

# On topological RNA interaction structures

Jing Qin<sup>a,b</sup>, Christian M. Reidys<sup>c,\*</sup>

<sup>a</sup>Max Planck Institute for Mathematics in the Sciences,  
Inselstraße 22, D-04103 Leipzig, Germany

<sup>b</sup>Bioinformatics Group, Department of Computer Science,  
Interdisciplinary Center for Bioinformatics,

University of Leipzig, Härtelstrasse 16-18, D-04107 Leipzig, Germany.

<sup>c</sup>University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

---

## Abstract

Recently a folding algorithm of topological RNA pseudoknot structures has been presented [24]. This algorithm folds single stranded  $\gamma$ -structures, i.e. RNA structures composed by distinct motifs of bounded topological genus. In this paper, we study the two backbone analogue of  $\gamma$ -structures: the RNA  $\gamma$ -interaction structures. These are RNA-RNA interaction structures that are constructed by a finite number of building blocks over two and one backbone having genus at most  $\gamma$ . Properties of  $\gamma$ -interaction structures are of practical interest since they are the targets of topological interaction structure folding algorithms. We show that the generating function of  $\gamma$ -interaction structures is algebraic, which implies that the numbers of interaction structures can be computed recursively. We furthermore obtain simple asymptotic formulas for 0- and 1-interaction structures. The simplest class are the 0-interaction structures, which represent the two backbone analogue of secondary structures.

*Keywords:* RNA-RNA interaction,  $\gamma$ -interaction structure, Shape, Symbolic enumeration, Singularity analysis, RNA secondary structure

*2000 MSC:* 05A16, 92E10

---

## 1. Introduction

The fact that nucleic acids are capable of recognizing other nucleic acids with very high specificity by means of complementary base pairing is the basis of many basic laboratory techniques starting from designed RNA and DNA primers. Not surprisingly, direct interactions of nucleic acids are also an important mechanism in gene regulation. The most prominent examples are certainly microRNAs and siRNAs which bind their targets by forming more or less perfect helical regions. More complex structure of the RNA-RNA binding are the

---

\*Corresponding author. [duck@santafe.edu](mailto:duck@santafe.edu)

Email address: [duck@santafe.edu](mailto:duck@santafe.edu) (Christian M. Reidys)

rule for snoRNAs and in particular for a wide variety of small bacterial sRNAs. RNA-RNA binding is not limited to small structured RNAs addressing protein-coding mRNAs. In many cases, the regulatory effect of bacterial sRNA-mRNA interactions can be understood in terms of the underlying interaction structure. Namely, depending on the location of the binding site conformational changes can lead to the sequestration or exposure of the Shine-Dalgarno sequence and hence disable or enable translation.

Interactions between nucleic acids can be understood to a good approximation at the level of secondary structures (patterns of base pairs) or extended secondary structures. From a technical point of view, therefore the analysis, prediction, and comparison of RNA-RNA and RNA-DNA interaction structures is formalized as a suite of related combinatorial (optimization) problems that make genome-wide studies computationally feasible and sets RNA bioinformatics apart from other areas of structural biology in which three-dimensional shapes and atomic detail cannot be neglected as easily. Despite this favorable starting point and the success of the combinatorial approach to structures of individual nucleic acid molecules, we nevertheless lack both theoretical insights and practical predictive power in the realm of interactions.

The simplest approach for folding RNA-RNA interaction structures concatenates two (or more) interacting sequences one after another remembering the specific merge point (cut-point) and then employs the standard secondary structure folding algorithm on a single strand. This approach falls short predicting many genuine features such as kissing-hairpin loops. The paradigm of concatenation has also been generalized to include cross-serial interactions [25]. The resulting model, however, still does not generate all relevant interaction structures [5]. An alternative is to neglect any internal base pairings in either strand, i.e., to compute the minimum free energy (MFE) secondary structure of hybridization of otherwise unstructured RNAs. [16, 15] and [6] restrict interactions to a single interval that remains unpaired in the secondary structure for each partner.

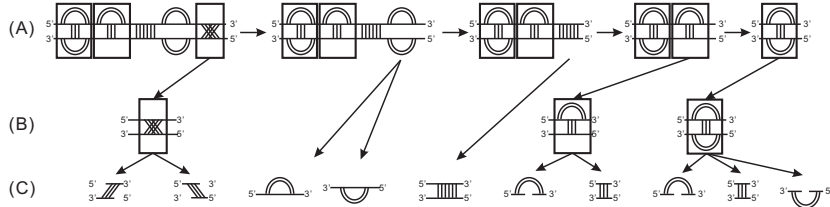
A different approach was taken independently by [23] and [1] who proposed MFE folding algorithms for predicting the *AP-structure* of two interacting RNA molecules. In this model, the intramolecular structures of each partner are pseudoknot-free, the intermolecular binding pairs are non-crossing, and there is an ad-hoc exclusion of so called “zig-zag” motifs. The RNA-RNA interaction structures of [11, 1, 9, 3] have the following features: (a) when drawing the two backbones on top of each other, all base pairs are non-crossing, i.e., no pseudoknots formed by internal or external arcs are allowed and (b) zig-zag motifs are disallowed.

In this paper we consider a novel filtration of RNA-RNA interaction structures. This filtration is based on the topological genus of certain motifs contained therein. The classification and expansion of pseudoknotted RNA structures in terms of the topological genus of an associated fatgraph or double line graph were first proposed by [18] and [4], although fatgraphs were applied to RNA secondary structures already by [22] and [19]. The first enumerative results initiated by [18, 26] are an application of matrix models from theoretical

physics. Genus as well as other topological invariants of fatgraphs were introduced and studied as descriptors of proteins in [21].

In [24], RNA  $\gamma$ -structures were introduced. The idea of  $\gamma$ -structures is to consider RNA structures obtained by building blocks, having topological genus at most  $\gamma$ . From this point of view, RNA secondary structures can be seen to be obtained by the building block that is simply an arc by means of concatenation and nesting. Accordingly, secondary structures are equivalent to 0-structures over one backbone. We call these blocks in case of genus greater or equal to 1 irreducible shadows. In this sense, topological genus controls the complexity of the cross-serial interactions expressed in these irreducible shadows. In [24] but already implicit in [27], it was proved that for *any* genus, there are only *finitely* many irreducible shadows.

RNA interaction structures of fixed topological genus [2] have, as their one backbone counterparts, only finitely many irreducible shadows. Following the identification of these [2] develops the concept of RNA  $\gamma$ -interaction structures. Note that the genus of  $\gamma$ -interaction structures is not fixed. For the simplest class, the 0-interaction structures, we already have in anticipation of folding algorithms an unambiguous grammars [2], see Fig. 1. While genus zero struc-

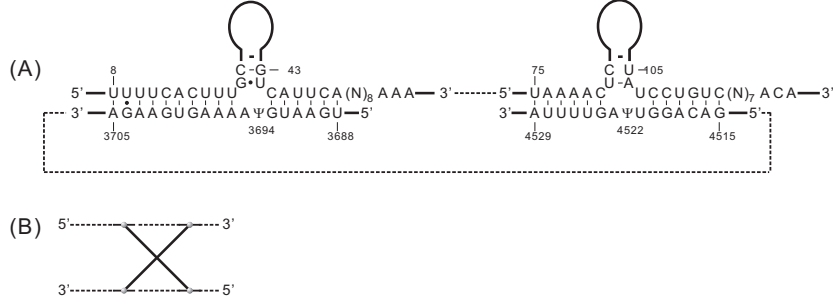


**Fig. 1:** An unambiguous grammar of 0-interaction structures, see [2] for details.

tures are exactly RNA secondary structure that only contain noncrossing arcs, genus zero interaction structures exhibit crossings, see Fig. 1. Such cross serial interactions however can only occur for exterior arcs, i.e. arcs that connect the two backbones.

The idea of topological RNA interaction structures is a topic of considerable practical interest, since RNA interaction structures that do not belong to the AP-class exist. For instance, the integral RNA (hTER) of the human telomerase ribonucleoprotein has a conserved secondary structure that contains a potential pseudoknot [13]. There is evidence that the two conserved complementary sequences of one stem of the hTER pseudoknot domain can pair intermolecularly in vitro, and that formation of this stem as part of a novel “transpseudoknot” is required for the telomerase to be active in its dimeric form, see Fig. 2.

Our main result is the computation of the generating function of  $\gamma$ -interaction structures. To prove this, we employ  $\gamma$ -structures over one backbone without isolated points ( $\gamma$ -matchings), whose generating function,  $\mathbf{H}_\gamma(u)$ , was derived in [8]. One key ingredient in the computation is a new polynomial,  $\mathbf{I}_{g,1}(z)$  whose coefficients are the numbers of irreducible shadows of genus  $g$  over one back-



**Fig. 2:** (A) Homo sapiens ACA27 snoRNA. This H/ACA box RNA was cloned [12, 17] from a HeLa cell extract immunoprecipitated with an anti-GAR1 antibody. (B) The structure contains two crossing hybrids, which cannot be found in AP-structures.

bone. We develop here the two backbone analogues of these polynomials and use them as cornerstones of our combinatorial constructions. In Theorem 3 we prove that the generating function of  $\gamma$ -interaction structures is given by

$$\mathbf{Q}_\gamma(u) = \frac{\mathbf{H}_\gamma^2(u) \cdot \left( u\mathbf{H}_\gamma^2(u) + \sum_{g \leq \gamma} \mathbf{I}_{g,2} \left( \frac{u \cdot \mathbf{H}_\gamma^2(u)}{1-u} \right) \right)}{1 - u\mathbf{H}_\gamma^2(u) - \sum_{g \leq \gamma} \mathbf{I}_{g,2} \left( \frac{u \cdot \mathbf{H}_\gamma^2(u)}{1-u} \right)},$$

where the  $\mathbf{I}_{g,2}(u)$  are the above mentioned polynomials whose coefficients count certain diagrams of genus  $g$  over two backbones.

## 2. Background and notations

### 2.1. Diagrams

A diagram is a labeled graph over the vertex set  $[n] = \{1, \dots, n\}$  in which each vertex has degree  $\leq 3$ , represented by drawing its vertices in a horizontal line and its edges  $(i, j)$ , where  $i < j$ , in the upper half-plane. A backbone is a sequence of consecutive integers contained in  $[n]$ . A diagram over  $b$  backbones is a diagram together with a partition of  $[n]$  into  $b$  backbones.

An interval  $[i, i + 1]$  is a gap interval if there exists a pair of subsequent backbones  $B_1$  and  $B_2$  such that  $i(j)$  is the rightmost(leftmost) vertex of  $B_1(B_2)$ . The vertex  $i$  is referred to as cut vertex. There are of course exactly  $(b - 1)$  such cut vertices.

We call backbone edges  $B$ -arcs and any other edge simply an arc. We shall distinguish exterior and interior arcs, where the former connect different backbones and the latter are within a specific one. Diagrams over multiple backbones without exterior arcs are simply disjoint unions of diagrams over one backbone.

The vertices and arcs of a diagram correspond to nucleotides and base pairs, respectively. For a diagram over  $b$  backbones, the leftmost vertex of each backbone denotes the 5' end of the RNA sequence, while the rightmost vertex denotes the 3' end.

The particular case  $b = 2$  is referred to as RNA interaction structures [10, 11]. As mentioned above, interaction structures are oftentimes represented alternatively by drawing the two backbones  $R$  and  $S$  on top of each other, indexing the vertices  $R_1$  to be the 5' end of  $R$  and  $S_1$  to be the 3' of  $S$ . In the following we shall denote the set of diagrams over one and two backbones by  $\mathbb{D}$  and  $\mathbb{E}$ , respectively.

A vertex  $i$  is isolated if it is not incident to any arc (except of backbone arcs). A diagram is connected if and only if it is connected as a combinatorial graph (i.e. employing arcs as well as backbone arcs). A diagram that does not contain any isolated vertices is called a matching, see Fig. 3.

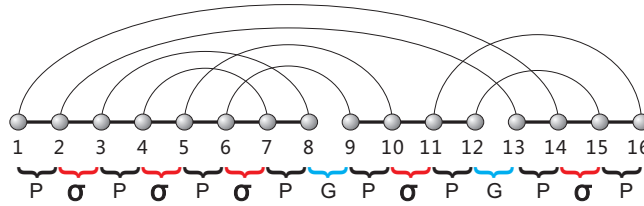
An interior stack of length  $\tau$  on  $[i, j] = \{i, i+1, \dots, j-1, j\}$  denoted by  $ST_{i,j}^\tau$  is a maximal set (with respect to set inclusion) of exactly  $\tau$  “parallel” interior arcs, namely

$$((i, j), (i+1, j-1), \dots, (i+(\tau-1), j-(\tau-1))).$$

Furthermore, an interior stack is  $\tau$ -canonical if it contains at least  $\tau$  interior arcs. Exterior stacks on  $[i, j]$  ( $SE_{i,j}$ ) and  $\tau$ -canonical exterior stacks,  $SE_{i,j}^\tau$ , can be similarly defined. We oftentimes simply write stack,  $ST_{i,j}^k$ , on  $[i, j]$  of length  $k$  if its particular type is immaterial. A stack  $ST_{i,j}^k$  on  $[i, j]$  of length  $k$  naturally induces  $(k-1)$  pairs of intervals of the form

$$([i+\ell, i+\ell+1], [j-\ell-1, j-\ell]) \quad \text{where } 0 \leq \ell \leq k-2.$$

Any of these  $2(k-1)$  intervals is referred to as a  $P$ -interval. Any interval other than a gap or  $P$ -interval is called a  $\sigma$  interval. Clearly, a diagram over  $[n]$ , contains  $(n-1)$  intervals and we distinguish the three types: gap intervals,  $P$ -intervals and  $\sigma$ -interval, see Fig. 3.

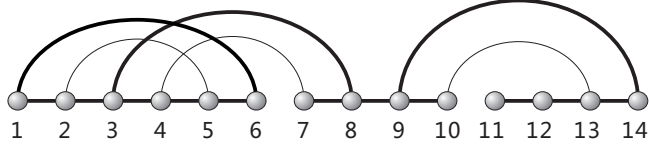


**Fig. 3:** The three types of intervals: gap intervals (**G**),  $\sigma$ -intervals ( $\sigma$ ) and  $P$ -intervals (**P**).

Let  $\prec$  be the partial order on arcs given by  $(i, j) \prec (i', j')$  if and only if  $i' \leq i$  and  $j \leq j'$ . Any diagram has a unique set of maximal arcs, cf. Fig. 4.

## 2.2. From diagrams to topological surfaces

One approach for deriving meaningful filtration of RNA structure is to pass from diagrams to topological surfaces [14]. It is natural to make this transition

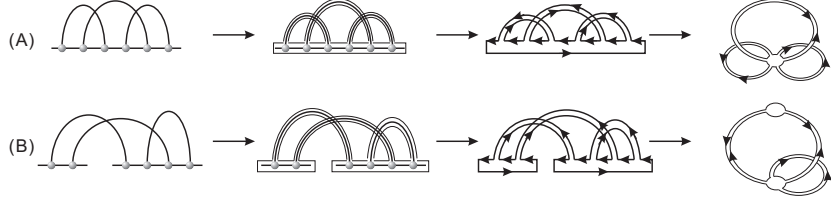


**Fig. 4:** Maximal arcs: the maximal arcs (1,6), (3,8) and (9,14) of a diagram of [14] (bold).

from combinatorics to topology via fatgraphs [21, 20]. A fatgraph  $\mathbb{G}$ , sometimes also called “ribbon graph” or “map”, is a graph  $G$  together with a collection of cyclic orderings, called a fattening, one such ordering on the half-edges incident on each vertex. Each fatgraph  $\mathbb{G}$  determines an oriented surface  $F(\mathbb{G})$  as follows: let  $V(G)$  be the set of  $G$ -vertices and  $E(G)$  be the set of  $G$ -edges. For each  $v \in V(G)$ , consider an oriented surface isomorphic to a polygon  $P_v$  with  $2k$  sides containing  $v$  in its interior where  $k$  is the valence of  $v$ . The incident edges of  $v$  are also incident to a univalent vertex contained in alternating sides of  $P_v$ , which are identified with the incident half-edges in the natural way so that the induced counter-clockwise cyclic ordering on the boundary of  $P_v$  agrees with the fattening of  $\mathbb{G}$  about  $v$ . The surface  $F(\mathbb{G})$  is the quotient of the disjoint union  $\sqcup_{v \in V(G)} P_v$ , where the frontier edges, which are oriented with the polygons on their left, are identified by an orientation-reversing homeomorphism if the corresponding half-edges lie in a common edge of  $G$ . This defines the oriented surface  $F(\mathbb{G})$ , which is connected if and only if  $G$  is and is uniquely determined in this case by its genus  $g = g(G) \geq 0$  and number  $r = r(G) \geq 1$  of boundary components. Since  $F(\mathbb{G})$  contains  $G$  as a deformation retract, they share the Euler characteristic  $v - e$ , and the genus of  $F(\mathbb{G})$  is given by  $2 - 2g - r = v - e$ .

For an RNA diagram, we may draw a representation as usual so that the backbone is a horizontal line oriented from left to right, and the arcs lie in the upper half-plane. This determines a unique fattening on any diagram, cf. the leftmost two panels in Fig. 5 for the fatgraph and its corresponding surface. Each boundary component of  $F(\mathbb{G})$  determines a closed edge-path or cycle on  $G$ , oriented with the surface lying on its left. In particular, a neighborhood of each edge inherits an orientation from that of  $F(\mathbb{G})$  which combine to give the oriented cycles as depicted in the third panel of Fig. 5. Without affecting topological type of the constructed surface, one may collapse each backbone to a single vertex with the induced fattening called the polygonal model of the RNA, as illustrated in the rightmost panels in Fig. 5. It is the orientation of each backbone from the 5' end to the 3' end that allows us to transform the fatgraph of an RNA-structure or RNA-interaction into a fatgraph with one or two vertices.

This backbone-collapse preserves orientation, Euler characteristic and genus by construction. It is reversible by inflating each vertex to form a backbone. Using the collapsed fatgraph representation, we see that for a connected diagram over  $b$  backbones, the genus  $g$  of the surface (with boundary) is determined by



**Fig. 5:** (A) The fatgraph of a diagram and its reduction to a single vertex. Contracting the backbone of a diagram into a single vertex decreases the length of the boundary components and preserves the genus. (B) Inflation of edges and vertices to ribbons and discs, as well as walking along the boundary components. Here we have six vertices, seven edges and one boundary component. The corresponding surface has Euler characteristic  $\chi = v - e = -1$  and  $g = 1$ . At the last step, we collapse each backbone into a single disc again preserving genus. The backbone of the polymer can be recovered by inflating each disk to a backbone segment.

the number  $n$  of arcs as well as the number  $r$  of boundary components, namely,  $2 - 2g - r = v - e = b - n$ , cf. Fig. 5.

Diagrams over one and two backbones are related by gluing, i.e., we have the mapping

$$\alpha: \mathbb{E} \rightarrow \mathbb{D},$$

where  $\alpha(E)$  is obtained by keeping all arcs in  $E$  and connecting the 3' end of  $R$  and the 5' end of  $S$ , cf. Fig. 6 (A).

In addition to gluing, there is another mapping: given two diagrams over two backbones,  $E_1, E_2 \in \mathbb{E}$  we can insert  $E_2$  into the gap of  $E_1$  by concatenating the backbones  $R_2$  and  $R_1$  and  $S_1$  and  $S_2$  preserving orientation.; see Figure 6 (B). This composition is again a diagram over two backbones denoted by  $E_1 \bullet E_2$ :

$$\mu: \mathbb{E} \times \mathbb{E} \longrightarrow \mathbb{E}, \quad \mu(E_1, E_2) = E_1 \bullet E_2. \quad (2.1)$$

It is straightforward to see that  $\bullet$  is an associative product with unit given by the diagram over two empty backbones. The product  $\bullet$  is not commutative.

### 2.3. Shadows

A shadow is a diagram with no non-crossing arcs or isolated vertices in which each stack has size one. The shadow of a diagram is obtained by removing all non-crossing arcs, deleting all isolated vertices and collapsing each induced stack to a single arc cf. Fig. 7. We shall denote the shadow of a diagram  $X$  by  $sd(X)$ , note that  $sd^2(X) = sd(X)$ . Projecting into the shadow does not affect genus, i.e.,  $g(X) = g(sd(X))$ . In case there are no crossing arcs,  $sd(X)$  becomes an empty diagram on the same number of backbones as  $X$  as in Figure 7 (C). By definition, any empty backbone contributes one boundary component. For example, for a diagram  $X$  over  $b$  backbones that contains no crossing arcs,  $sd(X)$  is a sequence of  $b$  empty backbones with  $b$  boundary components.

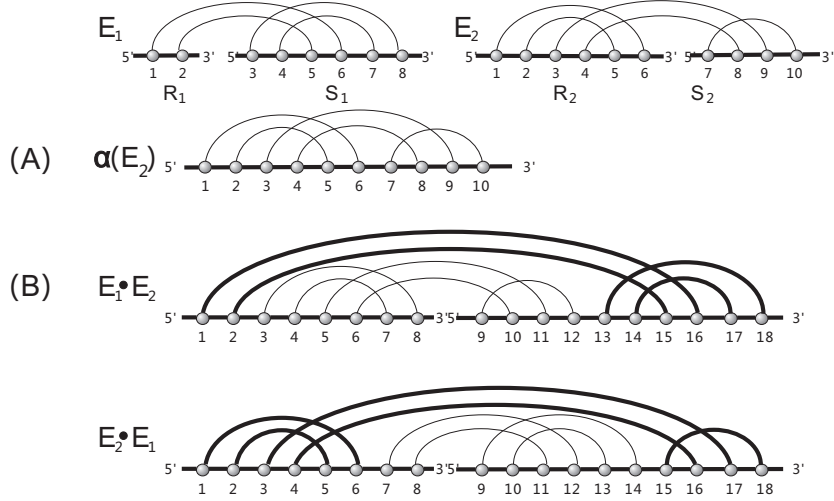


Fig. 6: (A) Mapping a diagram over two backbones into a diagram over one backbone by gluing. (B) the product  $\bullet$ .

**Theorem 1.** [2]. *A shadow over two backbones has the following properties:*  
**(a)** *a shadow of genus  $g \geq 1$  over two backbones contains at least  $(2g + 1)$  and at most  $6(g + 1) - 2$  arcs; a shadow of genus 0 has at least 2 and at most 4 arcs. in particular, the set of such shadows is finite;*  
**(b)** *for any  $(2g + 1) \leq \ell \leq 6(g + 1) - 2$  in case of  $g \geq 1$  and  $2 \leq \ell \leq 4$  in case of  $g = 0$ , there exists some shadow over two backbones with genus  $g$  containing exactly  $\ell$  arcs.*

#### 2.4. Irreducibility

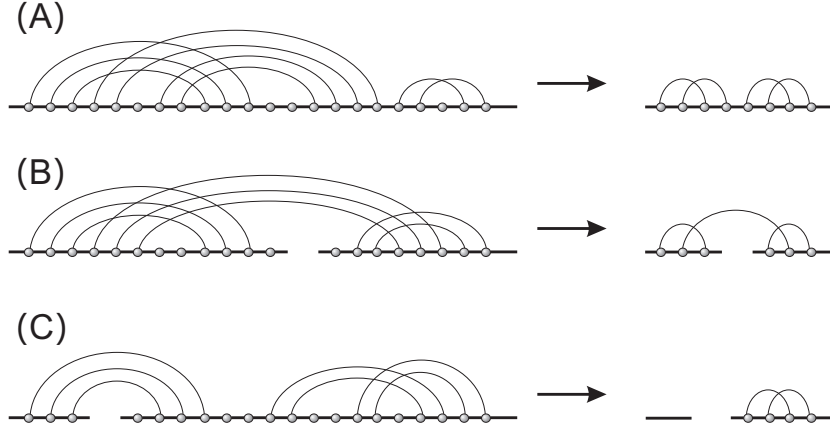
A diagram  $E$  over  $b$  backbones is called irreducible if and only if it is connected and for any two arcs,  $\alpha_1, \alpha_k$  contained in  $E$ , there exists a sequence of arcs  $(\alpha_1, \alpha_2, \dots, \alpha_{k-1}, \alpha_k)$  such that  $(\alpha_i, \alpha_{i+1})$  are crossing. As proved in [2], there is the following corollary of Theorem 1.

**Corollary 1.** *An irreducible shadow having genus  $g = 0$  over two backbones contains at least 2 and at most 4 arcs, and for and  $2 \leq \ell \leq 4$ , there exists an irreducible shadow of genus  $g = 0$  over two backbones having exactly  $\ell$  arcs. An irreducible shadow having genus  $g \geq 1$  has the following properties:*

- (a)** *every irreducible shadow with genus  $g$  over two backbones contains at least  $2g + 1$  and at most  $6(g + 1) - 2$  arcs;*
- (b)** *for arbitrary genus  $g$  and any  $2g + 1 \leq \ell \leq 6g - 2$ , there exists an irreducible shadow of genus  $g$  over one backbone having exactly  $\ell$  arcs.*

Let  $X$  be a diagram. We call  $S'$  an irreducible shadow of  $X$  (irreducible  $X$ -shadow) if and only if  $S'$  is an irreducible shadow and any arc in  $S'$  is contained





**Fig. 7:** Shadows: (A) a diagram over one backbone and its shadow, (B) a diagram over two backbones whose shadow is again a diagram over two backbones and (C) a diagram whose shadow with an empty backbone.

in  $X$ .  $S'$  is a  $(g, b, m)$ -shadow if  $S'$  is a diagram over  $b$  backbones having genus  $g$  and  $m$  arcs. The set of irreducible  $(g, b, m)$ -shadows is denoted by  $\mathcal{I}_{g,b,m}$ . Let  $\mathcal{I}_{g,b} = \cup_m \mathcal{I}_{g,b,m}$ .

According to Theorem 1, the generating function  $\mathbf{I}_{g,b}(u) = \sum \mathbf{i}_{g,b}(m)u^m$  of the combinatorial class  $\mathcal{I}_{g,b}$  is a polynomial. For  $0 \leq g \leq 1$  and  $1 \leq b \leq 2$ , we explicitly list the generating polynomials for  $\mathbf{I}_{g,b}(u)$  as follows.

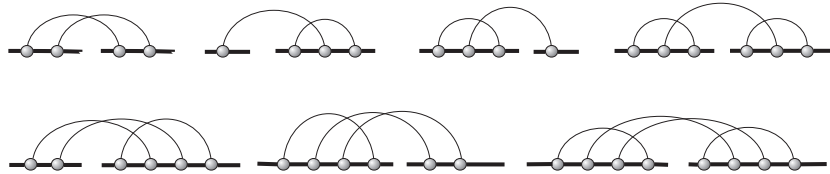
$$\mathbf{I}_{1,1}(z) = z^2 + 2z^3 + z^4,$$

$$\mathbf{I}_{2,1}(z) = 17z^4 + 160z^5 + 566z^6 + 1004z^7 + 961z^8 + 476z^9 + 96z^{10},$$

$$\mathbf{I}_{0,2}(u) = 3u^2 + 3u^3 + u^4,$$

$$\mathbf{I}_{1,2}(u) = 11u^3 + 137u^4 + 656u^5 + 1520u^6 + 1951u^7 + 1456u^8 + 572u^9 + 96u^{10}.$$

We display the seven irreducible shadows over two backbones of genus 0 in Fig. 8.



**Fig. 8:** There are exactly seven irreducible shadows over two backbones of genus 0.

### 2.5. $\gamma$ -structures

A diagram is a  $\gamma$ -structure if and only if it is connected and all its irreducible shadows have genus at most  $\gamma$ . A  $\gamma$ -matching is a  $\gamma$ -structure which is also a matching. The combinatorial classes of  $\gamma$ -matchings over one and two backbones are denoted by  $\mathcal{H}_\gamma$  and  $\mathcal{Q}_\gamma$  respectively. We also call the  $\gamma$ -structure over two backbones ( $\mathcal{Q}_\gamma$ )  $\gamma$ -interaction structures. The generating function of  $\mathcal{H}_\gamma$  denoted by  $\mathbf{H}_\gamma(u)$  has been derived in [8] as follows:

**Theorem 2.** *Let  $R = \mathbb{Z}[u]$ . Then  $\mathbf{H}_\gamma(u)$ , satisfies*

$$\mathbf{H}_\gamma(u)^{-1} = 1 - \left( u \mathbf{H}_\gamma(u) + \mathbf{H}_\gamma^{-1}(u) \sum_{g \leq \gamma} \mathbf{I}_{g,1} \left( \frac{u \mathbf{H}_\gamma^2(u)}{1 - u \mathbf{H}_\gamma^2(u)} \right) \right). \quad (2.2)$$

Furthermore, eq. (2.2) determines  $\mathbf{H}_\gamma(u)$  uniquely.

We denote the generating function of a set of diagrams  $\mathcal{K}$  by  $\mathbf{K}(u) = \sum_n \mathbf{k}(n)u^n$ , where  $\mathbf{k}(n)$  denotes the number of  $\mathcal{K}$ -diagrams that contain exactly  $n$  arcs. In our proofs we shall recruit the following combinatorial classes:  $\mathcal{E}$  (neutral class which include only one element of size 0),  $\mathcal{Z}$  (vertices) and  $\mathcal{R}$  (arcs).

### 3. $\gamma$ -matchings over two backbones

Let  $\mathcal{X}_\gamma$  denote the set of ordered pairs of  $\mathcal{H}_\gamma$  structures. The ordering is naturally induced by that of the respective backbones. Clearly, we have

$$\mathcal{X}_\gamma = \mathcal{H}_\gamma \times \mathcal{H}_\gamma. \quad (3.1)$$

Let next  $\mathcal{A}_\gamma$  denote the set of  $\gamma$ -matchings over two backbones that contains only exterior arcs that are noncrossing. Given a diagram  $s$  in  $\mathcal{A}_\gamma$ , let  $s'$  denote the diagram obtained by collapsing each stack in  $s$  to a single arc.

Let  $\mathbf{Level}(s)$  denote the number of exterior arcs  $s'$ . Since any  $\gamma$ -matching  $s$  over two backbones is connected, we have  $\mathbf{Level}(s) \geq 1$  for any  $s \in \mathcal{A}_\gamma$ . Let  $\mathbf{a}_\gamma(n)$  denote the number of  $\mathcal{A}_\gamma$ -matchings with  $n$  arcs.

**Lemma 1.** *The generating function  $\mathbf{A}_\gamma(u) = \sum_n \mathbf{a}_\gamma(n)u^n$  is given by*

$$\mathbf{A}_\gamma(u) = \frac{u \cdot \mathbf{H}_\gamma^4(u)}{1 - u \cdot \mathbf{H}_\gamma^2(u)}. \quad (3.2)$$

*Proof.* Let  $\mathcal{A}_{\gamma,\ell}$  denote the set of  $\mathcal{A}_\gamma$ -structures  $s$  such that  $\mathbf{Level}(s) = \ell$ . Let  $s^*$  denote the  $\mathcal{A}_{\gamma,\ell}$ -matching that contains only  $\ell$  noncrossing exterior arcs. Any  $\mathcal{A}_{\gamma,\ell}$  matching  $s$  can be obtained from  $s^*$  through the following procedure, cf. Fig. 9.

First, the diagram  $s^*$  has  $\ell - 1$  pairs of  $P$ -intervals of the form  $([i, i + 1], [2\ell - i, 2\ell - i + 1])$  for each  $i \in \{1, 2, \dots, \ell\}$ . Inserting nonempty  $\gamma$ -matchings into at

least one of the two intervals of each pair  $([i, i + 1], [2\ell - i, 2\ell - i + 1])$  we obtain the class

$$\mathcal{U}^\ell \times (\mathcal{H}_\gamma^2 - \epsilon)^{\ell-1}.$$

Second we inflate any exterior arc in the so derived matching into an exterior stack and thereby derive the class

$$\mathcal{N}_{\gamma,\ell} = (\mathcal{U} \times \mathbf{SEQ}(\mathcal{U}))^\ell \times (\mathcal{H}_\gamma^2 - \epsilon)^{\ell-1}.$$

Since by construction,

$$\mathcal{A}_{\gamma,\ell} = \mathcal{X}_\gamma \bullet \mathcal{N}_{\gamma,\ell} \bullet \mathcal{X}_\gamma,$$

we have

$$\mathcal{A}_{\gamma,\ell} = \mathcal{H}_\gamma^2 \times (\mathcal{U} \times \mathbf{SEQ}(\mathcal{U}))^\ell \times (\mathcal{H}_\gamma^2 - \epsilon)^{\ell-1} \times \mathcal{H}_\gamma^2. \quad (3.3)$$

It remains to sum over all  $\ell \geq 1$  and the lemma follows.  $\square$

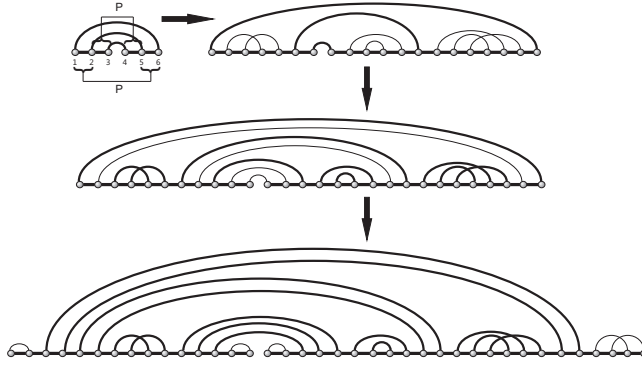


Fig. 9: The construction of  $\mathcal{A}_{\gamma,\ell}$  in case of  $\ell = 3$ .

**Theorem 3.** *The generating function of  $\gamma$ -matchings over two backbones,  $\mathbf{Q}_\gamma(u)$  satisfies*

$$\mathbf{Q}_\gamma(u) = \frac{\mathbf{H}_\gamma^2(u) \cdot \left( u\mathbf{H}_\gamma^2(u) + \sum_{g \leq \gamma} \mathbf{I}_{g,2} \left( \frac{u \cdot \mathbf{H}_\gamma^2(u)}{1-u} \right) \right)}{1 - u\mathbf{H}_\gamma^2(u) - \sum_{g \leq \gamma} \mathbf{I}_{g,2} \left( \frac{u \cdot \mathbf{H}_\gamma^2(u)}{1-u} \right)}. \quad (3.4)$$

In Tab. 1, we tabular  $\mathbf{q}_0(n)$ .

Table 1: The numbers of 0-matchings over two backbones.

$n$	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 10$
$\mathbf{q}_0(n)$	1	8	54	328	1882	10503	57845	316508	1725866	9392543

*Proof.* Let  $s$  be an arbitrary  $\mathcal{Q}_\gamma$ -matching. Consider the set  $\mathcal{Sh}_2(s)$  of irreducible shadows which contain at least one exterior  $s$ -arc. We observe that there exists exactly one element in  $\mathcal{Sh}_2(s)$  consisting of maximal arcs, i.e. all elements in  $\mathcal{Sh}_2(s)$  are nested. This fact can be understood by noting that two diagrams over two backbones, denoted by  $E_1$  and  $E_2$ , cannot be concatenated without generating a diagram over three backbones. It is straightforward to verify that the only way to combine such diagrams is via  $\bullet$ -product.

Let  $t = |\mathcal{Sh}_2(s)|$ , then

$$\mathcal{Q}_\gamma = \bigcup_{t=0} \mathcal{Q}_{\gamma,t}, \quad (3.5)$$

where  $\mathcal{Q}_\gamma$  and  $\mathcal{Q}_{\gamma,t}$  denote the set of  $\gamma$ -matchings over two backbones and the set of  $\gamma$ -matchings over two backbones containing  $t$  nested irreducible shadows over two backbones. In case of  $t = 0$   $\mathcal{Q}_{\gamma,0} = \mathcal{A}_\gamma$ . Suppose next  $t = 1$ . Let  $\phi$  be a fixed irreducible shadow of genus  $g$  having  $m$  arcs and let  $\mathcal{W}_\phi$  be the set of  $(\gamma)$ -matchings over two backbones,  $s$ , such that  $\mathcal{Sh}_2(s) = \{\phi\}$ . Any  $\mathcal{W}_\phi$ -diagram can be constructed as follows.

Let  $\mathcal{AX}_\gamma$  denote the set of all possible diagrams such that  $d \in \mathcal{AX}_\gamma$  is either a  $\mathcal{A}_\gamma$ -matching or a  $\mathcal{X}_\gamma$ -matching

$$\mathcal{AX}_\gamma = \mathcal{A}_\gamma + \mathcal{X}_\gamma.$$

Consider all possible diagrams obtained by the construction

$$\mathcal{AX}_\gamma \bullet \phi \bullet \mathcal{AX}_\gamma.$$

Next, we insert  $\mathcal{H}_\gamma$ -matching into the  $(2m - 2)$   $\sigma$ -intervals of the  $\phi$  and proceed by inflating each arc in  $\phi$  into a stack. Accordingly, we arrive at

$$\mathcal{W}_\phi = (\mathcal{U} \times \mathbf{SEQ}(\mathcal{U}))^m \times \mathcal{H}_\gamma^{2m-2} \times (\mathcal{A}_\gamma + \mathcal{H}_\gamma^2)^2. \quad (3.6)$$

A key observation is now that eq. (3.6) recruits only the number of arcs in  $\phi$  and does not depend on their particular configuration. Thus for any other fixed irreducible shadow  $\varrho$  of genus  $g$  over two backbones having  $m$  arcs, we have  $\mathbf{W}_\phi(u) = \mathbf{W}_\varrho(u)$ . Thus

$$\mathbf{Q}_{\gamma,1}(u) = \sum_{g \leq \gamma} \sum_{\tau \in \mathcal{I}_{g,2}} \mathbf{W}_\tau(u)$$

and consequently

$$\mathbf{Q}_{\gamma,1}(u) = \sum_{g \leq \gamma} \sum_m \mathbf{i}_{g,2}(m) \left( \frac{u}{1-u} \right)^m (\mathbf{A}_\gamma(u) + \mathbf{H}_\gamma^2(u))^2 \cdot \mathbf{H}_\gamma^{2m-2}(u). \quad (3.7)$$

Note that the summation over  $m$  is indeed finite, see Theorem 1.

In the case of  $t > 1$ , we proceed by induction on  $t$ . Let  $\hat{\mathcal{Q}}_{\gamma,t}$  denote the subset of  $\mathcal{Q}_\gamma$ -matchings  $s$  such that the maximal arcs of  $s$  form exactly an irreducible

shadow  $sh$  in  $\mathcal{S}h_2(s)$ . In the following, we refer  $sh$  as 2-shadow of  $s$ . Clearly, each  $\mathcal{Q}_\gamma$ -matching could be obtained via the  $\bullet$ -product  $\hat{\mathcal{Q}}_{\gamma,t} \bullet \mathcal{A}\mathcal{X}_\gamma$ . I.e.

$$\mathcal{Q}_{\gamma,t} = \hat{\mathcal{Q}}_{\gamma,t} \times (\mathcal{H}_\gamma^2 + \mathcal{A}_\gamma). \quad (3.8)$$

Set  $\varrho$  to be an irreducible shadow in  $\mathcal{I}_{g,2,m}$ , where  $g \leq \gamma$ . Let  $\hat{\mathcal{Q}}_{\gamma,t}^e$  denote the set of all possible  $\hat{\mathcal{Q}}_{\gamma,t}$ -matchings,  $\hat{s}$ , with exactly one 2-shadow  $\varrho$ . Any  $\hat{\mathcal{Q}}_{\gamma,t}^e$ -matching,  $\hat{s}$  can be obtained as follows: (1) nest exactly one  $\mathcal{Q}_{\gamma,t-1}$ -matching into  $\varrho$  via  $\bullet$ -product, (2) insert  $\mathcal{H}_\gamma$ -matchings into the  $(2m-2)$   $\sigma$ -intervals of  $\varrho$  and (3) inflate each arc in  $\varrho$  into a stack. Accordingly, we have

$$\hat{\mathcal{Q}}_{\gamma,t}^e = \mathcal{Q}_{\gamma,t-1} \times (\mathcal{U} \times \mathbf{SEQ}(\mathcal{U}))^m \times \mathcal{H}_\gamma^{2m-2}. \quad (3.9)$$

Since eqn. (3.9) depends only on the number of arcs in  $\varrho$  we derive

$$\hat{\mathcal{Q}}_{\gamma,t} = \mathcal{Q}_{\gamma,t-1} \times \sum_{g \leq \gamma} \sum_m \mathbf{i}_{g,2}(m) (\mathcal{U} \times \mathbf{SEQ}(\mathcal{U}))^m \times \mathcal{H}_\gamma^{2m-2}. \quad (3.10)$$

In view of eqn. (3.8) we obtain

$$\mathbf{Q}_{\gamma,t}(u) = \mathbf{Q}_{\gamma,t-1}(u) \cdot (\mathbf{H}_\gamma^2(u) + \mathbf{A}_\gamma(u)) \cdot \frac{1}{\mathbf{H}_\gamma^2(u)} \sum_{g \leq \gamma} \mathbf{I}_{g,2} \left( \frac{u \cdot \mathbf{H}_\gamma^2(u)}{1-u} \right). \quad (3.11)$$

Together with eqn. (3.7), we derive for  $t \geq 1$

$$\mathbf{Q}_{\gamma,t}(u) = \left( \frac{\mathbf{I}_{g,2} \left( \frac{u \cdot \mathbf{H}_\gamma^2(u)}{1-u} \right)}{\mathbf{H}_\gamma^2(u)} \right)^t \cdot (\mathbf{H}_\gamma^2(u) + \mathbf{A}_\gamma(u))^{t+1}. \quad (3.12)$$

In view of  $\mathbf{Q}_\gamma(u) = \sum_{t=0} \mathbf{Q}_{\gamma,t}(u)$  and  $\mathbf{Q}_{\gamma,0}(u) = \mathbf{A}_\gamma(u)$ , the theorem follows.  $\square$

**Theorem 4.** For  $\gamma = 0, 1$  the coefficients of  $\mathbf{Q}_\gamma(u)$  are asymptotically given by

$$[u^n] \mathbf{Q}_\gamma(u) \sim k_\gamma (\delta_\gamma^{-1})^n \quad (3.13)$$

for some constant  $k_\gamma > 0$ . Explicitly,  $\delta_0^{-1} \approx 5.4252$  and  $\delta_1^{-1} \approx 8.7266$ .

*Proof.* Suppose first  $\gamma = 0$ . Since  $\mathbf{H}_0(u) = \frac{1-\sqrt{1-4u}}{2u}$ , we have

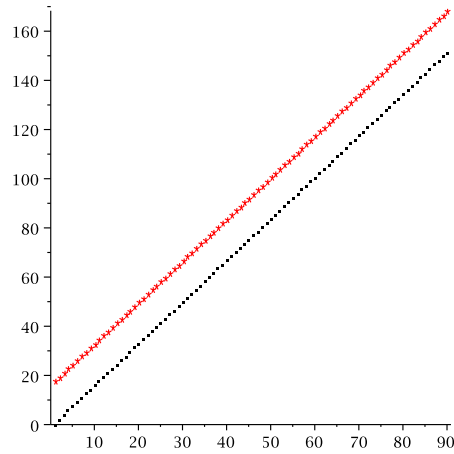
$$\mathbf{Q}_0(u) = \frac{\mathbf{P}_1(u) + \mathbf{P}_2(u) \cdot \sqrt{1-4u}}{u^2 \cdot \mathbf{P}_3(u)}, \quad (3.14)$$

where  $\mathbf{P}_1(u)$ ,  $\mathbf{P}_2(u)$  and  $\mathbf{P}_3(u)$  are explicit polynomials with nonzero constants. Let  $\mathbf{Z}(f)$  denotes the set of zeros of the function  $f$ . Consider the complex plane slit along the ray  $(1/4, +\infty)$ , then  $\mathbf{Q}_0(u)$  is analytic in the slit plane, except perhaps at  $u \in \mathbf{Z}(\mathbf{P}_3(u))$  and  $u = 0$ . Clearly  $u = 0$  is a removable singularity and

$\mathbf{Q}_0(u)$  is extended setting  $\mathbf{Q}_0(0) = 0$ . Since  $\mathbf{Q}_0(u)$  is a combinatorial generating function it is a power series with only nonnegative coefficients, Pringsheim's theorem guarantees the existence of some real dominant singularity,  $\delta_0$ . To locate  $\delta_0$  we observe that since  $\mathbf{P}_3(0) < 0$  and  $\mathbf{P}_3(1/4) > 0$ , there exists a unique smallest root of  $\mathbf{P}_3(u)$  in the interval  $(0, 1/4)$  and this root is a real dominant singularity,  $\tau_0$ . By explicit computation one verifies that  $\tau_0$  is indeed unique. Furthermore since  $\frac{\partial \mathbf{P}_3(u)}{\partial u} > 0$  for  $u \in [0, 1/4]$ , this root is simple, whence basic singularity analysis [7] implies that the subexponential factor of asymptotic expression is 1. Using the Maple command `fsolve`, we compute  $\delta_0 \approx 0.184325$ , i.e.

$$[u^n] \mathbf{Q}_0(u) \approx k_0 \cdot (5.4252)^n, \quad (3.15)$$

where  $k_0$  is some positive constant. In Fig. 10, we showcase the quality of the asymptotic formula for  $\gamma = 0$ .



**Fig. 10:** Exact enumeration versus asymptotic formula. The number of  $\mathcal{Q}_\gamma$ -matchings (black dots) versus the asymptotic formula  $\text{const} \cdot 5.4252^n$  (red asterisks). We separate the two data-sets setting the constant to  $10^7$ .

Let next  $\gamma = 1$  and  $\delta_1$  denote the dominant real singularity of

$$\mathbf{Q}_1(u) = \frac{\mathbf{H}_1^2(u) \cdot \left( u\mathbf{H}_1^2(u) + \sum_{g \leq \gamma} \mathbf{I}_{g,2} \left( \frac{u \cdot \mathbf{H}_1^2(u)}{1-u} \right) \right)}{1 - u\mathbf{H}_1^2(u) - \sum_{g \leq \gamma} \mathbf{I}_{g,2} \left( \frac{u \cdot \mathbf{H}_1^2(u)}{1-u} \right)},$$

whose existence is implied by Pringsheim's theorem. According to [8]  $\mathbf{H}_1(u)$  has its unique dominant singularity  $\mu_1$  at  $\approx 0.1207 > 2/17$ . We set

$$\mathbf{N}_1(u) = u\mathbf{H}_1^2(u) + \sum_{g \leq \gamma} \mathbf{I}_{g,2} \left( \frac{u \cdot \mathbf{H}_1^2(u)}{1-u} \right),$$

$\mathbf{F}_1(u) = \frac{1}{1-\mathbf{N}_1(u)}$  and  $\mathbf{L}_1(u) = \frac{1}{1-u}$ . By construction we then have

$$\mathbf{F}_1(u) = \mathbf{L}_1(\mathbf{N}_1(u)).$$

Clearly,  $\mathbf{L}_1(u)$  and  $\mathbf{N}_1(u)$  are power series with non-negative coefficients and in addition  $\mathbf{N}_1(0) = 0$  whence the composition is a well-defined power series. As for a dominant real singularity of  $\mathbf{N}_1(u)$ ,  $\rho_{\mathbf{N}_1}$ , we have  $\rho_{\mathbf{N}_1} = \rho_{\mathbf{H}_\gamma^2(u)}$ . We observe  $\mathbf{N}_1(\frac{2}{17}) > 1$  which implies that there exists some smallest  $\delta_1 \in (0, \frac{2}{17})$  such that  $\mathbf{N}_1(\delta_1) = 1$ . Since  $\mu_1 > 2/17 > \delta_1$  the composition power series  $\mathbf{L}_1(\mathbf{N}_1(u))$  belongs to the supercritical paradigm [7], i.e. its singular expansion is that of the outer function whose dominant singularity is encountered first. This evidently is the singular expansion of  $\mathbf{L}_1(u)$  leading (again) to the subexponential factor of 1. Using Maple command `fsolve`, we find  $\delta_1 \approx 0.1145925596 < \mu_1$  and derive

$$[u^n]\mathbf{F}_1(u) = c_1 \cdot \delta_1^{-n} \approx c_1 \cdot (8.7266)^n, \quad (3.16)$$

for some  $c_1 > 0$  and consequently

$$[u^n]\mathbf{Q}_1(u)k_1 \cdot \delta_1^{-n} \approx k_1 \cdot (8.7266)^n, \quad (3.17)$$

where  $k_{1,2}$  is some positive constant. This completes the proof of the theorem.  $\square$

#### 4. $\gamma$ -interaction structures

The enumeration of  $\gamma$ -interaction structures is obtained using *shapes*: A matching  $X$  is a shape if each stack in  $X$  is of length exactly one. Given an arbitrary matching  $s$ , its shape is obtained by iteratively collapsing each stack to a single arc and then removing any isolated vertices from the obtained diagram as illustrated in Fig. 11.

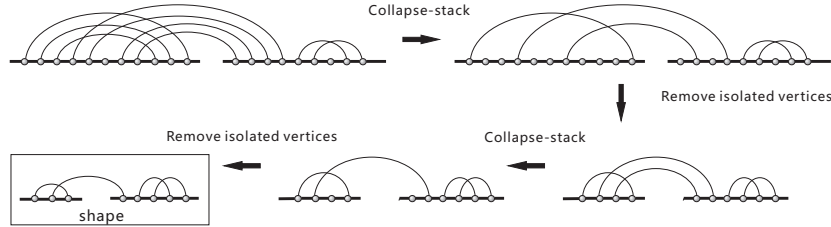


Fig. 11: From a diagram to its shape.

Let  $\mathcal{S}_\gamma$  denote the set of shapes that are  $\mathcal{Q}_\gamma$ -matchings and let  $\mathcal{Q}_\gamma(n, m)$  denote the combinatorial class of  $\mathcal{Q}_\gamma$ -matchings over  $2n$  vertices with  $m$  interior arcs of length 1 (1-arcs). Furthermore, let  $\mathcal{S}_\gamma(n, m)$  denote the class of all  $\mathcal{S}_\gamma$ -shapes over  $2n$  vertices with  $m$  1-arcs with generating function  $\mathbf{S}_\gamma(u, e)$ .

**Lemma 2.** For any  $\gamma \geq 0$ , we have

$$\mathbf{S}_\gamma(u, e) = \frac{(1+u)^2}{(1+2u-ue)^2} \mathbf{Q}_\gamma \left( \frac{u(1+u)}{(1+2u-ue)^2} \right). \quad (4.1)$$

*Proof.* We first claim that

$$\mathbf{Q}_\gamma(x, y) = \frac{1}{(x+1-yx)^2} \mathbf{Q}_\gamma \left( \frac{x}{(x+1-yx)^2} \right). \quad (4.2)$$

Consider the number of all possible  $\mathcal{Q}_\gamma(s+1, m+1)$ -matchings with exactly one labeled 1-arc. There are two ways to enumerate this. On one hand, choose an arbitrary  $\xi \in \mathcal{Q}_\gamma(s+1, m+1)$  and label one of its 1-arcs. Since we can label any of the  $(m+1)$  1-arcs of  $\xi$ ,  $(m+1) \mathbf{h}_\gamma(s+1, m+1)$  different such labeled linear arc diagrams arise. On the other hand, to produce  $\xi$  with this labeling, we can add one labeled 1-arc to an element of  $\mathcal{Q}_\gamma(s, m+1)$  by inserting a parallel copy of an existing 1-arc or by inserting a new labeled 1-arc into an element of  $\mathcal{Q}_\gamma(s, m)$ , where we may only insert the 1-arc between two vertices not already forming a 1-arc. It follows that we have the recursion

$$(m+1) \mathbf{q}_\gamma(n+1, m+1) = (m+1) \mathbf{q}_\gamma(n, m+1) + (2n+2-m) \mathbf{q}_\gamma(n, m)$$

or equivalently the PDE

$$\frac{\partial \mathbf{Q}_\gamma(x, y)}{\partial y} = x \frac{\partial \mathbf{Q}_\gamma(x, y)}{\partial y} + 2x^2 \frac{\partial \mathbf{Q}_\gamma(x, y)}{\partial x} + 2x \mathbf{Q}_\gamma(x, y) - xy \frac{\partial \mathbf{Q}_\gamma(x, y)}{\partial y}, \quad (4.3)$$

which is thus satisfied by  $\mathbf{Q}_\gamma(x, y)$ .

On the other hand,

$$\mathbf{Q}_\gamma^*(x, y) = \frac{1}{(x+1-yx)^2} \mathbf{Q}_\gamma \left( \frac{x}{(x+1-yx)^2} \right)$$

is also a solution of eq. (4.3), which specializes to  $\mathbf{Q}_\gamma(x) = \mathbf{Q}_\gamma^*(x, 1)$ , and moreover, we have  $\mathbf{q}_\gamma^*(n, m) = [x^n y^m] \mathbf{Q}_\gamma^*(x, y) = 0$ , for  $m > n$ . Indeed, the first assertion is easily verified directly, the specialization is obvious, and the fact that  $y$  only appears in the power series  $\mathbf{Q}_\gamma^*(x, y)$  in the form of products  $xy$  implies that  $\mathbf{q}_\gamma^*(n, m) = 0$ , for  $m > n$ . Thus, the coefficients  $\mathbf{q}_\gamma^*(n, m)$  satisfy the same recursion and initial conditions as  $\mathbf{q}_\gamma(n, m)$ , and hence by induction on  $n$ , we conclude  $\mathbf{q}_\gamma^*(n, m) = \mathbf{q}_\gamma(n, m)$ , for  $n, m \geq 0$ . This proves that  $\mathbf{Q}_\gamma(n, m)$  indeed satisfies eq. (4.2) as was claimed.

To complete the proof of eq. (4.1), we will define a projection

$$\vartheta : \mathcal{Q}_\gamma \rightarrow \mathcal{S}_\gamma$$

which project an arbitrary  $\mathcal{Q}_\gamma$ -matching  $s$  into its unique shape  $s^*$ . Clearly,  $\vartheta$  is surjective and affects neither irreducible shadows nor the number of 1-arcs. I.e. it can be naturally restricted to

$$\vartheta_m : \bigcup_n \mathcal{Q}_\gamma(n, m) \rightarrow \bigcup_n \mathcal{S}_\gamma(n, m).$$



Let us consider a fixed  $\gamma$ -shape,  $\lambda$ , having  $s$  arcs, of which  $t$  are 1-arcs and the generating function  $\mathbf{Q}_\gamma^\lambda(x, y)$ , counting  $\gamma$ -matchings that project into  $\lambda$  via  $\vartheta_m$ . Then

$$\mathbf{Q}_\gamma^\lambda(x, y) = \left( \frac{x}{1-x} \right)^s y^t,$$

which shows that  $\mathbf{Q}_\gamma^\lambda(x, y)$  depends only on the total number of arcs and number of 1-arcs in  $\lambda$ . Consequently,

$$\mathbf{Q}_\gamma(x, y) = \sum_{s \geq 0} \sum_{m=0}^s \mathbf{s}_\gamma(s, m) \left( \frac{x}{1-x} \right)^s y^m = \mathbf{S}_\gamma \left( \frac{x}{1-x}, y \right). \quad (4.4)$$

Setting  $u = \frac{x}{1-x}$ , i.e.,  $x = \frac{u}{1+u}$ , and  $e = y$ , we arrive at

$$\mathbf{S}_\gamma(u, e) = \frac{(1+u)^2}{(1+2u-ue)^2} \mathbf{Q}_\gamma \left( \frac{u(1+u)}{(1+2u-ue)^2} \right),$$

as required.  $\square$

Using symbolic enumeration we can conclude from Lemma 2:

**Lemma 3.** *Let  $\lambda$  be a fixed  $\gamma$ -shape with  $s \geq 1$  arcs and  $m \geq 0$  1-arcs. Then the generating function of  $\tau$ -canonical  $\gamma$ -diagrams containing no 1-arc that have shape  $\lambda$  is given by*

$$\mathbf{Q}_{\tau, \gamma}^\lambda(z) = (1-z)^{-2} \left( \frac{z^{2\tau}}{(1-z^2)(1-z)^2 - (2z-z^2)z^{2\tau}} \right)^s z^m.$$

In particular,  $\mathbf{Q}_{\tau, \gamma}^\lambda(z)$  depends only upon the number of arcs and 1-arcs in  $\lambda$ .

Our main result about enumerating  $\tau$ -canonical  $\gamma$ -interaction structures follows.

**Theorem 5.** *Suppose  $\gamma \geq 0$  and  $\tau \geq 1$  and let  $u_\tau(z) = \frac{(z^2)^{\tau-1}}{z^{2\tau} - z^2 + 1}$ . Then the generating function  $\mathbf{Q}_{\tau, \gamma}(z)$  is algebraic and given by*

$$\mathbf{Q}_{\tau, \gamma}(z) = \frac{1}{(u_\tau(z)z^2 - z + 1)^2} \mathbf{Q}_\gamma \left( \frac{u_\tau(z)z^2}{(u_\tau(z)z^2 - z + 1)^2} \right). \quad (4.5)$$

In particular for  $\gamma = 0$  we have

$$[z^n] \mathbf{Q}_{1,0}(z) \sim k_{1,0} (\rho_{1,0}^{-1})^n, \quad \text{and} \quad [z^n] \mathbf{Q}_{2,0}(z) \sim k_{2,0} (\rho_{2,0}^{-1})^n$$

and for  $\gamma = 1$  we have

$$[z^n] \mathbf{Q}_{1,1}(z) \sim k_{1,1} (\rho_{1,1}^{-1})^n, \quad \text{and} \quad [z^n] \mathbf{Q}_{2,1}(z) \sim k_{2,1} (\rho_{2,1}^{-1})^n$$

for some constants  $k_1, k_2$ , where  $\rho_{0,1}^{-1} = 2.9954$ ,  $\rho_{0,2}^{-1} = 2.0935$ ,  $\rho_{1,1}^{-1} = 3.6825$  and  $\rho_{1,2}^{-1} = 2.2992$ .

*Proof.* According to Lemma 3 and  $\mathbf{Q}_{\tau,\gamma}(z) = \sum_{\lambda \in \mathcal{S}_\gamma} \mathbf{Q}_{\tau,\gamma}^\lambda(z)$ , we obtain

$$\mathbf{Q}_{\tau,\gamma}(z) = (1-z)^{-2} \cdot \mathbf{S}_\gamma \left( \frac{z^{2\tau}}{(1-z^2)(1-z)^2 - (2z-z^2)z^{2\tau}}, z \right). \quad (4.6)$$

Furthermore, eqn. (4.5) follows by substituting  $u = \frac{z^{2\tau}}{(1-z^2)(1-z)^2 - (2z-z^2)z^{2\tau}}$  and  $e = z$  in eqn. (4.1).

Clearly  $\mathbf{Q}_{\tau,\gamma}(z)$  is algebraic according to eqn. (4.1). Furthermore, since the composition scheme is supscritical in case of  $0 \leq \gamma \leq 1$  and  $1 \leq \tau \leq 2$ , the singularity type is that of the external function, i.e.  $\mathbf{Q}_\gamma(z)$ , whence the theorem.  $\square$

## 5. Discussion

Our main result opens the door to interesting research. Namely, it is of great importance to introduce the topological genus as an additional filtration into  $\mathbf{Q}_\gamma(u)$  and to thereby pass to a bivariate generating function. A result of [2] indicates how this can be derived. There it is proved how to compute the topological genus of a  $\gamma$ -interaction structure. The latter formula is in difference to the case of a single backbone not simply the sum of the genera of its irreducible shadows. This computation can be weaved into the combinatorial construction presented here in order to refine our results by the topological genus.

The derivation of such bivariate algebraic equations is instrumental for obtaining recursions for computing shadows of genus  $g$  from those of smaller genera. Similar to the Zagier-Harer generating function [27] it is a fascinating prospect to derive a recursion for the polynomials  $\mathbf{I}_{g,2}(z)$ . Here it will be vital to obtain hints for bijective proof hidden in the algebraic formulas. Common factors of these polynomials whose coefficients count numbers of irreducible shadows of fixed genus will be the key for deeper understanding. Accordingly it is of importance to compute the polynomials  $\mathbf{I}_{g,2}(z)$  explicitly for all relevant genera. Once the algebraic recursions for these polynomials are obtained it is of importance to derive bijective proofs. Results along these lines will have profound algorithmic impact and offer novel insights in how to fold interaction structures faster.

## References

- [1] C. Alkan, E. Karakoc, J. H. Nadeau, S. C. Sahinalp, and K. Z. Zhang. RNA-RNA interaction prediction and antisense RNA target search. *J. Comput. Biol.*, 13:267–282, 2006.
- [2] J. E. Andersen, F. W. D. Huang, R. C. Penner, and C. M. Reidys. *Topology of RNA-RNA interaction structures*. 2012.
- [3] S. Bernhart, H. Tafer, U. Mückstein, C. Flamm, P. F. Stadler, and I. L. Hofacker. Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol. Biol.*, 1:3, 2006.

- [4] M. Bon, G. Vernizzi, H. Orland, and A. Zee. Topological classification of RNA structures. *J. Mol. Biol.*, 379:900–911, 2008.
- [5] A. Busch, A. S. Richter, and R. Backofen. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, 24:2849–2856, 2008.
- [6] A. Busch, A. S. Richter, and R. Backofen. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, 24:2849–2856, 2008.
- [7] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press New York, 2009.
- [8] H. S. W. Han and C. M. Reidys. Combinatorics of  $\gamma$ -structures. submitted. 2012.
- [9] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, 125:167–188, 1994.
- [10] F. W. D. Huang, J. Qin, P. F. Stadler, and C. M. Reidys. Partition function and base pairing probabilities for RNA-RNA interaction prediction. *Bioinformatics*, 25(20):2646–2654, 2009.
- [11] F. W. D. Huang, J. Qin, C. M. Reidys, and P. F. Stadler. Target prediction and a statistical sampling algorithm for RNA-RNA interaction. *Bioinformatics*, 26:175–181, 2010.
- [12] A. M. Kiss, B. E. Jady, E. Bertrand, and T. Kiss. Human box H/ACA pseudouridylation guide RNA machinery. *Mol. Cell Biol.*, 24:5797–5807, 2004.
- [13] H. Ly, L. Xu, M. A. Rivera, T. G. Parslow, and E. H. Blackburn. A role for a novel ‘trans-pseudoknot’ RNA-RNA interaction in the functional dimerization of human telomerase. *Genes & Dev.*, 17:1078–1083, 2003.
- [14] W. S. Massey. *Algebraic Topology: An Introduction*. Springer-Verlag, New York, 1967.
- [15] U. Mückstein, H. Tafer, S. H. Bernhard, M. Hernandez-Rosales, J. Vogel, P. F. Stadler, and I. L. Hofacker. Translational control by RNA-RNA interaction: Improved computation of RNA-RNA binding thermodynamics. In Mourad Elloumi, Josef Küng, Michal Linial, Robert F. Murphy, Kristan Schneider, and Cristian Toma Toma, editors, *Bioinformatics Research and Development — BIRD 2008*, volume 13 of *Comm. Comp. Inf. Sci.*, pages 114–127, Berlin, 2008. Springer.
- [16] U. Mückstein, H. Tafer, J. Hackermüller, S. H. Bernhard, P. F. Stadler, and I. L. Hofacker. Thermodynamics of RNA-RNA binding. *Bioinformatics*, 22:1177–1182, 2006.

- [17] J. Ofengand and A. Bakin. Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaeobacteria, mitochondria and chloroplasts. *J. Mol. Biol.*, 266:246–268, 1997.
- [18] H. Orland and A. Zee. RNA folding and large  $n$  matrix theory. *Nuclear Physics B*, 620:456–476, 2002.
- [19] R. C. Penner. Cell decomposition and compactification of Riemann’s moduli space in decorated Teichmüller theory. In Nils Tongring and R. C. Penner, editors, *Woods Hole Mathematics-perspectives in math and physics*, pages 263–301. World Scientific, Singapore, 2004. arXiv: math.GT/0306190.
- [20] R. C. Penner. *Decorated Teichmüller Theory*. European Mathematical Society, Zürich, 2011.
- [21] R. C. Penner, Michael Knudsen, Carsten Wiuf, and Jørgen Ellegaard Andersen. Fatgraph models of proteins. *Comm. Pure Appl. Math.*, 63:1249–1297, 2010.
- [22] R. C. Penner and M. S. Waterman. Spaces of RNA secondary structures. *Adv. Math.*, 101:31–49, 1993.
- [23] D. D. Pervouchine. IRIS: Intermolecular RNA interaction search. *Proc. Genome Informatics*, 15:92–101, 2004.
- [24] C. M. Reidys, F. W. D. Huang, J. E. Andersen, R. C. Penner, P. F. Stadler, and M. E. Nebel. Topology and prediction of RNA pseudoknots. *Bioinformatics*, 27:1076–1085, 2011.
- [25] E. Rivas and S. R. Eddy. A dynamic programming algorithms for RNA structure prediction including pseudoknots. *J. Mol. Biol.*, 285:2053–2068, 1999.
- [26] G. Vernizzi and H. Orland. Large- $N$  random matrices for RNA folding. *Acta Phys. Polon.*, 36:2821–2827, 2005.
- [27] D. Zagier. On the distribution of the number of cycles of elements in symmetric groups. *Nieuw Arch. Wisk. IV*, 13:489–495, 1995.