Combinatorics of γ -Structures

HILLARY S.W. HAN, THOMAS J.X. LI, and CHRISTIAN M. REIDYS

ABSTRACT

In this article we study canonical γ -structures, a class of RNA pseudoknot structures that plays a key role in the context of polynomial time folding of RNA pseudoknot structures. A γ -structure is composed of specific building blocks that have topological genus less than or equal to γ , where composition means concatenation and nesting of such blocks. Our main result is the derivation of the generating function of γ -structures via symbolic enumeration using so called irreducible shadows. We furthermore recursively compute the generating polynomials of irreducible shadows of genus $\leq \gamma$. The γ -structures are constructed via γ -matchings. For $1 \leq \gamma \leq 10$, we compute Puiseux expansions at the unique, dominant singularities, allowing us to derive simple asymptotic formulas for the number of γ -structures.

Key words: generating function, irreducible shadow, shape, γ -structure.

1. INTRODUCTION AND BACKGROUND

AN RNA SEQUENCE IS A LINEAR, ORIENTED SEQUENCE of the nucleotides (bases) **A,U,G,C**. These sequences “fold” by establishing bonds between pairs of nucleotides. These bonds cannot form arbitrarily; a nucleotide can at most establish one bond, and the global conformation of an RNA molecule is determined by topological constraints encoded at the level of secondary structure, that is, by the mutual arrangements of the base pairs (Bailor et al., 2010).

Secondary structures can be interpreted as (partial) matchings in a graph of permissible base pairs (Tabaska et al., 1998). When represented as a diagram, that is, as a graph whose vertices are drawn on a horizontal line with arcs in the upper halfplane, one refers to a secondary structure with crossing arcs as a pseudoknot structure.

Folded configurations exhibit the stacking of adjacent base pairs and specific minimum arc-length conditions (Smith and Waterman, 1978), where a stack is a sequence of parallel arcs $[(i, j), (i + 1, j - 1), \dots, (i + \tau, j - \tau)]$.

The topological classification of RNA structures (Bon et al., 2008); Andersen et al., 2013) has recently been translated into an efficient folding algorithm (Reidys et al., 2011). This algorithm *a priori* folds into a novel class of pseudoknot structures, the γ -structures. These γ -structures differ from pseudoknotted RNA structures of fixed topological genus of an associated fatgraph or double-line graph (Orland and Zee, 2002; Bon et al., 2008), because they have an arbitrarily high genus. They are composed of irreducible sub-diagrams whose individual genus is bounded by γ and contain no bonds of length one (1-arcs; see Section 2 for details).

Nebel and Weinberg (2011) study a plethora of RNA structures. The authors study asymptotic expansions for $\gamma = 1$ and find that in the limit of large n there are $j_1 n^{-\frac{3}{2}}(\varrho_{1,1}^{-1})^n$, 1-structures, where $\varrho_{1,1}^{-1} = 3.8782$ and j_1 is some positive constant.

In this article we study canonical γ -structures, that is, partial matchings composed by irreducible motifs of genus $\leq \gamma$ without isolated arcs and 1-arcs. These motifs are called irreducible shadows. We first establish a functional relationship between the generating function of γ -matchings and that of irreducible shadows. Via this relation, we identify a polynomial $P_\gamma(u, X)$, whose unique solution equals the generating function of γ -matchings. We then derive a recurrence of the generating function of irreducible shadows using the Harer-Zagier recurrence (Harer and Zagier, 1986). The generating function of γ -matchings is then expanded at its unique dominant singularity as a Puiseux-series. This implies, by means of transfer theorems (Flajolet and Sedgewick, 2009), simple asymptotic formulas for the numbers of γ -matchings.

γ -Matchings are the stepping stone to derive via Lemma 3 the further refined, bivariate generating function of γ -shapes, that is, γ -matchings containing only stacks composed by a single arc. This generating function keeps additional track of the 1-arcs, which are vital for the later inflation into γ -structures. We then compute the generating function of τ -canonical γ -structures, inflating γ -shapes by means of symbolic enumeration.

2. SOME BASIC FACTS

2.1. γ -Diagrams

A *diagram* is a labeled graph over the vertex set $[n] = \{1, \dots, n\}$, in which each vertex has degree ≤ 3 , represented by drawing its vertices in a horizontal line. The backbone of a diagram is the sequence of consecutive integers $(1, \dots, n)$ together with the edges $\{(i, i+1) \mid 1 \leq i \leq n-1\}$. The arcs of a diagram, (i, j) , where $i < j$, are drawn in the upper half-plane. We shall distinguish the backbone edge $\{i, i+1\}$ from the arc $(i, i+1)$, which we refer to as a 1-arc.

A *stack* of length τ is a maximal sequence of “parallel” arcs,

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

A stack of length $\geq \tau$ is called a τ -canonical stack; that is, a stack of length zero is an isolated arc. The particular arc $(1, n)$ is called a *rainbow*, and an arc is called *maximal* if it is maximal with respect to the partial order $(i, j) \leq (i', j')$ iff $i' \leq i \wedge j \leq j'$ (Fig. 1).

A stack of length τ $[(i, j), (i+1, j-1), \dots, (i+\tau, j-\tau)]$ induces a sequence of pairs $[[i, i+1], [j, j-1]], [(i+1, i+2], [j-1, j-2]] \dots [(i+\tau-1, i+\tau], [j-\tau, j-\tau+1]]$. We call any of these 2τ intervals a *P-interval*. The interval $[i+\tau, j-\tau]$ is called a *σ -interval* (Fig. 2).

We shall consider diagrams as fatgraphs, $\mathbb{G}$; that is, graphs G together with a collection of cyclic orderings, called *fattenings*, one such ordering on the half-edges incident on each vertex. Each fatgraph $\mathbb{G}$ determines an oriented surface $F(\mathbb{G})$ (Loebl and Moffatt, 2008; Penner et al., 2010), which is connected if G is and has some associated genus $g(G) \geq 0$ and number $r(G) \geq 1$ of boundary components. Clearly, $F(\mathbb{G})$ contains G as a deformation retract (Massey, 1967). Fatgraphs were first applied to RNA secondary structures in Penner and Waterman (1993) and Penner (2004).

A diagram $\mathbb{G}$ hence determines a unique surface $F(\mathbb{G})$ (with boundary). Filling the boundary components with discs we can pass from $F(\mathbb{G})$ to a surface without boundaries. Euler characteristic, χ , and genus, g , of this surface is given by $\chi = v - e + r$ and $g = 1 - \frac{1}{2}\chi$, respectively, where v, e, r is the number of discs,

FIG. 1. (a) A diagram containing a rainbow (in boldface), three stacks $((5, 9), (6, 8)), ((10, 15), (11, 14), (12, 13))$, and $((1, 19), (2, 18), (3, 17), (4, 16))$. (b) The maximal arcs of a diagram displayed in (boldface).

FIG. 2. A schematic showing σ - and P -intervals.

ribbons, and boundary components in $\mathbb{G}$ (Massey 1967). The *genus* of a diagram is that of its associated surface without boundary.

The *shadow* of a diagram of genus g is obtained by removing all noncrossing arcs, deleting all isolated vertices, and collapsing all induced stacks (i.e., maximal subsets of subsequent, parallel arcs) to single arcs (Fig. 3). We denote shadows by σ .

The shadow of a diagram $\mathbb{G}$, $\sigma(\mathbb{G})$, can possibly be empty. Furthermore, projecting into the shadow does not affect genus. Any shadow of genus g over one backbone contains at least $2g$ and at most $(6g - 2)$ arcs. In particular, for fixed genus g , there exist only finitely many shadows (Reidys et al., 2011; Andersen et al., 2012). In Figure 4, we display the four shadows of genus one.

A diagram is called *irreducible* if and only if for any two arcs, α_1, α_k contained in E , there exists a sequence of arcs $(\alpha_1, \alpha_2, \dots, \alpha_{k-1}, \alpha_k)$ such that (α_i, α_{i+1}) are crossing. Irreducibility is equivalent to the concept of primitivity introduced by (Bon et al., 2008), inspired by the work of Dyson (1949). According to Andersen et al. (2012), for arbitrary genus g and $2g \leq \ell \leq (6g - 2)$, there exists an irreducible shadow of genus g having exactly ℓ arcs. We may reuse Figure 4 as an illustration of this result since the four shadows of genus one are all irreducible.

Let $\mathbf{i}_g(m)$ denote the number of irreducible shadows of genus g with m arcs. Since for fixed genus g there exist only finitely many shadows, we have the generating polynomial of irreducible shadows of genus g

$$\mathbf{I}_g(z) = \sum_{m=2g}^{6g-2} \mathbf{i}_g(m) z^m.$$

For instance, for genus 1 and 2 we have

$$\begin{aligned} \mathbf{I}_1(z) &= z^2(1+z)^2, \\ \mathbf{I}_2(z) &= z^4(1+z)^4(17+92z+96z^2). \end{aligned}$$

FIG. 3. The shadow is obtained by removing all noncrossing arcs and isolated points and collapsing all stacks and resulting stacks into single arcs.

FIG. 4. The four shadows of genus one.

The shadow $\sigma(\mathbb{G})$ of a diagram $\mathbb{G}$ decomposes into a set of irreducible shadows. We shall call these shadows *irreducible $\mathbb{G}$ -shadows*.

Any diagram $\mathbb{G}$ can iteratively be decomposed by first removing all noncrossing arcs as well as isolated vertices, second by collapsing any stacks, and third by removing irreducible $\mathbb{G}$ -shadows iteratively as follows (Fig. 5):

- one removes (i.e., cuts the backbone at two points and after removal merges the cut-points) irreducible $\mathbb{G}$ -shadows from bottom to top, such that there exists no irreducible $\mathbb{G}$ -shadow nested within the one previously removed.
- if the removal of an irreducible $\mathbb{G}$ -shadow induces the formation of a stack, it is collapsed into a single arc.

A diagram, $\mathbb{G}$, is a γ -diagram if and only if for any irreducible $\mathbb{G}$ -shadow, $\mathbb{G}'$, $g(\mathbb{G}') \leq \gamma$ holds.

We denote the set of τ -canonical γ -diagrams by $\tilde{\mathcal{G}}_{\tau, \gamma}$. Such a diagram without arcs of the form $(i, i+1)$ (1-arcs) is called a τ -canonical γ -structure, and their set is denoted by $\mathcal{G}_{\tau, \gamma}$. A γ -matching is a γ -diagram that contains only vertices of degree three. A γ -shape is a γ -matching that contains only stacks of length zero. Let $\mathcal{H}_\gamma$ and $\mathcal{S}_\gamma$ denote the set of γ -matchings and γ -shapes, respectively.

2.2. Some generating functions

In this article we denote the ring of polynomials over a ring R by $R[X]$ and the ring of formal power series $\sum_{n \geq 0} a_n X^n$ by $R[[X]]$. $R[[X]]$ is a local ring with maximal ideal (X) , that is, any power series with nonzero constant term is invertible. A Puiseux-series (Wall, 2004) is a power series in fractional powers of X , that is, $\sum_{n \geq 0} a_n X^{n/k}$ for some fixed $k \in \mathbb{N}$.

FIG. 5. A diagram $\mathbb{G}$ is decomposed: We remove any noncrossing arcs and isolated points, collapse any stacks into single arcs, and finally remove irreducible $\mathbb{G}$ -shadows from bottom to top and collapse any stack generated in the process into a single arc.

We denote the generating functions of a set of diagrams $\mathcal{D}$ filtered by the number of arcs $\mathbf{D}(z) = \sum_{2n \geq 0} \mathbf{d}(n)z^n$. Similarly, a generating function of diagrams filtered by the length of the backbone is written as $\mathbf{D}(z) = \sum_{n \geq 0} \mathbf{d}(n)z^n$. In particular, the generating functions of γ -matchings and τ canonical γ -structures are given by

$$\mathbf{H}_\gamma(u) = \sum_{2n \geq 0} \mathbf{h}_\gamma(n)u^n, \quad \mathbf{G}_{\tau, \gamma}(z) = \sum_{n \geq 0} \mathbf{g}_{\tau, \gamma}(n)z^n.$$

Let $\mathcal{H}_\gamma(n, m) \supseteq \mathcal{S}_\gamma(n, m)$ denote the collections of all γ -matchings and γ -shapes on $2n \geq 0$ vertices containing $m \geq 0$ 1-arcs with generating functions

$$\mathbf{H}_\gamma(x, y) = \sum_{m, 2n \geq 0} \mathbf{h}_\gamma(n, m)x^n y^m, \quad \mathbf{S}_\gamma(x, y) = \sum_{m, 2n \geq 0} \mathbf{s}_\gamma(n, m)x^n y^m,$$

where $\mathbf{h}_\gamma(n, m) = \mathbf{s}_\gamma(n, m) = 0$ if $2\gamma > n$ or if $m > n$.

Furthermore, there is a natural projection ϑ from γ -matchings to γ -shapes defined by collapsing each non-empty stack onto a single arc

$$\vartheta : \mathcal{H}_\gamma \rightarrow \mathcal{S}_\gamma,$$

which is surjective and preserves irreducible shadows as well as the number of 1-arcs. ϑ restricts to a surjection

$$\vartheta : \sqcup_{n \geq 0} \mathcal{H}_\gamma(n, m) \rightarrow \sqcup_{n \geq 0} \mathcal{S}_\gamma(n, m),$$

which collapses each stack to an arc and preserves any irreducible shadow and also the number m of 1-arcs.

3. COMBINATORICS OF γ -MATCHINGS

In this section, we study γ -matchings.

Theorem 1. *Let $R = \mathbb{Z}[u]$. Then the following assertions hold:*

(a) *The generating function of γ -matchings, $\mathbf{H}_\gamma(u)$, satisfies*

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

equivalently,

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

In particular, there exists a polynomial $P_\gamma(u, X) \in R[X]$ of degree $(12\gamma - 2)$, whose coefficients are sums of $\mathbf{I}_g(z)$ coefficients, such that $P_\gamma(u, \mathbf{H}_\gamma(u)) = 0$.

(b) *Equation (3.1) determines $\mathbf{H}_\gamma(u)$ uniquely.*

Proof. We first prove (a). Let σ be a fixed irreducible shadow of genus g having m arcs. Let $\mathcal{V}_\sigma$ be the set of diagrams, generated by concatenating and nesting σ .

Claim 1:

$$\mathbf{V}_\sigma(u) = (1 - \mathbf{V}_\sigma(u)^{-1} (u \mathbf{V}_\sigma(u)^2)^m)^{-1}.$$

To prove Claim 1 we consider a $\mathcal{V}_\sigma$ -diagram. Clearly, its maximal arcs are contained in $t \geq 1$ copies of σ . These arcs induce exactly $(2m - 1)t$ σ -intervals, in each of which we find again an element of $\mathcal{V}_\sigma$, whence

$$\mathbf{V}_\sigma(u) = \sum_{t \geq 0} (u^m \mathbf{V}_\sigma(u)^{2m-1})^t$$

and Claim 1 follows.

Let $\mathcal{L}_\sigma$ be the set of diagrams having the fixed shape σ obtained by inflating σ -arcs into stacks, or symbolically, $\mathcal{U} \times \text{SEQ}(\mathcal{U})$. Here $\mathcal{U}$ and $\mathcal{R} = \text{SEQ}(\mathcal{U})$ denote the classes of arcs and sequences of arcs. Clearly, the associated generating function of $\mathcal{U} \times \mathcal{R}$ is $u(1-u)^{-1}$.

Note that each $\mathcal{L}_\sigma$ -diagram contains exactly $(2m-1)$ σ -intervals and an arbitrary number of pairs of P -intervals. Let $\mathcal{F}_\sigma$ denote the set of diagrams generated by concatenating and nesting $\mathcal{L}_\sigma$ -diagrams that contain no empty P -intervals. Let finally $\mathcal{W}_\sigma$ be the set of one-canonical diagrams, having shapes in $\mathcal{F}_\sigma$.

Claim 2.

$$\mathbf{W}_\sigma(u)^{-1} = 1 - \mathbf{W}_\sigma(u)^{-1} \left(\frac{\frac{u}{1-u} \mathbf{W}_\sigma^2(u)}{1 - \frac{u}{1-u} (\mathbf{W}_\sigma^2(u) - 1)} \right)^m. \quad (2)$$

We shall construct $\mathcal{W}_\sigma$ using arcs, $\mathcal{U}$; sequences of arcs, $\mathcal{R}$; induced arcs, $\mathcal{N}$; and sequence of induced arcs, $\mathcal{M}$. The class $\mathcal{F}_\sigma$ is obtained by concatenating and nesting $\mathcal{L}_\sigma$ -diagrams that do not contain any empty P -intervals (Fig. 6).

An induced arc, that is, an arc together with at least one nontrivial $\mathcal{F}_\sigma$ -diagram in either one or in both P -intervals

$$\mathcal{N} = \mathcal{U} \times ((\mathcal{F}_\sigma - 1) + (\mathcal{F}_\sigma - 1) + (\mathcal{F}_\sigma - 1)^2) = \mathcal{U} \times (\mathcal{F}_\sigma^2 - 1).$$

Clearly, we have for a single induced arc $\mathbf{N}(u) = u(\mathbf{F}_\sigma(u)^2 - 1)$ and for a sequence of induced arcs, $\mathcal{M} = \text{SEQ}(\mathcal{N})$, where

$$\mathbf{M}(u) = \frac{1}{1 - u(\mathbf{F}_\sigma(u)^2 - 1)}.$$

By construction, the maximal arcs of an $\mathcal{F}_\sigma$ -diagram coincide with those of its underlying $\mathcal{V}_\sigma$ -diagram. Therefore,

$$\mathcal{F}_\sigma = \sum_{t \geq 0} ((\mathcal{U} \times \mathcal{M})^m \mathcal{F}_\sigma^{2m-1})^t$$

with generating function

$$\mathbf{F}_\sigma(u) = \sum_{t \geq 0} \left(\left(\frac{u}{1 - u(\mathbf{F}_\sigma(u)^2 - 1)} \right)^m \mathbf{F}_\sigma(u)^{2m-1} \right)^t. \quad (3)$$

Next we inflate the arcs of the $\mathcal{F}_\sigma$ -diagram into stacks, $\mathcal{U} \times \mathcal{R}$.

This inflation process generates $\mathcal{W}_\sigma$ -diagrams, and any $\mathcal{W}_\sigma$ -diagram can be constructed from a unique fixed irreducible shadow σ of genus g with m arcs. We have

$$\mathbf{W}_\sigma(u) = \sum_{t \geq 0} \left(\left(\frac{\frac{u}{1-u}}{1 - \frac{u}{1-u} (\mathbf{W}_\sigma(u)^2 - 1)} \right)^m \mathbf{W}_\sigma(u)^{2m-1} \right)^t, \quad (4)$$

whence Claim 2.

Claim 3: Let M be the set of irreducible shadows of genus $g \leq \gamma$. Then

$$\mathbf{W}_M(u)^{-1} = 1 - \sum_{g \leq \gamma} \sum_{1 < m} \mathbf{i}_g(m) \mathbf{W}_M(u)^{-1} \left(\frac{u \mathbf{W}_M^2(u)}{1 - u \mathbf{W}_M^2(u)} \right)^m. \quad (5)$$

The maximal arcs of a $\mathcal{V}_M$ -structure, partition into the maximal arcs of t concatenated irreducible shadows $\sigma_1, \dots, \sigma_t$ and

$$\sum_{\substack{\{\sigma_1, \dots, \sigma_t\} \\ \sigma_i \in M}} 1 = \left(\sum_{g \leq \gamma} \sum_{1 < m} \mathbf{i}_g(m) \right)^t. \quad (6)$$

These maximal arcs induce exactly $(2m-1)t$ σ -intervals. In each σ -interval, we find again an element of $\mathcal{V}_M$. Thus for any σ_i having m arcs, we have $\mathcal{V}_M^{2m-1}$, which leads to the term $u^m \mathbf{V}_M(u)^{2m-1}$. It remains to

FIG. 6. First, a fixed irreducible shadow σ is inflated into a $\mathcal{L}_\sigma$ -diagram; second, we pass to an $\mathcal{F}_\sigma$ -diagram by inserting a nontrivial $\mathcal{L}_\sigma$ -diagram in one of the P -intervals.

sum over all t , that is, expressing all the decompositions of $\mathcal{V}_{M_\gamma}$ -structures into concatenated, irreducible shadows, and we obtain

$$\mathbf{V}_M(u) = \sum_{t \geq 0} \left(\sum_{g \leq \gamma} \sum_{m > 1} \mathbf{i}_g(m) u^m \mathbf{V}_M(u)^{2m-1} \right)^t. \quad (7)$$

The passage from $\mathcal{V}_M$ to $\mathcal{L}_M$, as well as that from $\mathcal{L}_M$ to $\mathcal{F}_M$, follows from Claim 2, whence

$$\mathbf{F}_M(u) = \sum_{t \geq 0} \left(\sum_{g \leq \gamma} \sum_{m > 1} \mathbf{i}_g(m) \mathbf{F}_M(u)^{-1} \left(\frac{u \mathbf{F}_M^2(u)}{1 - u(\mathbf{F}_M^2(u) - 1)} \right)^m \right)^t. \quad (8)$$

Here $\mathbf{F}_M(u)^{-1}$ exists in $\mathbb{C}[[u]]$, having a nonzero constant term. Next we inflate the arcs of the $\mathcal{F}_M$ -structure into stacks, obtaining

$$\mathbf{W}_M(u)^{-1} = 1 - \sum_{g \leq \gamma} \sum_{1 < m} \mathbf{i}_g(m) \mathbf{W}_M(u)^{-1} \left(\frac{u \mathbf{W}_M^2(u)}{1 - u(\mathbf{W}_M^2(u))} \right)^m. \quad (9)$$

We next derive the functional equation for $\mathbf{H}_\gamma(u)$ by incorporating noncrossing arcs. Since the maximal arcs composed of noncrossing arcs are exactly rainbows, the generating function of $\mathcal{H}_\gamma$ -diagrams nested in a rainbow is given by $u\mathbf{H}_\gamma(u)$. As in Claim 3, we conclude

$$\mathbf{H}_\gamma(u)^{-1} = 1 - \sum_{g \leq \gamma} \left(u \mathbf{H}_\gamma(u) + \mathbf{H}_\gamma(u)^{-1} \sum_{m > 1} \mathbf{i}_g(m) \vartheta(u)^m \right),$$

where

$$\vartheta(u) = \frac{u \mathbf{H}_\gamma^2(u)}{1 - u \mathbf{H}_\gamma^2(u)}.$$

Setting $w_u(X) = 1 - uX^2$, Equation 1 gives rise to the polynomial

$$P_\gamma(u, X) = w_u(X)^{\kappa_\gamma} (-1 + X - uX^2) - \sum_{g \leq \gamma} w_u(X)^{\kappa_\gamma} \mathbf{I}_g \left(\frac{uX^2}{w_u(X)} \right), \quad (10)$$

where $\kappa_\gamma = 6\gamma - 2$, $\deg(P_\gamma(u, X)) = (2 + 2\kappa_\gamma)$, $[X^{2+2\kappa_\gamma}]P_\gamma(u, X) = -u^{1+\kappa_\gamma}$, and $P_\gamma(u, \mathbf{H}_\gamma(u)) = 0$, whence (a).

It remains to prove (b). Since M is the finite set of irreducible shadows of genus $g \leq \gamma$, and any such shadow has $2g \leq m \leq \kappa_\gamma$ arcs (Andersen et al., 2012), any M -shadow has $\leq \kappa_\gamma$ arcs. Setting $v(u) = 1 - u \mathbf{H}_\gamma^2(u)$, Equation 1 implies

$$v(u)^{\kappa_\gamma} = \mathbf{H}_\gamma(u) v(u)^{\kappa_\gamma} - u \mathbf{H}_\gamma^2(u) v(u)^{\kappa_\gamma} - \sum_{g \leq \gamma} v(u)^{\kappa_\gamma} \mathbf{I}_g \left(\frac{u \mathbf{H}_\gamma^2(u)}{v(u)} \right)$$

and consequently

$$\begin{aligned} \mathbf{H}_\gamma(u) &= -\mathbf{H}_\gamma(u) \sum_{i=1}^{\kappa_\gamma} \binom{\kappa_\gamma}{i} 1^{\kappa_\gamma-i} (v(u)-1)^i + u \mathbf{H}_\gamma^2(u) v(u)^{\kappa_\gamma} + v(u)^{\kappa_\gamma} \\ &\quad + \sum_{g \leq \gamma} v(u)^{\kappa_\gamma} \mathbf{I}_g \left(\frac{u \mathbf{H}_\gamma^2(u)}{v(u)} \right). \end{aligned} \quad (11)$$

All coefficients of $\mathbf{H}_\gamma(u)$ in the RHS of Equation 11, are polynomials in u of degree ≥ 1 , whence any $[z^n]\mathbf{H}_\gamma(u)$ for $n \geq (\kappa_\gamma + 1)$ can be recursively computed. Accordingly, Equation 11 determines $\mathbf{H}_\gamma(u)$ uniquely. ■

4. IRREDUCIBLE SHADOWS

The bivariate generating function of irreducible shadows of genus g with m arcs is denoted by

$$\mathbf{I}(z, t) = \sum_{g \geq 1} \mathbf{I}_g(z) t^g = \sum_{g \geq 1} \sum_{m=2g}^{6g-2} \mathbf{i}_g(m) z^m t^g.$$

Let $\mathbf{c}_g(m)$ denote the number of matchings of genus g with m arcs. We have the generating function of matchings of genus g

$$\mathbf{C}_g(z) = \sum_{m \geq 2g} \mathbf{c}_g(m) z^m.$$

The bivariate generating function of matchings of genus g with m arcs is denoted by

$$\mathbf{C}(z, t) = \sum_{g \geq 0} \mathbf{C}_g(z) t^g = \sum_{g \geq 0} \sum_{m \geq 2g} \mathbf{c}_g(m) z^m t^g.$$

Theorem 2. *The generating functions $\mathbf{C}(z, t)$ and $\mathbf{I}(z, t)$ satisfy*

$$\mathbf{C}(z, t)^{-1} = 1 - \left(z \mathbf{C}(z, t) + \mathbf{C}(z, t)^{-1} \mathbf{I} \left(\frac{z \mathbf{C}(z, t)^2}{1 - z \mathbf{C}(z, t)^2}, t \right) \right),$$

equivalently,

$$\mathbf{C}(z, t) - z \mathbf{C}(z, t)^2 - \mathbf{I} \left(\frac{z \mathbf{C}(z, t)^2}{1 - z \mathbf{C}(z, t)^2}, t \right) = 1. \quad (12)$$

Proof. We distinguish the classes of blocks into two categories characterized by the unique component containing all maximal arcs (maximal component). Namely, ■

- blocks whose maximal component contains only one arc,
- blocks whose maximal component is an irreducible (nonempty) matching.

In the first case, the removal of the maximal component (one arc) generates again an arbitrary matching, which translates into the term

$$z \mathbf{C}(z, t).$$

Let $\mathbf{T}(z, t)$ denote the (genus-filtered) generating function of blocks of the second type. The decomposition of matchings into a sequence of blocks implies

$$\mathbf{C}(z, t)^{-1} = 1 - (z \mathbf{C}(z, t) + \mathbf{T}(z, t)).$$

Let σ be a fixed irreducible shadow of genus g having n arcs. Let $\mathbf{T}_\sigma(z, t)$ be the generating function of blocks, having σ as the shadow of its unique maximal component. Then we have

$$\mathbf{T}(z, t) = \sum_{\sigma \in \mathcal{J}} \mathbf{T}_\sigma(z, t),$$

where $\mathcal{J}$ denotes the set of irreducible shadows.

We shall construct $\mathcal{J}_\sigma$ in three steps using arcs, $\mathcal{R}$; sequences of arcs, $\mathcal{K}$; induced arcs, $\mathcal{N}$; sequence of induced arcs, $\mathcal{M}$; and arbitrary matchings, $\mathcal{C}$.

FIG. 7. Step I: Inflation of each arc in σ into a sequence of induced arcs.

FIG. 8. Step II: Inflation of each arc in the component with shadow σ into stacks.

Step I: We inflate each arc in σ into a sequence of induced arcs (Fig. 7). An induced arc, that is an arc together with at least one nontrivial match in either one or in both P -intervals

$$\mathcal{N} = \mathcal{R} \times ((\mathcal{C} - 1) + (\mathcal{C} - 1) + (\mathcal{C} - 1)^2) = \mathcal{R} \times (\mathcal{C}^2 - 1).$$

Clearly, we have for a single induced arc $\mathbf{N}(z, t) = z (\mathbf{C}(z, t)^2 - 1)$, guaranteed by the additivity of genus, and for a sequence of induced arcs, $\mathcal{M} = \text{SEQ}(\mathcal{N})$, where

$$\mathbf{M}(z, t) = \frac{1}{1 - z(\mathbf{C}(z, t)^2 - 1)}.$$

Inflating each arc into a sequence of induced arcs, $R^n \times \mathcal{M}^n$, gives the corresponding generating function

$$z^n \mathbf{M}(z, t)^n = \left(\frac{z}{1 - z(\mathbf{C}(z, t)^2 - 1)} \right)^n,$$

since the genus is additive.

Step II: We inflate each arc in the component with shadow σ into stacks (Fig. 8). The corresponding generating function is

$$\left(\frac{\frac{z}{1-z}}{1 - \frac{z}{1-z}(\mathbf{C}(z, t)^2 - 1)} \right)^n = \left(\frac{z}{1 - z\mathbf{C}(z, t)^2} \right)^n$$

Step III: We insert additional matchings at exactly $(2n - 1)$ σ -intervals (Fig. 9). Accordingly, the generating function is $\mathbf{C}(z, t)^{2n-1}$.

Combining these three steps and utilizing additivity of the genus, we arrive at

$$\begin{aligned} \mathbf{T}_\sigma(z, t) &= t^g \left(\frac{z}{1 - z\mathbf{C}(z, t)^2} \right)^n \mathbf{C}(z, t)^{2n-1} \\ &= t^g \mathbf{C}(z, t)^{-1} \left(\frac{z\mathbf{C}(z, t)^2}{1 - z\mathbf{C}(z, t)^2} \right)^n. \end{aligned}$$

Therefore

$$\begin{aligned} \mathbf{T}(z, t) &= \sum_{\sigma \in \mathcal{J}} \mathbf{T}_\sigma(z, t) \\ &= \sum_{g, n} \mathbf{i}_g(n) t^g \mathbf{C}(z, t)^{-1} \left(\frac{z\mathbf{C}(z, t)^2}{1 - z\mathbf{C}(z, t)^2} \right)^n. \end{aligned}$$

We derive

$$\mathbf{T}(z, t) = \mathbf{C}(z, t)^{-1} \mathbf{I} \left(\frac{z\mathbf{C}(z, t)^2}{1 - z\mathbf{C}(z, t)^2}, t \right),$$

completing the proof of Equation 12. ■

Now we can derive a recursion for $\mathbf{I}_g(z)$ from Theorem 2.

FIG. 9. Step III: Insertion of additional matchings at exactly $(2n - 1)$ σ -intervals.

Corollary 1. For $g \geq 1$, $\mathbf{I}_g(z)$ satisfies the following recursion

$$\begin{aligned} \mathbf{I}_g(z) = & \mathbf{C}_g(\theta(z)) - \theta(z) \sum_{i=0}^g \mathbf{C}_i(\theta(z)) \mathbf{C}_{g-i}(\theta(z)) \\ & - \sum_{j=1}^{g-1} [t^{g-j}] \mathbf{I}_j \left(\frac{\theta(z) \left(\sum_{k=0}^{g-j} \mathbf{C}_k(\theta(z)) t^k \right)^2}{1 - \theta(z) \left(\sum_{k=0}^{g-j} \mathbf{C}_k(\theta(z)) t^k \right)^2} \right), \end{aligned}$$

where $\theta(z) = \frac{z(z+1)}{(2z+1)^2}$.

Proof. We compute the coefficient of t^g on both sides of Equation 12

$$\begin{aligned} [t^g] \mathbf{C}(t, z) - z [t^g] \mathbf{C}(t, z)^2 - [t^g] \mathbf{I} \left(\frac{z \mathbf{C}(t, z)^2}{1 - z \mathbf{C}(t, z)^2}, t \right) &= 0 \\ \mathbf{C}_g(z) - z \sum_{i=0}^g \mathbf{C}_i(z) \mathbf{C}_{g-i}(z) - \sum_{j=1}^g [t^{g-j}] \mathbf{I}_j \left(\frac{z \mathbf{C}(t, z)^2}{1 - z \mathbf{C}(t, z)^2} \right) &= 0 \\ \mathbf{C}_g(z) - z \sum_{i=0}^g \mathbf{C}_i(z) \mathbf{C}_{g-i}(z) - \sum_{j=1}^{g-1} [t^{g-j}] \mathbf{I}_j \left(\frac{z \mathbf{C}(t, z)^2}{1 - z \mathbf{C}(t, z)^2} \right) \\ &= [t^0] \mathbf{I}_g \left(\frac{z \mathbf{C}(t, z)^2}{1 - z \mathbf{C}(t, z)^2} \right) \end{aligned}$$

Note that

$$[t^{g-j}] \mathbf{I}_j \left(\frac{z \mathbf{C}(t, z)^2}{1 - z \mathbf{C}(t, z)^2} \right) = [t^{g-j}] \mathbf{I}_j \left(\frac{z \left(\sum_{k=0}^{g-j} \mathbf{C}_k(z) t^k \right)^2}{1 - z \left(\sum_{k=0}^{g-j} \mathbf{C}_k(z) t^k \right)^2} \right).$$

Hence,

$$\begin{aligned} \mathbf{C}_g(z) - z \sum_{i=0}^g \mathbf{C}_i(z) \mathbf{C}_{g-i}(z) - \sum_{j=1}^{g-1} [t^{g-j}] \mathbf{I}_j \left(\frac{z \left(\sum_{k=0}^{g-j} \mathbf{C}_k(z) t^k \right)^2}{1 - z \left(\sum_{k=0}^{g-j} \mathbf{C}_k(z) t^k \right)^2} \right) \\ = \mathbf{I}_g \left(\frac{z \mathbf{C}_0(z)^2}{1 - z \mathbf{C}_0(z)^2} \right) \end{aligned}$$

Setting $y = \frac{z \mathbf{C}_0(z)^2}{1 - z \mathbf{C}_0(z)^2}$, we have $z = \theta(y) = \frac{y(y+1)}{(2y+1)^2}$. Then we derive

$$\begin{aligned} \mathbf{I}_g(y) = & \mathbf{C}_g(\theta(y)) - \theta(y) \sum_{i=0}^g \mathbf{C}_i(\theta(y)) \mathbf{C}_{g-i}(\theta(y)) \\ & - \sum_{j=1}^{g-1} [t^{g-j}] \mathbf{I}_j \left(\frac{\theta(y) \left(\sum_{k=0}^{g-j} \mathbf{C}_k(\theta(y)) t^k \right)^2}{1 - \theta(y) \left(\sum_{k=0}^{g-j} \mathbf{C}_k(\theta(y)) t^k \right)^2} \right) \end{aligned}$$

completing the proof. ■

A seminal result (Harer and Zagier 1986), computes a recursion and generating function for the number $\mathbf{c}_g(m)$ as follows:

Lemma 1. (Harer and Zagier 1986) The $\mathbf{c}_g(m)$ satisfy the recursion

$$(m+1) \mathbf{c}_g(m) = 2(2m-1) \mathbf{c}_g(m-1) + (2m-1)(m-1)(2m-3) \mathbf{c}_{g-1}(m-2), \quad (13)$$

where $\mathbf{c}_g(m) = 0$ for $2g > m$.

The recursion Equation 13 is equivalent to the ODE

$$z(1-4z) \frac{d}{dz} \mathbf{C}_g(z) + (1-2z) \mathbf{C}_g(z) = \Phi_{g-1}(z) \quad (14)$$

where

$$\Phi_{g-1}(z) = z^2 \left(4z^3 \frac{d^3}{dz^3} \mathbf{C}_{g-1}(z) + 24z^2 \frac{d^2}{dz^2} \mathbf{C}_{g-1}(z) + 27z \frac{d}{dz} \mathbf{C}_{g-1}(z) + 3\mathbf{C}_{g-1}(z) \right)$$

with initial condition $\mathbf{C}_g(0) = 0$ since $r = n + 1 - 2g$ has no positive solution $r > 0$ for $n < 2g$. Therefore, we can recursively compute $\mathbf{C}_g(z)$ by solving Equation (14) via Maple.

Theorem 3. (Andersen et al., 2013) For any $g \geq 1$ the generating function $\mathbf{C}_g(z)$ is given by

$$\mathbf{C}_g(z) = Q_g(z) \frac{\sqrt{1-4z}}{(1-4z)^{3g}}, \quad (15)$$

where $Q_g(z)$ is a polynomial with integral coefficients of degree at most $(3g - 1)$, $Q_g(1/4) \neq 0$, $[z^{2g}]Q_g(z) \neq 0$ and $[z^h]Q_g(z) = 0$ for $0 \leq h \leq 2g - 1$.

The recursion Equation (14) permits the calculation of the polynomials $Q_g(z)$, the first five of which are given as follows (Andersen et al., 2013):

$$\begin{aligned} Q_1(z) &= z^2, \\ Q_2(z) &= 21z^4(z+1) \\ Q_3(z) &= 11z^6(158z^2 + 558z + 135), \\ Q_4(z) &= 143z^8(2339z^3 + 18378z^2 + 13689z + 1575), \\ Q_5(z) &= 88179z^{10}(1354z^4 + 18908z^3 + 28764z^2 + 9660z + 675). \end{aligned}$$

Applying Corollary 1 together with the generating function $\mathbf{C}_g(z)$, we recursively compute $\mathbf{I}_g(z)$. For example, for $g = 1$,

$$\begin{aligned} \mathbf{I}_1(z) &= \mathbf{C}_1(\theta(z)) - 2\theta(z) \mathbf{C}_0(\theta(z))\mathbf{C}_1(\theta(z)) \\ &= z^2(1+z)^2. \end{aligned}$$

For $1 \leq g \leq 8$, we list $\mathbf{I}_g(z)$ as follows:

$$\begin{aligned} \mathbf{I}_1(z) &= z^2(1+z)^2, \\ \mathbf{I}_2(z) &= z^4(1+z)^4(17+92z+96z^2) \\ \mathbf{I}_3(z) &= z^6(1+z)^6(1259+15928z+61850z^2+92736z^3+47040z^4) \\ \mathbf{I}_4(z) &= z^8(1+z)^8(200589+4245684z+31264164z^2+107622740z^3 \\ &\quad + 188262816z^4+161967360z^5+54333440z^6) \\ \mathbf{I}_5(z) &= z^{10}(1+z)^{10}(54766516+1681752448z+19092044658z^2 \\ &\quad + 109184482584z^3+353376676011z^4 \\ &\quad + 675135053568z^5+753610999040z^6 \\ &\quad + 453941596160z^7+113867919360z^8) \\ \mathbf{I}_6(z) &= 3z^{12}(1+z)^{12}(7613067765+312905543772z+4932317894440z^2 \\ &\quad + 40797413383380z^3+200964285178270z^4+626595744773516z^5 \\ &\quad + 1268150755326432z^6+1660845652501760z^7+1357241056522240z^8 \\ &\quad + 628740761518080z^9+126004558299136z^{10}) \end{aligned}$$

$$\begin{aligned}
\mathbf{I}_7(z) &= z^{14}(1+z)^{14}(13532959408258 + 706557271551408 z \\
&\quad + 14506513039164060 z^2 + 160434554727348896 z^3 \\
&\quad + 1089075339931680039 z^4 + 4857650169218369856 z^5 \\
&\quad + 14771712773087154704 z^6 + 31138771188689736192 z^7 \\
&\quad + 45486763075779571200 z^8 + 45167296685229793280 z^9 \\
&\quad + 29078583024627105792 z^{10} \\
&\quad + 10941912454886326272 z^{11} + 1826131581135486976 z^{12}) \\
\mathbf{I}_8(z) &= z^{16}(1+z)^{16}(10826939105517381 + 692156096364848676 z \\
&\quad + 17724869034206737356 z^2 \\
&\quad + 249069951630509297956 z^3 + 2192230050291936695620 z^4 \\
&\quad + 12980362620620450943588 z^5 + 53923920139564145104556 z^6 \\
&\quad + 161060520394034807160164 z^7 + 349969438514715552162336 z^8 \\
&\quad + 553647075623879302120960 z^9 + 630641488385967162351616 z^{10} \\
&\quad + 503519879227179011162112 z^{11} + 267275771110990512783360 z^{12} \\
&\quad + 84670509266097640833024 z^{13} + 12107536630199227514880 z^{14})
\end{aligned}$$

We conjecture that the polynomial $\mathbf{I}_g(z)$, for arbitrary g , has $z^{2g} (1+z)^{2g}$ as a factor.

5. ASYMPTOTICS OF γ -MATCHINGS

Let us begin recalling the following result of Flajolet and Sedgewick (2009):

Theorem 4. *Let $y(u) = \sum_{n \geq 0} y_n u^n$ be a generating function, analytic at 0, satisfying a polynomial equation $\Phi(u, y) = 0$. Let ρ be the real dominant singularity of $y(u)$. Define the resultant of $\Phi(u, y)$ and $\frac{\partial}{\partial y} \Phi(u, y)$ as polynomial in y*

$$\Delta(u) = \mathbf{R} \left(\Phi(u, y), \frac{\partial}{\partial y} \Phi(u, y), y \right).$$

(1) *The dominant singularity ρ is unique and a root of the resultant $\Delta(u)$ and there exists $\pi = y(\rho)$, satisfying the system of equations,*

$$\Phi(\rho, \pi) = 0, \quad \Phi_y(\rho, \pi) = 0. \quad (16)$$

(2) *If $\Phi(u, y)$ satisfies the conditions:*

$$\Phi_u(\rho, \pi) \neq 0, \quad \Phi_{yy}(\rho, \pi) \neq 0, \quad (17)$$

then $y(u)$ has the following expansion at ρ

$$y(u) = \pi + \lambda(\rho - u)^{\frac{1}{2}} + O(\rho - u),$$

for some nonzero constant λ

Further the coefficients of $y(u)$ satisfy

$$[u^n]y(u) \sim c n^{-\frac{3}{2}} \rho^{-n}, \quad n \rightarrow \infty,$$

for some constant $c > 0$.

Proof. The proof of (1) can be found in Flajolet and Sedgewick (2009) or Hille (1962, pp. 103). To prove (2), let $\Psi(u, y) = \Phi(\rho - u, \pi - y)$. Immediately, we have $\Psi(0, 0) = 0$. Puiseux's theorem (Wall, 2004) guarantees a solution of $y - \pi$ in terms of a Puiseux series in $\rho - u$. Note that Equations (16) and (17) are equivalent to

$$\Psi(0, 0) = 0, \quad \Psi_y(0, 0) = 0, \quad \Psi_u(0, 0) \neq 0, \quad \Psi_{yy}(0, 0) \neq 0.$$

Then we apply Newton's polygon method to determine the type of expansion and find the first exponent of u to be $\frac{1}{2}$. Therefore, the Puiseux series expansion of $y(u)$ has the required form. The asymptotics of the coefficients follow from Equation (18) as a straightforward application of the transfer theorem (Flajolet and Sedgewick, 2009, pp. 389, Theorem VI.3). ■

Combining Theorem 1 and Theorem 4, the asymptotic analysis of $\mathbf{H}_\gamma(u)$ follows.

Theorem 5. For $1 \leq \gamma \leq 10$, let

$$\Delta_\gamma(u) = \mathbf{R} \left(P_\gamma(u, X), \frac{\partial}{\partial X} P_\gamma(u, X), X \right)$$

the resultant of $P_\gamma(u, X)$ and $\frac{\partial}{\partial X} P_\gamma(u, X)$ as polynomials in X , and ρ_γ denote the real dominant singularity of $\mathbf{H}_\gamma(u)$.

(a) The dominant singularity ρ_γ is unique and a root of $\Delta_\gamma(u)$,

(b) at ρ_γ we have

$$\mathbf{H}_\gamma(u) = \pi_\gamma + \lambda_\gamma (\rho_\gamma - u)^{\frac{1}{2}} + O(\rho_\gamma - u), \text{ for some nonzero constant } \lambda_\gamma.$$

(c) the coefficients of $\mathbf{H}_\gamma(u)$ are asymptotically given by

$$[u^n] \mathbf{H}_\gamma(u) \sim c_\gamma n^{-3/2} \rho_\gamma^{-n}$$

for some $c_\gamma > 0$.

Proof. Pringsheim's theorem (Flajolet and Sedgewick, 2009), pp. 240) guarantees that for any γ , $\mathbf{H}_\gamma(u)$ has a dominant real singularity $\rho_\gamma > 0$. To prove the singular expansion of the function and asymptotic of the coefficients, we verify $P_\gamma(u, X)$, for $1 \leq \gamma \leq 10$, satisfy the condition of Theorem 4, and the results follow. ■

6. COMBINATORICS OF γ -DIAGRAMS

Lemma 2. For any $\gamma \geq 1$, we have

$$\mathbf{S}_\gamma(u, e) = \frac{1+u}{1+2u-ue} \mathbf{H}_\gamma \left(\frac{u(1+u)}{(1+2u-ue)^2} \right). \quad (19)$$

The proof of Lemma 2 can be obtained by standard symbolic method.

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

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

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

Our main result about enumerating τ -canonical γ -structures follows.

Theorem 6. Suppose $\gamma, \tau \geq 1$ and let $u_\tau(z) = \frac{(z^2)^\tau - 1}{z^{2\tau} - z^2 + 1}$. Then the generating function $\mathbf{G}_{\tau, \gamma}(z)$ is algebraic and given by

$$\mathbf{G}_{\tau, \gamma}(z) = \frac{1}{u_\tau(z)z^2 - z + 1} \mathbf{H}_\gamma \left(\frac{u_\tau(z)z^2}{(u_\tau(z)z^2 - z + 1)^2} \right). \quad (20)$$

TABLE 1. THE EXPONENTIAL GROWTH RATES OF $\rho_{s,i}^{-1}$, FOR $1 \leq s, i \leq 2$

(s, i)	(1, 1)	(2, 1)	(1, 2)	(2, 2)
$\rho_{s,i}^{-1}$	3.6005	2.2759	3.8846	2.3553

In particular for $1 \leq s, i \leq 2$ we have

$$[z^n] \mathbf{G}_{s,i}(z) \sim k_{s,i} n^{-\frac{3}{2}} (\rho_{s,i}^{-1})^n,$$

for some constants $k_{s,i} > 0$, for $\rho_{s,i}^{-1}$, we have Table 1.

Proof. Since each γ -diagram has a unique γ -shape, λ , having some number $m \geq 0$ of 1-arcs, we have

$$\mathbf{G}_{\tau,\gamma}(z) = \sum_{m \geq 0} \sum_{\substack{\lambda \text{ } \gamma\text{-shape} \\ \text{having } m \text{ 1-arcs}}} \mathbf{G}_{\tau,\gamma}^{\lambda}(z). \quad (21)$$

According to Lemma 3, $\mathbf{G}_{\tau,\gamma}^{\lambda}(z)$ only depends on the number of arcs and 1-arcs of λ , and we can therefore express

$$\begin{aligned} \mathbf{G}_{\tau,\gamma}(z) &= \frac{1}{z-1} \mathbf{S}_{\gamma} \left(\frac{z^{2\tau}}{(1-z^2)(1-z)^2 - (2z-z^2)z^{2\tau}}, z \right) \\ &= \frac{1}{(1-z) + u_{\tau}(z)z^2} \mathbf{H}_{\gamma} \left(\frac{z^2 u_{\tau}(z)}{((1-z) + u_{\tau}(z)z^2)^2} \right), \end{aligned}$$

using Lemma 2 in order to confirm Equation 23, where the second equality follows from direct computation. Let

$$\theta_{\tau}(z) = \frac{z^2 u_{\tau}(z)}{((1-z) + u_{\tau}(z)z^2)^2}$$

denote the argument of $\mathbf{H}_{\gamma}$ in this expression. By definition we have $\theta(z) \in \mathbb{C}(z)$. Since $\theta_{\sigma}(0) = 0$ the composition $\mathbf{H}_{\gamma}(\theta(z))$ is well defined as a power series. Obviously, $P_{\gamma}(z, \mathbf{H}_{\gamma}(z)) = 0$ guarantees $P_{\gamma}(\theta_{\tau}(z), \mathbf{H}_{\gamma}(\theta_{\tau}(z))) = 0$. We have the following Hasse diagram of fields:

from which we immediately conclude that $\mathbf{G}_{\tau,\gamma}(z)$ is algebraic. Pringsheim's theorem (Flajolet and Sedgewick, 2009) guarantees that for any γ , $\tau \geq 1$, $G_{\tau,\gamma}(z)$ has a dominant real singularity $\rho_{\tau,\gamma} > 0$.

According to Theorem 5 we have

$$\mathbf{H}_i(z) = \pi_1 + \sum_{j \geq 1} a_{j,i} \left((\mu_i - z)^{1/2} \right)^j \quad \text{and} \quad [z^n] \mathbf{H}_i(z) \sim k_i n^{-3/2} (\mu_i^{-1})^n.$$

For $\tau = 1, 2$, we verify directly that $\rho_{1,i}$ and $\rho_{2,i}$ are the unique solutions of minimum modulus of $\theta_1(z) = \mu_i$ and $\theta_2(z) = \mu_i$. These solutions are strictly smaller than any other singularities of $\theta_1(z)$ and $\theta_2(z)$ and furthermore satisfy $\theta'_1(\rho_{1,i}) \neq 0$ as well as $\theta'_2(\rho_{2,i}) \neq 0$. It follows that $\mathbf{G}_{1,i}(z)$ and $\mathbf{G}_{2,i}(z)$ are governed by the supercritical paradigm (Flajolet and Sedgewick, 2009), which in turn implies

$$[z^n] \mathbf{G}_{s,i}(z) \sim k_{s,i} n^{-3/2} (\rho_{s,i}^{-1})^n, \quad (22)$$

where $s = 1, 2$ and $k_{s,i}$ is some positive constant. ■

Theorem 6 has its analogue for τ -canonical, γ -diagrams containing 1-arcs. The asymptotic formula in case of $\tau = 1$, $\gamma = 1$,

$$[z^n]\tilde{\mathbf{G}}_{1,1}(z) \sim j_1 n^{-\frac{3}{2}}(\varrho_{1,1}^{-1})^n$$

is due to Nebel and Weinberg (2011), who used the explicit grammar developed in Reidys et al. (2011) to obtain an algebraic equation for $\tilde{\mathbf{G}}_{1,1}(z)$.

Corollary 2. Suppose γ , $\tau \geq 1$ and let $u_\tau(z) = \frac{(z^2)^\tau - 1}{z^{2\tau} - z^2 + 1}$. Then the generating function of τ -canonical γ -diagrams containing 1-arcs, $\tilde{\mathbf{G}}_{\tau,\gamma}(z)$, is algebraic and

$$\tilde{\mathbf{G}}_{\tau,\gamma}(z) = \mathbf{H}_\gamma \left(\frac{u_\tau(z)z^2}{(1-z)^2} \right). \quad (23)$$

In particular for $\gamma = 1$, we have

$$[z^n]\tilde{\mathbf{G}}_{1,1}(z) \sim j_1 n^{-\frac{3}{2}}(\varrho_{1,1}^{-1})^n, \quad \text{and} \quad [z^n]\tilde{\mathbf{G}}_{2,1}(z) \sim j_2 n^{-\frac{3}{2}}(\varrho_{2,1}^{-1})^n$$

for some constants j_1, j_2 , where $\varrho_{1,1}^{-1} = 3.8782$ and $\varrho_{2,1}^{-1} = 2.3361$.

Proof. Let λ be a fixed γ -shape with $s \geq 1$ -arcs and $m \geq 0$ 1-arcs. Then the generating function of τ -canonical γ -diagrams containing 1-arcs that have shape λ containing 1-arcs is given by

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

■

7. DISCUSSION

The symbolic approach based on γ -matchings allows not only for computing the generating function of canonical γ -structures. On the basis of Theorem 6, it is possible to obtain a plethora of statistics of γ -structures by means of combinatorial markers.

For instance, we can analogously compute the bivariate generating function of τ canonical γ -structures over n vertices, containing exactly m arcs, $\mathbf{A}_{\tau,\gamma}(z, t)$ as

$$\mathbf{A}_{\tau,\gamma}(z, t) = \frac{1}{u_\tau(z, t)z^2 - z + 1} \mathbf{H}_\gamma \left(\frac{u_\tau(z, t)z^2}{(u_\tau(z, t)z^2 - z + 1)^2} \right) \quad (24)$$

where $u_\tau(z, t)$ is given by

$$u_\tau(z, t) = \frac{t(tz^2)^{\tau-1}}{(tz^2)^\tau - tz^2 + 1}.$$

This bivariate generating function is the key to obtain a central limit theorem for the distribution of arc-numbers in γ -structures (Bender, 1973) on the basis of the Lévy-Cramér theorem on limit distributions (Feller, 1991).

Statistical properties of γ -structures play a key role for quantifying algorithmic improvements via sparsifications (Busch et al., 2008; Möhl et al., 2010; Wexler, 2007). The key property here is the *polymer-zeta property* (Kabakcioglu and Stella, 2008; Kafri et al., 2000), which states that the probability of an arc of length ℓ is bounded by $k\ell^c$, where k is some positive constant and $c > 1$. Polymer-zeta stems from the theory of self-avoiding walks (Vanderzande, 1998) and has only been empirically established for the simplest class of RNA structures, namely those of genus zero. It turns out, however, that the polymer-zeta property is genuinely a combinatorial property of a structure class. Moreover, our results allow for quantifying the effect of sparsifications of folding algorithms into γ -structures (Andersen et al., 2012; Huang and Reidys, 2012).

We finally remark that around 98% of RNA pseudoknot structures catalogued in databases are, in fact, canonical 1-structures. RNA pseudoknot structures like the HDV virus exhibiting irreducible shadows of genus two are relatively rare.

ACKNOWLEDGMENTS

We thank Fenix W.D. Huang for discussions and comments. We furthermore acknowledge the financial support of the Future and Emerging Technologies (FET) program within the Seventh Framework Programme (FP7) for Research of the European Commission, under the FET-Proactive grant agreement TOPDRIM, number FP7-ICT-318121.

AUTHOR DISCLOSURE STATEMENT

The authors declare that no competing financial interests exist.

REFERENCES

- Andersen, J.E., Huang, F.W.D., Penner, R.C., et al. 2012. Topology of RNA-interaction structures. *J. Comput. Biol.* 19, 928–943.
- Andersen, J.E., Penner, R.C., Reidys, C.M., et al. 2013. Topological classification and enumeration of rna structures by genus. *J. Math. Biol.* 67, 1261–1278.
- Bailor, M.H., Sun, X., and Al-Hashimi, H.M. 2010. Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science* 327, 202–206.
- Bender, E.A. 1973. Central and local limit theorems applied to asymptotic enumeration. *J. Combin. Theory A* 15, 91–111.
- Bon, M., Vernizzi, G., Orland, H., et al. 2008. Topological classification of RNA structures. *J. Mol. Biol.* 379, 900–911.
- Busch, A., Richter, A.S., and Backofen, R. 2008. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* 24, 2849–2856.
- Dyson, F.J. 1949. The S matrix in quantum electrodynamics. *Phys. Rev.* 75, 1736–1755.
- Feller, W. 1991. *An Introduction to Probability Theory and Its Application*. Addison-Wesley Publishing Company Inc., New York.
- Flajolet, P., and Sedgewick, R. 2009. *Analytic Combinatorics*. Cambridge University Press, New York.
- Harer, J., and Zagier, D. 1986. The euler characteristic of the moduli space of curves. *Invent. Math.* 85, 457–486.
- Hille, E. 1962. *Analytic Function Theory, Volume II*. Chelsea Publishing Company.
- Huang, F.W.D., and Reidys, C.M. 2012. On the combinatorics of sparsification. *Algorithm. Mol. Biol.* 7, 28.
- Kabakcioglu, A., and Stella, A.L. 2008. A scale-free network hidden in the collapsing polymer. *Bioinformatics*.
- Kafri, Y., Mukamel, D., and Peliti, L. 2000. Why is the DNA denaturation transition first order? *Phys. Rev. Lett.* 85, 4988–4991.
- Loebl, M., and Moffatt, I. 2008. The chromatic polynomial of fatgraphs and its categorification. *Adv. Math.* 217, 1558–1587.
- Massey, W.S. 1967. *Algebraic Topology: An Introduction*. Springer-Verlag, New York.
- Möhl, R., Salari, R., Will, S., et al. 2010. Sparsification of RNA structure prediction including pseudoknots. *Algorithms for Molecular Biology* 5, 39.
- Nebel, M.E., and Weinberg, F. 2011. Algebraic and combinatorial properties of common RNA pseudoknot classes with applications. *J. Comput. Biol.* 19, 1134–1150.
- Orland, H., and Zee, A. 2002. RNA folding and large n matrix theory. *Nuclear Physics B* 620, 456–476.
- Penner, R.C. 2004. Cell decomposition and compactification of Riemann’s moduli space in decorated Teichmüller theory, 263–301. In Tongring N., and Penner, R.C., eds. *Woods Hole Mathematics-Perspectives in Math and Physics*, pages 263–301. World Scientific, Singapore.
- Penner, R.C., Knudsen, M., Wiuf, C., et al. 2010. Fatgraph models of proteins. *Comm. Pure Appl. Math.* 63, 1249–1297.
- Penner, R.C., and Waterman, M.S. 1993. Spaces of RNA secondary structures. *Adv. Math.* 101, 31–49.
- Reidys, C.M., Huang, F.W.D., Andersen, J.E., et al. 2011. Topology and prediction of RNA pseudoknots. *Bioinformatics* 27, 1076–1085.
- Smith, T.F., and Waterman, M.S. 1978. RNA secondary structure. *Math. Biol.* 42, 31–49.
- Tabaska, J.E., Cary, R.B., Gabow, H.N., et al. 1998. An RNA folding method capable of identifying pseudoknots and base triples. *Bioinformatics* 14, 691–699.

- Vanderzande, C. 1998. *Lattice Models of Polymers*. Cambridge University Press, New York.
- Wall, C.T.C. 2004. *Singular Points of Plane Curves*. Cambridge University Press, New York.
- Wexler, Y., Zilberstein, C., Ziv-ukelson, M. 2007. A study of accessible motifs and RNA folding complexity. *J. Comput. Biol.* 14, 6.

Address correspondence to:
Prof. Christian Michael Reidys
Department of Mathematics and
Computer Science
University of Southern Denmark
Campusvej 55
DK-5230, Odense M
Denmark

E-mail: duck@santafe.edu