition from 2D to 3D structures quite easily [7, 8]. Here, one of tasks is the selection and enumeration of such constructions built from equal regular pentagons.

Say, take an annulus whose connected part is formed from regular pentagons wherein two adjacent pentagons have an edge in common, and no two inadjacent pentagons or three arbitrary ones share a common point. While considering possible annuli of some bigger size, one may observe that there are also some approximate geometric solutions of a problem, when there is a slight noncoincidence of two edges (of two pentagons) which might yet "flexibly" be deformed to fuse and reform a common edge. (In perspective, we mean, of course, a full-size problem of geometrically-consistent linking of respective cage molecules comprising a molecular ring.) What do such approximate solutions mean for chemistry? As is known, all molecules as well as crystals are systems of vibrating atoms wherein the latter vibrate about some geometrically averaged coordinate points. Clearly, exact 'frozen' coordinates of atoms simply do not exist. Therefore, a question about whether approximate geometric versions of pentagonal annuli are suitable as models for nanorings-may-be may have also a positive answer. But in any such case, the conditions imposed on a problem must simultaneously conform to specific geometric and chemical requirements. Thereby, the problem cannot be solved overall and should be divided into separate smaller ones, each of which obey its own constraints.

Pentagonal chains are constructed in a similar way, with the only distinction that, therein, there are two end pentagons having each only one adjacent neighbor. The exhaustive construction of all of pentagonal chains and annuli of given size is a difficult combinatorial problem. Indeed, the simpler model of selfavoiding walks on the square grid [9, 10] has been proven to be a $\mathcal{NP}$-hard problem [9]. (See the vast bibliography on the Internet.) Here, the possibility of building borane molecules by analogy with the mathematical problem in question raises the possibilty of using quantum chemistry for helping in solving problems in mathematics.

2. Preliminaries

A *pentagonal annulus* (p.a.) has a connected part formed from regular pentagons wherein two adjacent pentagons have an edge in common, and no two inadjacent pentagons or three arbitrary ones share a common point. A *pentagonal chain* (p.c.) is constructed in a similar way, with the only distinction that, therein, there are two end pentagons having each only one adjacent neighbor.

Each p.a. (res. p.c.) made up from $n \geq 2$ pentagons is obtained by addition of the n-th pentagon to a fixed end (res. to both ends) of a certain p.c. composed of $n-1$ pentagons. In general, there are only two possibilities to attach the next pentagon to a fixed end of a growing

chain or yet to the seed pentagon providing two 'front' edges for contacts. Namely, making a turn to the 'left' or to the 'right', which we denote by characters l and r, respectively. Let $\mathcal{A} = \{l, r\}$. If one starts with a rooted pentagon (fixed in the plane), all thus derived chains of length $p \geq 3$ can be represented as $sl\ldots a_p$ or $sr\ldots a'_p$, where s is an ancillary character to temporarily mark the first (seed, or start) pentagon, and $a_p, a'_p \in \mathcal{A}$. The start character s is to be substituted by a proper character, l or r, if another character is added on the left (say, under concatenation of strings); so, also $s \in \mathcal{A}$.

Let $\mathcal{A}^*$ and $\mathcal{A}^+$ be the free monoid and free semigroup over the alphabet $\mathcal{A}$, where $\mathcal{A}^+$ is the subsemigroup of $\mathcal{A}^*$ containing all elements except the empty string (or $\mathcal{A}^+ = \mathcal{A}^* \setminus 1$). Denote by $\mathcal{A}^+_p$ the set of all words (strings) of length $p \geq 1$ over $\mathcal{A}$; $|\mathcal{A}^+_p| = 2^{p-1}$. Let $\mathfrak{C}$ and $\mathfrak{A}$ be the sets of all target p.c.'s and p.a.'s, respectively; $\mathfrak{C}, \mathfrak{A} \subset \mathcal{A}^+$.

The set $\mathfrak{C}$ (or $\mathfrak{A}$) of words is called a *language*. A language $\mathfrak{C}$ (but not $\mathfrak{A}$) is called *factorial* [11], since $\mathfrak{C} = F(\mathfrak{C})$, where $F(\mathfrak{C})$ denotes the set of all the factors of all words of $\mathfrak{C}$. (All linear factors of cyclic words from $\mathfrak{A}$ belong to $\mathfrak{C}$, too.) A language $\mathfrak{C}$ is factorial if and only if $\mathfrak{C} = \mathcal{A}^* \setminus \mathcal{J}$, where $\mathcal{J}$ is a two-sided ideal of $\mathcal{A}^*$. Here, $\mathcal{J}$ is the set of all forbidden words (representing selfcrossing p.c.'s); and, in particular, $\mathcal{J} \supset \mathfrak{A}$. The *monoid $Q(\mathfrak{C})$ of the factors of $\mathfrak{C}$* is defined as the epimorphic image of the Rees quotient monoid $\mathcal{A}^*/\mathcal{J}$ of $\mathcal{A}^*$ by the ideal $\mathcal{J} = \mathcal{A}^* \setminus F(\mathfrak{C})$ (see [11–13]). The monoid $Q(\mathfrak{C})$ is isomorphic to the monoid having as support $F(\mathfrak{C}) \cup \{0\}$ endowed with the product ($\circ$), defined as $a_1 \circ a_2 = 0$ if $a_1 \circ a_2 \notin F(\mathfrak{C})$ and $a_1 \circ a_2 = a_1 a_2$, otherwise.

Our further exposition needs an excursion into geometry. Let the inradius of a regular pentagon be equal to $1/2$; thus, the distance between the centers of two adjacent pentagons in plane is equal to 1. We assume that the center of the first (seed) pentagon is in the origin of a rectangular coordinate system, and one of its vertices lies on the y-axis. Here, we call the orientation of this pentagon *positive* if this vertex is the apex and *negative* if the nadir. Apparently, the positive and negative orientations of a pentagon can swap their places after its rotation by π radians around the 5-fold symmetry axis.

At any step of the construction of pentagonal chains and annuli, there may be used only 10 vectors v_k ($k \in [1, 10]$) (of unit length) connecting the centers of adjacent pentagons. One way to realize which vectors are these is to construct a regular 10-annulus (from ten regular pentagons) and consider ten vectors of shifts between the centers of adjacent pentagons therein. Simple manipulations with additional pentagons may demonstrate that there are no other vectors than these. In order to prove this rigorously, one my apply the following technical lemma:

Lemma 1. *Each pair of adjacent regular pentagons in plane (sharing just one edge) have opposite orientations.*

Proof. This is an obvious geometric conclusion. □

One evident corollary is:

Corollary 1.1. *Each connected construction in plane built from regular pentagons (wherein any pair thereof share just one or no edge) has just two orientations of pentagons, and all pentagons with either orientation has no point in common.*

Lemma 1 has another elementary but useful corollary:

Corollary 1.2. *Every cyclic sequence of regular pentagons in plane where every two consecutive pentagons are adjacent (in particular, an annulus) has an even length.*

Proof. By virtue of Lemma 1, any sequence of pairwise adjacent pentagons is a one where orientations of pentagons alternate. This is obeyed in a closed sequence iff it uses equal numbers of positive and negative orientations of pentagons. Hence, we arrive at the proof. □

A further generalization of Corollary 1.2 is possible. First, we need to introduce some graph-theoretic notions. Let $\Gamma(V, E)$ be a connected graph with the vertex set V and edge set E ($|V| = n > 0; |E| = m > 0$). We associate the set V with a countable set of pentagons in a plane, while two vertices of V are adjacent in Γ iff the respective two pentagons are adjacent (share just one edge). The graph Γ may correspond to any process of successive addition of subsequent pentagons, where the n-th pentagon is obligatorily adjacent to the $(n-1)$-th pentagon and may also allow an arbitrary nonadjacency intersection with earlier pentagons. The target sets $\mathfrak{C}$ and $\mathfrak{A}$ of pentagonal constructions are here only a particular case.

We state the following:

Proposition 2. *Let $\Gamma(V, E)$ ($|V| = n > 0; |E| = m > 0$) be a connected graph as above. Then, Γ is a bipartite graph.*

Proof. A graph is bipartite iff all its cycles have only even lengths. By virtue of Corollary 1.2, this condition is obeyed by $\Gamma(V, E)$. Whence the proof is immediate. □

As to an easy calculation of the 10 shift vectors v_k ($k \in [1, 10]$), one may use for them ten vectors of unit length radiating from the center of a pentagon – five directed to its vertices and the other five perpendicularly to its edges (where the latter five can be obtained by the inversion of the former five, or vice versa). We have the following general formula for the *shift vectors*:

$$v_k = \left[x \cos\left(\frac{\pi k}{5}\right) + y \sin\left(\frac{\pi k}{5}\right) \right] = \left\{ x \mathrm{Re}\left[\exp\left(\frac{\pi k i}{5}\right) \right] + y \mathrm{Im}\left[\exp\left(\frac{\pi k i}{5}\right) \right] \right\} \quad (k \in [1, 10]; i = \sqrt{-1}),$$

(1)

where $v_{k+5} = -v_k$ ($k \in [1, 5]$).

Since addition and subtraction of complex numbers do not confuse imaginary parts with real parts thereof, vectors (1) may be replaced in certain circumstances by respective coordinates in the complex plane, *viz.*:

$$v_k \mapsto u_k = t \left[\exp\left(\frac{\pi k i}{5}\right) \right] \quad (k \in [1, 10]),$$

(2)

where t may be used as a dumb variable "for counting" vectors or as a norm. However, the numbers (2) contain less information than vectors (1), because they mix 'x-component' with 'y-component' (of u_k's) under multiplication, which is used in our calculations below.

5

The set $\{v_1, v_2, \ldots, v_{10}\}$ of all these vectors has its intrinsic symmetry group D_5 and also generates a countably infinite, commutative additive group

$$H_+ = \langle v_1, v_2, \ldots, v_{10} \rangle = \langle v_1, v_2, \ldots, v_5 \rangle \tag{3}$$

of vectors, not all of which are relevant to construction of sets $\mathfrak{C}$ and $\mathfrak{A}$. Along with this additive group, we use herein for our analytical convenience also its multiplicative representation:

$$H_\times = \langle \exp(v_1), \exp(v_2), \ldots, \exp(v_{10}) \rangle \Big|_{\exp(v_k) = \exp[x \cos(\frac{\pi k}{5}) + y \sin(\frac{\pi k}{5})]} \quad (k \in [1,10]) \cdot \tag{4}$$

The form of $H_\times$ allows to more conveniently perform symbolic multiplication of matrix entries using the program packages Maple.

The number of vectors used for generating the group H_+ in (3) is 5 but we consider plane objects in $\mathbb{E}^2$ only, where maximum two vectors may be linearly independent, while the others are expressed as a linear combination of the first two. Since the five vectors have noncointegral coefficients, such linear combinations may have irrational coefficients. However, with any given accuracy, irrational coefficients can be approximated by rational numbers. That is why, besides exact solutions comprising the sets $\mathfrak{C}$ and $\mathfrak{A}$, there may also exist approximate ones. Say, the last added pentagon completing construction of an annulus may have an edge which does not exactly coincide with an edge of the other end of a closing chain. If a preliminary agreement allows to construct annuli with some 'chemical admittances', a target object may also be obtained with some allowed deformation. So, this would not practically contradict the plans of designers of respective nanobjects from cages. In a more rigorous language, we state here the following:

Proposition 3. *Let $H_+ = \langle v_1, v_2, \ldots, v_5 \rangle$ be the infinite group of vectors as defined above, and let ϵ be an arbitrary positive number. Then, there exist infinitely many pairs of vectors $v, v' \in H_+$ such that $|v - v'| \leq \epsilon$.*

Proposition 3 predicts the existence of loci of closely lying (osculating) points generated by H^+ in the plane. Such loci are responsible for developing symmetric features observed in quasicrystals [14], which is supported in [15] (see p. 62–66 and Fig. 21). Here, it is interesting to recall that a plotted function $y = \sin(\xi x)$ (ξ to be varied) whose values are calculated under integer $x \in [1, 1000]$ just displays such an appealing picture, which so closely resembles a symmetric one (we recommend to see this).

Enumeration of chains and cyclic sequences constructed from polygons was a target subject of such papers as [16–20]. The authors applied different approaches, in particular, Elk [17] used code sequences of symbols (which is also a linguistic approach), while Cyvin, Cyvin, Brunvoll, and Dobrynin [20] used symmetry considerations. The authors of [20] so enumerated any chains of regular q-gons (with a fixed $q \geq 5$), considering only symmetry nonequivalent ones; however, they allowed chains to selfcross, if this may happen, – which just should be

excluded in our case. Their overall formula (as (12) in [20]) is:

$$I_r = \frac{1}{4}(q-3)^{r-2} + \frac{1}{8}[1 + (-1)^q] + \frac{1}{8}[1 - (-1)^q]\binom{2}{r} + \tag{5}$$

$$+ \frac{1}{4}\left\{1 + (-1)^q + \frac{1}{2}[1 - (-1)^q][1 + (-1)^r] + \right.$$

$$+ \left. [1 - (-1)^r]\lfloor(q-3)/2\rfloor\right\}(q-r)^{\lfloor r/2\rfloor-1} \quad (r > 1),$$

where r is the number of concatenated q-rings in a chain.

Here, we derive a special case of (5) for 5-gons:

$$\tilde{I}_{r+4} = 2^r + 2^{\lfloor r/2\rfloor} \quad (r \geq 0), \tag{6}$$

where the number $\tilde{I}_{r+4}$ is also the number of all caterpillars with $r + 4$ ($r \geq 0$) vertices, as showed Harary and Schwenk [22]. Recall that a *caterpillar* is a tree in which all vertices are within distance 1 of a central path. Thus, an enumeration task, as it was stated in [20], can also be reduced to enumeration of caterpillars; and in our case – to a special subclass thereof, which we cannot so far determine here. The very sequence of numbers $\tilde{I}_r$ ($r \geq 1$) for the first values of r is:

$$\tilde{I}_{r\geq1} : 1, 1, 1, 2, 3, 6, 10, 20, 36, 72, 136, 272, 528, 1056, 2080, 4160, 8256, \ldots \tag{7}$$

Another representation of the same result, as (6) is, may be given, due to a found recurrence:

$$\tilde{I}_{r+3} = 2(\tilde{I}_{r+2} + \tilde{I}_{r+1}) - 4\tilde{I}_r \quad (r \geq 3) \tag{8}$$

or in a shorter form:

$$\tilde{I}_{r+1} = 2\tilde{I}_r - [1 + (-1)^r]2^{\lfloor\frac{r}{2}\rfloor-3} \quad (r \geq 3). \tag{9}$$

Moreover, there exists a (reduced) transfer matrix corresponding to (8), whose characteristic polynomial is $x^3 - 2x^2 - 2x + 4 = 0$, with roots $x_1 = 2, x_2 = \sqrt{2}, x_3 = -\sqrt{2}$. Hence, as well as from (6), it follows that $\lim_{r\to\infty} \tilde{I}_{r+1}/\tilde{I}_r = 2$. Other versions are: $\lim_{r\to\infty} \tilde{I}_{r+4}/2^r = 1$ or $\lim_{r\to\infty} \tilde{I}_r/2^r = 1/16$.

Without symmetry considerations applied in [20], the respective sequence of numbers for $q = 5$ should correspond to a geometric progression; $1, 2, 4, \ldots$, with the r-th member equal to 2^{r-1} ($r \geq 1$). After comparing the members of both series (of that in [20] and of the last geometric one), it is clear that all symmetry-nonequivalent isomers of pentagonal chains of an essentially long length r comprise in limit just $1/8$ of all such chains produced by both the 'left' and 'right' attachments of the next pentagons. An analysis [20] of all possible symmetry groups of chains gives $D_{2h}, C_{2h}, C_{2v}, C_s$ ($|D_{2h}| = 8; |C_{2h}| = |C_{2v}| = 4; |C_s| = 2$). Here, the highest possible symmetry of a chain obeys the first group D_{2h} of order 8. Therefore, each nonsymmetric chain implies the existence of seven other chains which are symmetry-equivalent to it (or totally 8 such chains). Hence, it immediately follows a general conclusion that almost all chains are in

mass asymmetric ones. Only so the share of symmetry nonequivalent chains may comprise in mass 1/8 of the total amount of chains produced by the process. This may be used later.

In order to proceed, we need to introduce some notions. A *polygon P* is usually defined as a collection of n points $p_1, p_2, \ldots, p_n$ and n edges $p_1 p_2, p_2 p_3, \ldots, p_{n-1} p_n, p_n p_1$ such that no pair of nonconsecutive edges share a point. Let $p_1, p_2, \ldots, p_k$ be path or chain. A chain is called *monotone with respect to a line L* (p. 14 in [23]) if the projections of $p_1, p_2, \ldots, p_k$ onto L are ordered the same as in the chain; that is, there is no "doubling back" in the projection as the chain is traversed. The chain is *monotone* (p. 14 in [23]) if it is monotone with respect to at least one line.

Utilizing the above definition, we conclude that our target set of nonselfcrossing chains contains all monotone chains. We could enumerate the latter using a reduced transfer matrix which was constructed stepwise as follows. First we construct the entire transfer matrix whose weight entries are $\exp(v_i)$ ($i \in [1, 10]$), where shift vectors v_i's were introduced earlier in the text (see (1)). This basic matrix is relevant to generating all possible chains and/or annuli, without taking account symmetry. It is a weighted adjacency matrix of a (weighted) 10-cycle:

$$
\begin{bmatrix}
0 & \exp(v_2) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \exp(v_{10}) \\
\exp(v_1) & 0 & \exp(v_3) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \exp(v_2) & 0 & \exp(v_4) & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \exp(v_3) & 0 & \exp(v_5) & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \exp(v_4) & 0 & \exp(v_6) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \exp(v_5) & 0 & \exp(v_7) & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \exp(v_6) & 0 & \exp(v_8) & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \exp(v_7) & 0 & \exp(v_9) & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \exp(v_8) & 0 & \exp(v_{10}) \\
\exp(v_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \exp(v_9) & 0
\end{bmatrix}
\tag{10}
$$

The above matrix describes the general process which ensures correct attachment of the next pentagon, disregarding occurring selfcrossing of a chain. Its nonzero entries are functional weights which allow to watch the sum vector of shifts, corresponding to a total shift from the origin to the center of the last attached pentagon. When such a sum vector equals $v^* = (0, 0)$, the process returns us to the very first pentagon, thus indicating a closed walk – in particular, indicating a built annulus, but admitting in general multiple returns to the same point. But if we numerically sum the x- and y-coordinates, we obtain a zero sum of both mixed coordinates of all shift vector every time that the center of the last pentagon falls into line $y = -x$. Here, we need to turn from the above matrix to its two reductions.

The first reduced matrix takes into account the notion of the monotonicity of a chain. Accordingly, we nullify all entries which correspond to shift vectors v_i ($i \in [1, 10]$) having a nonnegative value of the x-coordinate; from (1), it follows that not nullified entries remain

those equal to $\exp(v_1), \exp(v_2), \exp(v_8), \exp(v_9), \exp(v_{10})$ (where these previously were above). We obtain a weighted adjacency matrix of the union of (i) a nonoriented path on 5 vertices, with two affluent arcs attached to its end points (that is, a mixed weakly connected graph on 7 vertices), and (ii) 3 isolated vertices:

$$
\begin{bmatrix}
0 & \exp(v_2) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \exp(v_{10}) \\
\exp(v_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \exp(v_2) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \exp(v_8) & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \exp(v_9) & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \exp(v_8) & 0 & \exp(v_{10}) \\
\exp(v_1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \exp(v_9) & 0
\end{bmatrix}
\tag{11}
$$

The second matrix excludes selfcrossings of a chain, while all nullified mixed coordinates (that is, $x + y = 0$) do not anymore correspond to cyclic constructions; the process generates only chains, and there is no even need in knowing the coordinates of the last pentagon's center, as such. Therefore, we reduce the second matrix to another one where all nonzero entries of the preceding one are replaced by 1's. Thereby we have (12).

$$
\begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{bmatrix}
\tag{12}
$$

The characteristic polynomial of the third matrix is $x^{10} - 4x^8 + 3x^6 = 0$, with nonzero roots $\pm\sqrt{3}, \pm 1$ and six 0's. Thus, we have the following recurrence:

$$
J_{r+4} = 4J_{r+2} - 3J_n \quad (r \geq 2),
\tag{13}
$$

with the first numbers

$$
J_{r \geq 1} : 1, 1, 1, 2, 3, 5, 9, 14, 27, 41, 81, 122, 243, 365, 729, 1094, \ldots .
\tag{14}
$$

It is easy to see shorter recurrences:

$$J_{r+2} = 3J_r - \frac{1}{2}[1 + (-1)^r] \quad (r \geq 2) \tag{15}$$

and (partial)

$$J_{2r+1} = 2J_{2r} - 1 \quad (r \geq 1). \tag{16}$$

From (15), in particular, it follows that $\lim_{r\to\infty} J_{r+2}/J_r = 3 < \lim_{r\to\infty} \tilde{I}_{r+2}/\tilde{I}_r = 4$. Thus, (13–16) also demonstrates an exponential growth but $\lim_{r\to\infty} J_r/\tilde{I}_r = 0$; therefore, the share of nonselfcrossing chains tends to 0, as r tends to ∞. In perspective, we are targeted at obtaining the numbers $\mathfrak{C}_r$ ($r \geq 1$) of all r-pentagon chains from $\mathfrak{C}$, for which $J_r \leq \mathfrak{C}_r \leq \tilde{I}_r$ and $3 \leq \lim_{r\to\infty} \mathfrak{C}_{r+2}/\mathfrak{C}_r \leq 4$ hold.

As it was stated, our target is, in particular, plane chains of pentagons. The first 9 numbers of such chains (but not the 10-th *et seq.*) are given in (7). Here, we add some subsequent numbers, which were found with a computer program:

$$C_{\geq 10} : 71, 134, 267, 515, 1021, 1992, 3954, 7763, 15354, 30211, 59722, 117633, 232102,$$
$$457057, 901140, 1774114, \ldots \quad (C_r \equiv \tilde{I}_r; r \in [1, 9]). \tag{17}$$

The numbers C_r from (7) and (17) allow to build a system of 13 simultaneous linear equations, in order to look for a linear recurrence, as was used above. However, in the last instance, no linear recurrence has been found. The only evident conclusion is that

$$C_r = a(r)\tilde{I}_r \quad (r \in [1, \infty)), \tag{18}$$

where $a(r) = C_r/\tilde{I}_r$ ($0 \leq a(r) \leq 1$) determines the share of plane pentagonal chains among all pentagonal chains (as the latter were determined in [20]). Let

$$a(r) = a(x)|_{x=r} \ (r \in [1, \infty)), \tag{19}$$

where $a(x)$ is a continuous nonincreasing function in x, which is assumed here to be smooth enough. We also use here its truncated inverse:

$$a^{-1}(x) = 1/a(x)\big|_{x \geq 10}; \tag{20}$$

$$a^{-1}(r) = \tilde{I}_r/C_r \quad (r \in [10, \infty)). \tag{21}$$

The following technical lemma plays an important role in our text:

Lemma 4. *Let $a^{-1}(x)$ be as above. Then, $a^{-1}(x)$ is a function rational under all natural values of x, which is expanded in a polynomial $a^{-1}(x) = \sum_{s=0}^{s_{max}} a_s x^{\alpha_s}$ in x, whose all coefficients a_s are rational and all powers α_s of x are natural numbers.*

Proof. Since a value of $a^{-1}(x)$ is a rational number under each $x = r$ ($r \in [1, \infty)$), the powers α_s in the expansion are all natural numbers – otherwise, such an expansion would give irrational values at some natural values of x, which is prohibited by the definition of $a^{-1}(x)$. In order to avoid any irrationality of $a^{-1}(x)$ under all natural x, the expansion cannot have irrational coefficients and be an infinite series, either. Hence, the overall proof follows. $\qquad\square$

Using the sequence (17) of known to us numbers and taking into account Lemma 4, we found the following approximate expansion:

$$a^{-1}(x) \approx 0.9915612179 - 0.003054367868x + 0.0004279348734x^2. \tag{22}$$

Hence, we obtain:

$$a(x) \approx (0.9915612179 - 0.003054367868x + 0.0004279348734x^2)^{-1}. \tag{23}$$

Though (23) seems a too simple approximation for $a(x)$, its worst interpolation value for C_r gives an error $< 1.5\%$ for $r \in [1, 9]$ and $\sim 1\%$ for $r \in [10, 25]$, then, $\sim 0.2\%$ for $r = 15$, $\sim 0.1\%$ for $r = 19$, and even $< 0.1\%$ for $r \in [19, 25]$. Since at some intermediate step of manipulations the interpolation was due to a straight-line plot, one may predict that (23) may allow a further acceptable extrapolation for some consecutive $r > 25$. All this allows to state:

Proposition 5. *Let* $a(x) \approx (0.9915612179 - 0.003054367868x + 0.0004279348734x^2)^{-1}$ *be an interpolation function for* $a(x)$ *(*$x \in [1, 25]$*). Then,*

$$C_r \approx \tilde{I}_r \left(0.9915612179 - 0.003054367868r + 0.0004279348734r^2 \right)^{-1} \quad (r \in [1, 25]). \tag{24}$$

Since (24) depends on exact numbers $\tilde{I}_r$ ($r \in \mathbb{N} \setminus 0$), this does not allow to estimate the numbers C_r for values r which are beyond our ability to calculate respective values of $\tilde{I}_r$. Therefore, it is worth giving here a purely asymptotic "independent" estimation. Earlier (see after (9)), there was established that $\lim_{r \to \infty} \tilde{I}_r / 2^r = 1/16$. Can this estimation be applied for finite values of r? *E. g.*, under $r = 25$, $2^{25}/16 = 2^{21} \approx 2097152$, while an exact number $\tilde{I}_{25} = 2098176$, which indicates an error $< 0.05\%$. From the last calculation, it follows that $C_{25} \approx 1773248$; in comparison with the exact value $C_{25} = 1774114$, this gives an error also $< 0.05\%$. Therefore, as a technical corollary of Proposition 5, we propose:

Corollary 5.1. *There exists the following approximation:*

$$C_r / 2^r \approx (15.86497949 - 0.04886988589r + 0.006846957974r^2)^{-1} \quad (r \in [1, 25]). \tag{25}$$

Another corollary is:

Corollary 5.2. *There exists the following approximate limit:*

$$\lim_{r \to \infty} \frac{C_r r^2}{2^r} \approx 146.0502611. \tag{26}$$

Corollary 5.2 (motivated by the above observations) gives an idea of the following conjecture:

Conjecture 6. *There exists the following exact limit:*

$$\lim_{r \to \infty} \frac{C_r r^2}{2^r} = const. \tag{27}$$

11

Acknowledgments

References

[1] V. R. Rosenfeld and D. J. Klein, "Enumeration of substitutional isomers with restrictive mutual positions of ligands. I. Overall counts", *J. Math. Chem. 51 (1)* (2013) 21–37.

[2] V. R. Rosenfeld and D. J. Klein, "Enumeration of substitutional isomers with restrictive mutual positions of ligands. II. Counts with restrictions on subsymmetry", *J. Math. Chem. 51 (1)* (2013) 239–264.

[3] V. R. Rosenfeld, J. M. Oliva, and D. J. Klein, "Chain carborane molecules", *Monatsh. Chemie 143 (3)* (2012) 361–364.

[4] V. R. Rosenfeld, D. J. Klein, and J. M. Oliva, "Enumeration of polycarborane isomers: Especially dicarboranes", *J. Math. Chem. 50 (7)* (2012) 2012–2022.

[5] J. M. Oliva, J. Rué, D. Hnyk, J. D. Kennedy, and V. R. Rosenfeld, "Borane polyhedra as building blocks for unknown but potentially isolatable new molecules – Extensions based on computations of the known $B_{18}H_{22}$ isomers, *Croat. Chem. Acta (86) (4)* (2013) 86 485–494.

[6] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*, Wiley, Chichester, England, 2004.

[7] D. Hnyk, M. McKee (Eds.), *Boron: The Fifth Element, Challenges and Advances in Computational Chemistry and Physics*, Vol. *20*, Series Editor: J. Leszczynski, Springer, 2015.

[8] R. N. Grimes, *Carboranes*, Academic Press, London, 2011.

[9] M. Liśkiewicz, M. Ogihara, and S. Toda, "The complexity of counting self-avoiding walks in subgraphs of two-dimensional grids and hypercubes", *Theoret. Comput. Sci. 304 (1–3)* (2003) 129–156.

[10] A. J. Guttmann, "Self-avoiding walks and polygons – An overview", *Asia Pac. Math. Newsl. 2 (4)* (2012) 1–10.

[11] A. De Luca and S. Varricchio, "Factorial languages whose growth function is quadratically upper bounded", *Inform. Process. Lett. 30* (1989) 283–288.

[12] A. De Luca and S. Varricchio, "Some combinatorial properties of the Thue-Morse sequence and a problem in semigroups", *Theoret. Comput. Sci. 63* (1989) 333–348.

[13] A. De Luca A. and S. Varricchio, "A combinatorial theorem on *p*-power-free words and an application to semigroups", *RAIRO Inform. Théor. Appl. 24 (3)* (1990) 205–228.

[14] See: *http://en.wikipedia.org/wiki/Quasicrystal*.

[15] V. I. Arnold, *Huygens and Barrow, Newton and Hooke – pioneers in mathematical analysis and catastrophe theory from evolvents to quasicrystals*, Birkhäuser Verlag, Basel, Boston, Berlin, 1990. The 1st edition in: Series The Modern Mathematics for Students", Nauka, Moscow, 1989 (in Russian).

[16] F. Harary and R. C. Read, "The enumeration of tree-like polyhexes", *Proc. Edinb. Math. Soc. (Series 2)*, *17 (1)* (1970) 1–13.

[17] S. B. Elk, "An algorithm to identify and count coplanar isomeric molecules formed by the linear fusion of cyclopentane modules", *J. Chem. Inf. Comput. Sci. 27 (2)* (1987) 67–69.

[18] J. Brunvoll, B. N. Cyvin, and S. J. Cyvin, "Enumeration and classification of coronoid hydrocarbons", *J. Chem. Inf. Comput. Sci. 27 (1)* (1987) 14–21.

[19] S. J. Cyvin, J. Brunvoll, and B. N. Cyvin, "Enumeration of unbranched catacondensed polygons systems: General solution for two kinds of polygons", *J. Chem. Inf. Comput. Sci. 37 (3)* (1997) 460–465.

[20] B. N. Cyvin, S. J. Cyvin, J. Brunvoll, and A. A. Dobrynin, "Enumeration of unbranched catacondensed systems of congruent polygons", *Vychisl. Sistemy* **155** (1996) 3–14.

[21] G. Brinkmann, A. A. Dobrynin, A. Krause, "Fast generation of polycyclic chains with arbitrary ring sizes", *MATCH Commun. Math. Comput. Chem.* **41** (2000) 137–144.

[22] F. Harary and A. J. Schwenk, "The number of caterpillars", *Discrete Math.* **6** (1973) 359–365.

[23] J. O'Rourke, *Art of Gallery Theorems and Algorithms*, Oxford University Press, New York, Oxford, 1987.