CANONICAL RNA PSEUDOKNOT STRUCTURES

GANG MA AND CHRISTIAN M. REIDYS *

ABSTRACT. In this paper we study k-noncrossing, canonical RNA pseudoknot structures with minimum arc-length ≥ 4. Let $\mathsf{T}_{k,\sigma}^{[4]}(n)$ denote the number of these structures. We derive exact enumeration results by computing the generating function $\mathbf{T}_{k,\sigma}^{[4]}(z) = \sum_n \mathsf{T}_{k,\sigma}^{[4]}(n)z^n$ and derive the asymptotic formulas $\mathsf{T}_{k,3}^{[4]}(n) \sim c_k\, n^{-(k-1)^2-\frac{k-1}{2}}(\gamma_{k,3}^{[4]})^{-n}$ for $k = 3, \ldots, 9$. In particular we have for $k = 3$, $\mathsf{T}_{3,3}^{[4]}(n) \sim c_3\, n^{-5}2.0348^n$. Our results prove that the set of biophysically relevant RNA pseudoknot structures is surprisingly small and suggest a new structure class as target for prediction algorithms.

1. INTRODUCTION

RNA pseudoknot structures have drawn a lot of attention over the last decade [1]. From micro-RNA binding to ribosomal frameshifts [20], we currently discover novel RNA functionalities at truly amazing rates. Our conceptional understanding of RNA pseudoknot structures has not kept up with this pace. Only recently the generating functions of k-noncrossing RNA structures of arc-length ≥ 2 [11], arc-length ≥ 4 [9] and canonical k-noncrossing RNA structures of arc-length ≥ 2 [13] have been derived. While these combinatorial results open new perspectives for the design of new folding algorithms, it has to be noted that realistic pseudoknot structures are subject to a minimum arc-length ≥ 4 and stack-length ≥ 3. Therefore the above structure classes are not "best possible". The lack of a transparent target class of RNA pseudoknot structures represents a problem for *ab initio* prediction algorithms. There are four algorithms, capable of the energy based prediction of certain pseudoknots in polynomial time: Rivas *et al.* (dynamic programming,

gap-matrices, $O(n^6)$ time and $O(n^4)$ space) [21], Uemura *et al.* ($O(n^5)$ time and $O(n^4)$ space, tree-adjoining grammars) [25], Akutsu [3] and Lyngso [18]. All of them follow the dynamic programming paradigm and none produces an easily specifiable class of pseudoknots as output.

In this paper we characterize a class of pseudoknot RNA structures in which bonds have a minimum length of four and stacks contains at least three base pairs. Our results show that this structure class is ideally suited as *a priori*-output for prediction algorithms. Tab.1 indicates that this class remains suitable even for more complex pseudoknots (specified in terms of larger sets of *mutually crossing* bonds). In fact, one can search RNA 3-noncrossing pseudoknot structure with arc-length ≥ 4 and stack-length $\sigma \geq 3$ for a sequence of length 100 w.r.t. a variety of objective functions (in particular loop-based minimum free energy models) on a 4-core PC in a few minutes [10].

In order to put our results into context, we turn the clock back by almost three decades. 1978 M. Waterman *et al.* [27, 28, 29, 30] began deriving the concepts for enumeration and prediction of RNA secondary structures. The latter represent arguably *the* prototype of prediction-targets of RNA structures. RNA secondary structures are coarse grained structures which can be represented as outer-planar graphs, diagrams, Motzkin-paths or words over ".", "(" and ")". Their decisive feature is that they have no two crossing bonds, see Fig.1. Let $\mathsf{T}_2^{[\lambda]}(n)$ denote the number of secondary structures with arc-length $\geq \lambda$ over $[n] = \{1, \ldots, n\}$. The key to RNA secondary structures is the following recursion for $\mathsf{T}_2^{[\lambda]}(n)$:

$$(1.1) \qquad \mathsf{T}_2^{[\lambda]}(n) = \mathsf{T}_2^{[\lambda]}(n-1) + \sum_{j=0}^{n-(\lambda+1)} \mathsf{T}_2^{[\lambda]}(n-2-j)\mathsf{T}_2^{[\lambda]}(j),$$

where $\mathsf{T}_2^{[\lambda]}(n) = 1$ for $0 \leq n \leq \lambda$. The latter follows from considering the concatenation of Motzkin-paths with minimum peak length $\lambda - 1$. Eq. (1.1) implies for the generating function $\mathbf{T}_2^{[\lambda]}(z) = \sum_{n\geq 0} \mathsf{T}_2^{[\lambda]}(n)z^n$ the functional equation

$$(1.2) \qquad z^2\,\mathbf{T}_2^{[\lambda]}(z)^2 - (1 - z + z^2 + \cdots + z^\lambda)\mathbf{T}_2^{[\lambda]}(z) + 1 = 0$$

from which eventually

$$\mathbf{T}_2^{[\lambda]}(z) = \frac{-1 + 2z - 2z^2 + z^{\lambda+1} + \sqrt{1 - 4z + 4z^2 - 2z^{\lambda+1} + 4z^{\lambda+2} - 4z^{\lambda+3} + z^{2\lambda+2}}}{2(z^3 - z^2)}$$

follows. Therefore, minimum arc-length restrictions do not impose particular difficulties for RNA secondary structures. In fact minimum stack size conditions can also be dealt with straightforwardly. We furthermore note that eq. (1.1) is a *constructive* recursion, i.e. it allows to inductively build secondary structures over $[n]$ from those over $[i]$, for all $i < n$.
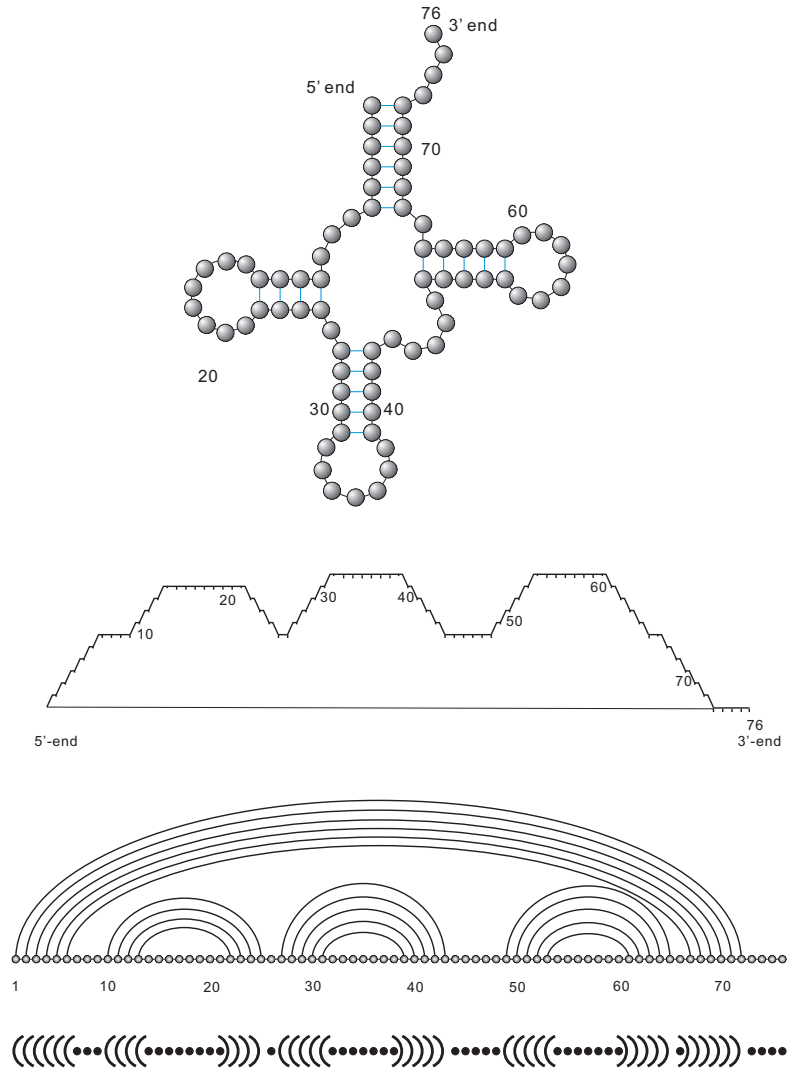
FIGURE 1. RNA secondary structures.

In order to analyze RNA structure with crossing bonds, we recall the notion of k-noncrossing diagrams [11]. A k-noncrossing diagram is a labeled graph over the vertex set $[n]$ with vertex degrees ≤ 1, represented by drawing its vertices $1, \ldots, n$ in a horizontal line and its arcs (i, j), where $i < j$, in the upper half-plane, containing at most $k - 1$ mutually crossing arcs. The vertices and arcs correspond to nucleotides and Watson-Crick (**A-U**, **G-C**) and (**U-G**) base pairs,
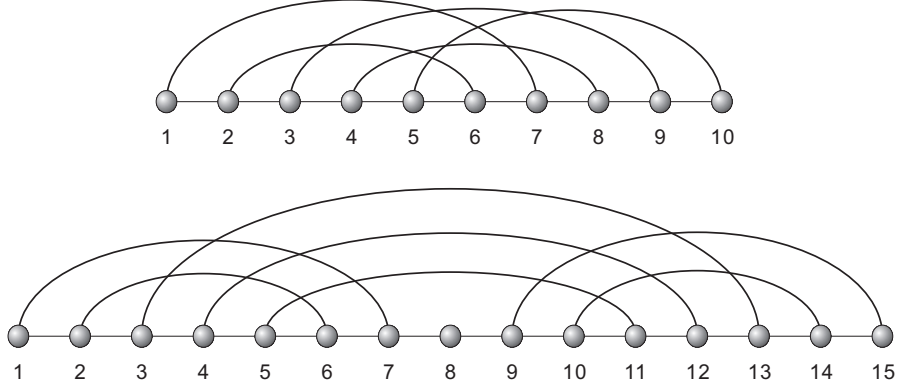
FIGURE 2. k-noncrossing diagrams: we display a 4-noncrossing, arc-length $\lambda \geq 4$ and $\sigma \geq 1$ (upper) and 3-noncrossing, $\lambda \geq 4$ and $\sigma \geq 2$ (lower) diagram.

respectively. Diagrams have the following three key parameters: the maximum number of mutually crossing arcs, $k-1$, the minimum arc-length, λ and minimum stack-length, σ ($\langle k, \lambda, \sigma \rangle$-diagrams). The length of an arc (i,j) is $j-i$ and a stack of length σ is a sequence of "parallel" arcs of the form $((i,j), (i+1, j-1), \ldots, (i+(\sigma-1), j-(\sigma-1)))$, see Fig.2. We call an arc of length λ a λ-arc. Let $T_{k,\sigma}^{[\lambda]}(n)$ denote the set of k-noncrossing diagrams with minimum arc- and stack-length λ and σ and let $\mathsf{T}_{k,\sigma}^{[\lambda]}(n)$ denote their number.

In the following, we shall identify pseudoknot RNA structures with k-noncrossing diagrams and refer to them as $\langle k, \lambda, \sigma \rangle$-structures. Pseudoknot RNA structures occur in functional RNA (RNAseP) [17], ribosomal RNA [16] and plant viral RNAs and vitro RNA evolution experiments have produced families of RNA structures with pseudoknot motifs [24]. In Fig.3 we give several representations of the UTR-pseudoknot of the mouse hepatitis virus. Due to the crossings of arcs pseudoknots differs considerably from secondary structures: pseudoknot RNA structures are inherently non-inductive and no analogue of eq. (1.1) exists. One key for the generating function of k-noncrossing RNA structures $\mathbf{T}_k^{[\lambda]}(z)$ was the bijection of Chen *et al.* [4] obtained in the context of k-noncrossing partitions. This bijection has been generalized to k-noncrossing tangled diagrams [5], a class of contact-structures tailored for expressing RNA tertiary interactions. Via the bijection k-noncrossing RNA structures can be identified with certain walks in $\mathbb{Z}^{k-1}$ that remain in the region

$$\{(x_1, \ldots, x_{k-1}) \in \mathbb{Z}^{k-1} \mid x_1 \geq x_2 \geq \ldots x_{k-1} \geq 0\}$$

FIGURE 3. UTR-pseudoknot structure of the mouse hepatitis virus.

starting and ending at 0, the boundaries of which are called walls. The enumeration of these walks is obtained employing the reflection principle. This method is due to André in 1887 [2] and has subsequently been generalized by Gessel and Zeilberger [7]. In the reflection principle "bad"-i.e. reflected- walks cancel themselves. In other words one enumerates all walks and due to cancellation only the ones survive that never touch the walls. Despite its beauty this method does not trigger any algorithmic intuition and is nonconstructive. Moreover, k-noncrossing RNA structures cannot directly be enumerated via the reflection principle: it does not preserve a minimum arc-length. In [11] it is shown how to eliminate specific classes of arcs after reflection. One nontrivial implication of this theory is that all generating functions for k-noncrossing RNA structures are D-finite, i.e. there exists a *nonconstructive* recurrence relation of finite length with polynomial coefficients for $\mathsf{T}_{k,\sigma}^{[\lambda]}(n)$. Note however, that although we can prove the existence of this recurrence

it is at present not known for *any* $k > 2$. In Fig.4 we illustrate the key steps for the enumeration of k-noncrossing RNA structures [11].
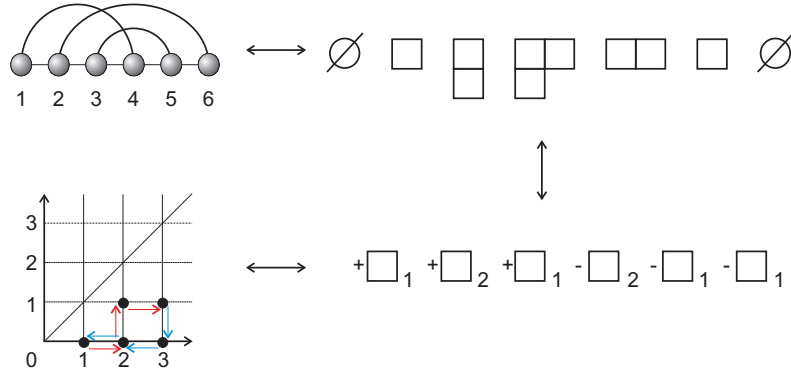

FIGURE 4. Exact enumeration of k-noncrossing RNA structures.

Once $\mathbf{T}_{k,\sigma}^{[4]}(z)$ is known we employ singularity analysis and study its dominant singularities, using Hankel contours. This Ansatz has been pioneered by P. Flajolet and A.M. Odlyzko [6]. Its basic idea is the construction of an "singular-analogue" of the Taylor-expansion. It can be shown that, under certain conditions, there exists an approximation, which is locally of the same order as the original function. The particular, local approximation allows then to derive the asymptotic form of the coefficients. In our situation all conditions for singularity analysis are met, since all our generating functions are D-finite [22, 31] and D-finite functions have an analytic continuation into any simply-connected domain containing zero.

We will compute $\mathbf{T}_{k,\sigma}^{[4]}(z)$ and show that $\mathbf{T}_{k,\sigma}^{[4]}(z)$ has an unique dominant singularity, whose type depends solely on the crossing number [12, 13]. Via singularity analysis will produce an array of exponential growth rates indexed by k and σ, summarized in Tab. 1. The ideas of this paper build on those of [11, 13]. In [13] core-structures are introduced via which σ-canonical k-noncrossing structures can be enumerated. $\langle k, 4, \sigma \rangle$-structures where $\sigma \geq 3$ can however not be enumerated via core-structures, see Fig.5. This is a result from the fact that the core-map, obtained by identifying stacks by single arcs does not preserve arc-length. Therefore we have to introduce a new set of k-noncrossing diagrams, denoted by $T_k^*(n, h)$. This class is designed for inducing a new type of cores, $C_k^*(n', h')$ (see Theorem 3). Then we proceed using ideas similar to those in [13] and prove

k	3	4	5	6	7	8	9
$\sigma = 3$	2.0348	2.2644	2.4432	2.5932	2.7243	2.8414	2.9480
$\sigma = 4$	1.7898	1.9370	2.0488	2.1407	2.2198	2.2896	2.3523
$\sigma = 5$	1.6465	1.7532	1.8330	1.8979	1.9532	2.0016	2.0449
$\sigma = 6$	1.5515	1.6345	1.6960	1.7457	1.7877	1.8243	1.8569
$\sigma = 7$	1.4834	1.5510	1.6008	1.6408	1.6745	1.7038	1.7297
$\sigma = 8$	1.4319	1.4888	1.5305	1.5639	1.5919	1.6162	1.6376
$\sigma = 9$	1.3915	1.4405	1.4763	1.5049	1.5288	1.5494	1.5677

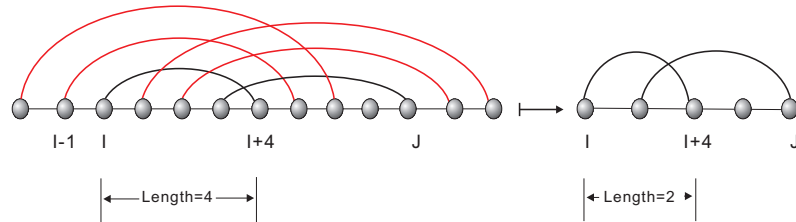TABLE 1. Exponential growth rates of $\langle k, 4, \sigma \rangle$-structures where $\sigma \geq 3$.

FIGURE 5. Core-structures will in general have 2-arcs: the structure $\delta \in T_{3,3}^{[4]}(12)$ (lhs) is mapped into its core $c(\delta)$ (rhs). Clearly δ has arc-length ≥ 4 and as a consequence of the collapse of the stack $((I+1, J+2), (I+2, J+1), (I+3, J))$ (the red arcs are being removed) into the arc $(I+3, J)$, $c(\delta)$ contains the arc $(I, I+4)$, which is, after relabeling, a 2-arc.

our exact enumeration result, Theorem 3. As for the singularity analysis the main contribution is Claim 1 of Theorem 4: a new functional equation for $\mathbf{T}_{k,\sigma}^{[4]}(z)$.

2. Preliminaries

In this Section we provide some background on the generating functions of k-noncrossing matchings [4, 15] and k-noncrossing RNA structures [11, 12]. We denote the set (number) of k-noncrossing RNA structures with arc-length $\geq \lambda$ and stack-size $\geq \sigma$ by $T_{k,\sigma}^{[\lambda]}(n)$ ($\mathsf{T}_{k,\sigma}^{[\lambda]}(n)$). By abuse of notation we omit the indices λ and σ in $T_{k,\sigma}^{[\lambda]}(n)$ ($\mathsf{T}_{k,\sigma}^{[\lambda]}(n)$) for $\lambda = 2$ and $\sigma = 1$. A k-noncrossing core-structure is a k-noncrossing RNA structures in which there exists *no* two arcs of the form $(i, j), (i+1, j-1)$. The set (number) of k-noncrossing core-structures and k-noncrossing core-structures with

exactly h arcs is denoted by $C_k(n)(\mathsf{C}_k(n))$ and $C_k(n,h)$ $(\mathsf{C}_k(n,h))$, respectively. Furthermore we denote by $f_k(n,\ell)$ the number of k-noncrossing diagrams with arbitrary arc-length and ℓ isolated vertices over n vertices and set $\mathsf{M}_k(n) = \sum_{\ell=0}^{n} f_k(n,\ell)$. That is, $\mathsf{M}_k(n)$ is the number of all k-noncrossing partial matchings. In Fig.6 we display the various types of diagrams involved.
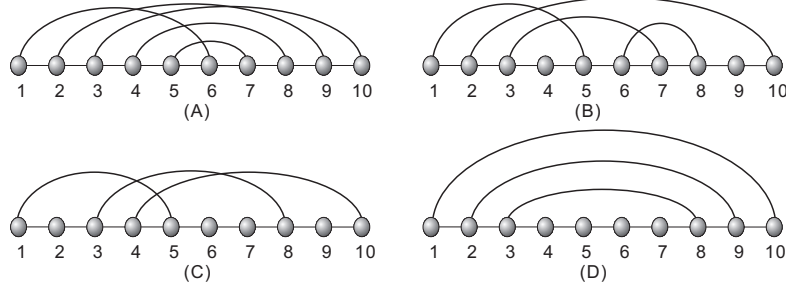

FIGURE 6. Basic diagram types: (A) 4-noncrossing matching (no isolated points), (B) 3-noncrossing partial matching (isolated points 4 and 9), (C) 4-noncrossing RNA structure with arc-length ≥ 4 and stack length ≥ 1, (D) RNA structure with arc-length ≥ 5 and stack-length ≥ 3.

2.1. k-noncrossing partial matchings and RNA structures.

The following identities are due to Grabiner and Magyar [8]

$$(2.1) \qquad \sum_{n\geq 0} f_k(n,0) \cdot \frac{x^n}{n!} \;=\; \det[I_{i-j}(2x) - I_{i+j}(2x)]\big|_{i,j=1}^{k-1}$$

$$(2.2) \qquad \sum_{n\geq 0} \left\{ \sum_{\ell=0}^{n} f_k(n,\ell) \right\} \cdot \frac{x^n}{n!} \;=\; e^x \det[I_{i-j}(2x) - I_{i+j}(2x)]\big|_{i,j=1}^{k-1} \;,$$

where $I_r(2x) = \sum_{j\geq 0} \frac{x^{2j+r}}{j!(r+j)!}$ denotes the hyperbolic Bessel function of the first kind of order r. Eq. (2.1) and (2.2) allow only "in principle" for explicit computation of the numbers $f_k(n,\ell)$ and in view of $f_k(n,\ell) = \binom{n}{\ell} f_k(n-\ell,0)$ everything can be reduced to (perfect) matchings, where we have the following situation: there exists an asymptotic approximation of the determinant of the hyperbolic Bessel function for general k due to [15] and employing the subtraction of singularities-principle [19] one can prove [15]

$$(2.3) \qquad \forall\, k \in \mathbb{N}; \qquad f_k(2n,0) \;\sim\; c_k\, n^{-((k-1)^2+(k-1)/2)}\, (2(k-1))^{2n}, \qquad c_k > 0 \;,$$

k	2	3	4	5	6	7	8	9	10
γ_k^{-1}	2.6180	4.7913	6.8541	8.8875	10.9083	12.9226	14.9330	16.9410	18.9472

TABLE 2. The exponential growth rates of $\langle k, 2, 1\rangle$-structures.

k	4	5	6	7	8	9
$\gamma_k^{[4]^{-1}}$	6.5290	8.6483	10.7176	12.7635	14.7963	16.8210

TABLE 3. The exponential growth rates of $\langle k, 4, 1\rangle$-structures.

where $\rho_k = \frac{1}{2(k-1)}$ is the dominant real singularity of $\sum_{n\geq 0} f_k(2n, 0)z^{2n}$. For $\langle k, 2, 1\rangle$-structures we have [11, 12]

$$(2.4) \qquad \mathsf{T}_k(n) \ = \ \sum_{b=0}^{\lfloor n/2\rfloor} (-1)^b \binom{n-b}{b} \mathsf{M}_k(n-2b)$$

$$(2.5) \qquad \mathsf{T}_k(n) \ \sim \ c_k\, n^{-((k-1)^2+(k-1)/2)}\,(\gamma_k)^{-n}, \qquad c_k > 0\,,$$

where γ_k is the unique, minimal solution of $\frac{z}{z^2-z+1} = \rho_k$, see Tab. 2. For $\langle k, 4, 1\rangle$-structures we have according to [9] the following exact enumeration result

$$(2.6) \qquad \mathsf{T}_k^{[4]}(n) \ = \ \sum_{b\leq\lfloor\frac{n}{2}\rfloor} (-1)^b\, \lambda(n, b)\, \mathsf{M}_k(n-2b), \qquad 4 \leq k \leq 9\,,$$

where $\lambda(n, b)$ denotes the number of way of selecting b arcs of length ≤ 3 over n vertices and

$$(2.7) \qquad \mathsf{T}_k^{[4]}(n) \sim c_k\, n^{-((k-1)^2+(k-1)/2)}\,\left(\gamma_k^{[4]}\right)^{-n}$$

where $\gamma_k^{[4]}$ is the unique positive, real solution of $\frac{z\, r_1(-z^2)}{1-zr_1(-z^2)} = \rho_k$ where $r_1(z)$ satisfies

$$u(z) \ = \ \sqrt{1 + 4z - 4z^2 - 6z^3 + 4z^4 + z^6}$$

$$r_1(z) \ = \ -\frac{-2z^2 + z^3 - 1 + u(z)}{2(1 - 2z - z^2 + z^4)}\,.$$

In Tab. 3 we present the exponential growth rates for $\mathsf{T}_k^{[4]}(n)$ for $k = 4, \ldots, 9$. For $\langle k, 2, \sigma\rangle$-structures we have according to [13]

$$(2.8) \qquad \mathbf{T}_{k,\sigma}(x) = \frac{1}{u_0 x^2 - x + 1} \sum_{n\geq 0} f_k(2n, 0) \left(\frac{\sqrt{u_0}\, x}{u_0 x^2 - x + 1}\right)^{2n}$$

where $u_0 = \frac{(x^2)^{\sigma-1}}{(x^2)^\sigma - x^2 + 1}$ and

$$(2.9) \qquad \mathsf{T}_{k,\sigma}(n) \sim c_k\, n^{-((k-1)^2+(k-1)/2)}\,\left(\gamma_k^{-1}\right)^n$$

k	2	3	4	5	6	7	8	9	10
$\sigma = 2$	1.9680	2.5881	3.0382	3.4138	3.7438	4.0420	4.3162	4.5715	4.8115
$\sigma = 3$	1.7160	2.0477	2.2704	2.4466	2.5955	2.7259	2.8427	2.9490	3.0469
$\sigma = 4$	1.5782	1.7984	1.9410	2.0511	2.1423	2.2209	2.2904	2.3529	2.4100
$\sigma = 5$	1.4899	1.6528	1.7561	1.8347	1.8991	1.9540	2.0022	2.0454	2.0845

TABLE 4. The exponential growth rates of $\langle k, 2, \sigma \rangle$-structures [13].

where $\gamma_{k,\sigma}$ is a positive real dominant singularity of $\sum_{n \geq 0} \mathsf{T}_{k,\sigma}(n) x^n$ and the minimal positive real solution of the equation

$$(2.10) \qquad \frac{\sqrt{\frac{(x^2)^{\sigma-1}}{(x^2)^\sigma - x^2 + 1}}\, x}{\left(\frac{(x^2)^{\sigma-1}}{(x^2)^\sigma - x^2 + 1} \right) x^2 - x + 1} = \rho_k.$$

In Table 4 we present the exponential growth rates of $\langle k, 2, \sigma \rangle$-structures.

2.2. **Singularity analysis.** Let us next recall some basic fact about analytic functions. Pfringsheim's Theorem [23] guarantees that each power series with positive coefficients has a positive real dominant singularity. This singularity plays a key role for the asymptotics of the coefficients. In the proof of Theorem 4 it will be important to deduce relations between the coefficients from functional equations of generating functions. The class of theorems that deal with such deductions are called transfer-theorems [6]. We consider a specific domain in which the functions in question are analytic and which is "slightly" bigger than their respective radius of convergence. It is tailored for extracting the coefficients via Cauchy's integral formula. Details on the method can be found in [6]. In case of D-finite functions we have analytic continuation in any simply connected domain containing zero [26] and all prerequisites of singularity analysis are met. To be precise, given two numbers ϕ, R, where $R > 1$ and $0 < \phi < \frac{\pi}{2}$ and $\rho \in \mathbb{R}$, the open domain $\Delta_\rho(\phi, R)$ is defined as

$$(2.11) \qquad \Delta_\rho(\phi, R) = \{ z \mid |z| < R, z \neq \rho, |\mathrm{Arg}(z - \rho)| > \phi \}$$

A domain is a Δ_ρ-domain if it is of the form $\Delta_\rho(\phi, R)$ for some R and ϕ. A function is Δ_ρ-analytic if it is analytic in some Δ_ρ-domain. We use the notation

$$(2.12) \qquad (f(z) = O\left(g(z) \right) \text{ as } z \to \rho) \quad \Longleftrightarrow \quad (f(z)/g(z) \text{ is bounded as } z \to \rho)$$

and if we write $f(z) = O(g(z))$ it is implicitly assumed that z tends to a (unique) singularity. $[z^n]\, f(z)$ denotes the coefficient of z^n in the power series expansion of $f(z)$ around 0.

k	$q_{0,k}(z)$	M_k
3	$(1/4 - 4z^2)\,z^2$	$\{\pm 1/4\}$
4	$(144\,z^4 - 40\,z^2 + 1)\,z^6$	$\{\pm 1/2, \pm 1/6\}$
5	$(-80\,z^2 + 1024\,z^4 + 1)\,z^8$	$\{\pm 1/4, \pm 1/8\}$
6	$(-4144\,z^4 + 140\,z^2 + 14400\,z^6 + 1)\,z^{10}$	$\{\pm 1/2, \pm 1/6, \pm 1/10, \}$
7	$(-1 - 12544\,z^4 + 224\,z^2 + 147456\,z^6)\,z^{12}$	$\{\pm 1/4, \pm 1/8, \pm 1/12\}$
8	$(1 - 336z^2 + 31584z^4 + 2822400z^8 - 826624z^6)z^{14}$	$\{\pm 1/2, \pm 1/6, \pm 1/10, \pm 1/14\}$
9	$-(-480z^2 + 1 + 69888z^4 + 37748736z^8 - 3358720z^6)z^{16}$	$\{\pm 1/4, \pm 1/8, \pm 1/12, \pm 1/16\}$

TABLE 5. The polynomials $q_{0,k}(z)$ and their nonzero roots.

Theorem 1. [6] *Let $f(z), g(z)$ be D-finite, Δ_ρ-analytic functions with unique dominant singularity ρ and suppose*

$$(2.13) \qquad\qquad f(z) = O(g(z)) \quad for \quad z \to \rho\ .$$

Then we have

$$(2.14) \qquad\qquad [z^n]f(z) = K\left(1 - O(\frac{1}{n})\right)[z^n]g(z)\ ,$$

where K is some constant.

Let $\mathbf{F}_k(z) = \sum_n f_k(2n,0)z^{2n}$, the ordinary generating function of k-noncrossing matchings. It follows from eq. (2.1) that the power series $\mathbf{F}_k(z)$ is D-finite, i.e. there exists some $e \in \mathbb{N}$ such that

$$(2.15) \qquad q_{0,k}(z)\frac{d^e}{dz^e}\mathbf{F}_k(z) + q_{1,k}(z)\frac{d^{e-1}}{dz^{e-1}}\mathbf{F}_k(z) + \cdots + q_{e,k}(z)\mathbf{F}_k(z) = 0\ ,$$

where $q_{j,k}(z)$ are polynomials. The key point is that any dominant singularity of $\mathbf{F}_k(z)$ is contained in the set of roots of $q_{0,k}(z)$ [22], which we denote by M_k. The polynomials $q_{0,k}(z)$ and their sets of roots for $k = 3, \ldots, 9$ are given in Table 5. Accordingly, $\mathbf{F}_k(z)$ has singularities $\pm\rho_k$, where $\rho_k = (2(k-1))^{-1}$.

As a consequence of Theorem 1, eq. (2.3) and the so called supercritical case of singularity analysis [6], VI.9., p. 400, we give the following result[14] tailored for our functional equations.

Theorem 2. *Suppose $\vartheta_\sigma(z)$ is algebraic over $K(z)$, regular for $|z| < \delta$ and satisfies $\vartheta_\sigma(0) = 0$. Suppose further $\gamma_{k,\sigma}$ is the unique solution with minimal modulus $< \delta$ of the two equations $\vartheta_\sigma(x) = \rho_k$ and $\vartheta_\sigma(x) = -\rho_k$. Then $\gamma_{k,\sigma}$ is the unique dominant singularity of $\mathbf{F}_k(\vartheta_\sigma(z))$ and*

$$(2.16) \qquad [z^n]\, \mathbf{F}_k(\vartheta_\sigma(z)) \sim c_k\, n^{-((k-1)^2 + (k-1)/2)} \left(\gamma_{k,\sigma}^{-1} \right)^n .$$

3. Exact Enumeration

In this section we present the exact enumeration of $\langle k, 4, \sigma \rangle$-structures, where $\sigma \geq 3$. The structure of our formula is analogous to the Möbius inversion formula proved in [13]: $\mathsf{T}_{k,\sigma}(n, h) = \sum_{b=\sigma-1}^{h-1} \binom{b+(2-\sigma)(h-b)-1}{h-b-1} \mathsf{C}_k(n-2b, h-b)$, which relates the number of all structures and the number of core-structures. As we pointed out in the introduction the latter cannot be used in order to enumerate k-noncrossing structures with arc-length ≥ 4, see Fig.5. We consider the arc-sets

$$\beta_2 = \{(i, i+2) \mid i+1 \text{ isolated}\} \quad \text{and} \quad \beta_3 = \{(i, i+3) \mid i+1, i+2 \text{ isolated}\}$$

and set $\beta = \beta_2 \cup \beta_3$. Furthermore

$$(3.1) \qquad C_k^*(n, h) \quad = \quad \{\delta \mid \delta \in C_k(n, h); \delta \text{ contains no 1-arc and no } \beta\text{-arc}\}$$

$$(3.2) \qquad T_k^*(n, h) \quad = \quad \{\delta \mid \delta \in T_k(n, h); \delta \text{ contains no 1-arc and no } \beta\text{-arc}\} .$$

Theorem 3. *Suppose we have $k, h, \sigma \in \mathbb{N}$, $k \geq 2$, $h \leq n/2$ and $\sigma \geq 3$. Then the number of $\langle k, 4, \sigma \rangle$-structures having exactly h arcs is given by*

$$(3.3) \qquad \mathsf{T}_{k,\sigma}^{[4]}(n, h) = \sum_{b=\sigma-1}^{h-1} \binom{b + (2-\sigma)(h-b) - 1}{h - b - 1} \mathsf{C}_k^*(n - 2b, h - b)$$

where $\mathsf{C}_k^*(n, h)$ satisfies $\mathsf{C}_k^*(n, 0) = 1$ and

$$(3.4) \qquad \mathsf{C}_k^*(n, h) = \sum_{b=0}^{h-1} (-1)^{h-b-1} \binom{h-1}{b} \mathsf{T}_k^*(n - 2h + 2b + 2, b + 1) \quad \text{for } h \geq 1 .$$

Furthermore, $\mathsf{T}_k^*(n, h)$ satisfies

$$(3.5) \; \mathsf{T}_k^*(n, h) = \sum_{0 \leq j_1 + j_2 + j_3 \leq h} (-1)^{j_1 + j_2 + j_3} \lambda(n, j_1, j_2, j_3)\, f_k(n - 2j_1 - 3j_2 - 4j_3, n - 2h - j_2 - 2j_3)$$

where

$$\lambda(n, j_1, j_2, j_3) = \binom{n - j_1 - 2j_2 - 3j_3}{j_1, j_2, j_3, n - 2j_1 - 3j_2 - 4j_3} .$$

n	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
$\mathsf{T}_{3,3}^{[4]}(n)$	1	2	4	8	15	28	52	96	176	316	557	965	1660	2860	4974	8754	15562
$\mathsf{T}_{3,4}^{[4]}(n)$	1	1	1	2	4	8	14	23	36	56	88	141	231	382	633	1038	1679

TABLE 6. Exact enumeration: $\mathsf{T}_{3,3}^{[4]}(n)$ and $\mathsf{T}_{3,4}^{[4]}(n)$ for $n \leq 24$, respectively.

In Tab.6 we display the first numbers of $\langle k, 4, 3 \rangle$-structures and $\langle k, 4, 4 \rangle$-structures, respectively.

Proof. We first show that there exists a mapping from $\langle k, 4, \sigma \rangle$-structures with h arcs over $[n]$ into $\dot{\bigcup}_{\sigma-1 \leq b \leq h-1} C_k^*(n - 2b, h - b)$:

$$(3.6) \qquad c \colon T_{k,\sigma}^{[4]}(n, h) \to \dot{\bigcup}_{\sigma-1 \leq b \leq h-1} C_k^*(n - 2b, h - b), \quad \delta \mapsto c(\delta)$$

which is obtained in two steps: first induce $c(\delta)$ by mapping arcs and isolated vertices as follows:

$$(3.7) \quad \forall \ell \geq \sigma - 1; \quad ((i - \ell, j + \ell), \ldots, (i, j)) \mapsto (i, j) \quad \text{and} \quad j \mapsto j \quad \text{if } j \text{ is an isolated vertex}$$

and second relabel the resulting diagram from left to right in increasing order, see Fig.7.

Claim 1. $c \colon T_{k,\sigma}^{[4]}(n, h) \longrightarrow \dot{\bigcup}_{\sigma-1 \leq b \leq h-1} C_k^*(n - 2b, h - b)$ is well-defined and surjective.
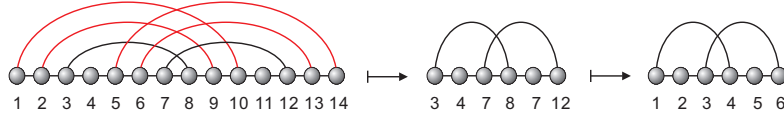

FIGURE 7. The mapping $c \colon T_{k,\sigma}^{[4]}(n, h) \longrightarrow \dot{\bigcup}_{\sigma-1 \leq b \leq h-1} C_k^*(n - 2b, h - b)$ is obtained in two steps: first contraction of the stacks while keeping isolated points and secondly relabeling of the resulting diagram.

By construction, c does not change the crossing number. Since $T_{k,\sigma}^{[4]}(n)$ contains only arcs of length ≥ 4 we derive $c(T_{k,\sigma}^{[4]}(n)) \subset C_k^*(n - 2b, h - b)$. Therefore c is well-defined. It remains to show that c is surjective. For this purpose let $\delta \in C_k^*(n - 2b, h - b)$ and set $a = b - (\sigma - 1)(h - b)$. We proceed constructing a k-noncrossing structure $\tilde{\delta}$ in three steps:

Step 1. replace each label i by r_i, where $r_i \leq r_s$ if and only if $i \leq s$.

Step 2. replace the leftmost arc (r_p, r_q) by the sequence of arcs

$$(3.8) \qquad ((\tau_p - ([\sigma - 1] + a), \tau_q + ([\sigma - 1] + a)), \ldots, (\tau_p, \tau_q))$$

replace any other arc (r_p, r_q) by the sequence

$$(3.9) \qquad ((\tau_p - [\sigma - 1], \tau_q + [\sigma - 1]), \ldots, (\tau_p, \tau_q))$$

and each isolated vertex r_s by τ_s.

Step 3. Set for $x, y \in \mathbb{Z}$, $\tau_b + y \le \tau_c + x$ if and only if $(b < c)$ or $(b = c$ and $y \le x)$. By construction, $\le$ is a linear order over

$$n - 2b + 2(h - b)(\sigma - 1) + 2a = n - 2b + 2(h - b)(\sigma - 1) + 2(b - (\sigma - 1)(h - b)) = n$$

elements, which we then label from 1 to n (left to right) in increasing order. It is straightforward to verify that $c(\tilde{\delta}) = \delta$ holds. It remains to show that $\tilde{\delta} \in T_{k,\sigma}^{[4]}(n)$. Suppose *a contrario* $\tilde{\delta}$ contains an arc $(i, i + 2)$. Since $\sigma \ge 3$ we can then conclude that $i + 1$ is necessarily isolated. The arc $(i, i + 2)$ is mapped by c into $(j, j + 2)$ with isolated point $j + 1$, which is impossible by definition of $C_k^*(n', h')$. It follows similarly that an arc of the form $(i, i + 3)$ cannot be contained in $\tilde{\delta}$ and Claim 1 follows.

Labeling the h arcs of $\delta \in T_{k,\sigma}^{[4]}(n, h)$ from left to right and keeping track of multiplicities gives rise to the map

(3.10)

$$f_{k,\sigma} : T_{k,\sigma}^{[4]}(n, h) \to \dot{\bigcup}_{\sigma - 1 \le b \le h - 1} \left[C_k^*(n - 2b, h - b) \times \left\{ (a_j)_{1 \le j \le h - b} \,\Big|\, \sum_{j=1}^{h-b} a_j = b, \ a_j \ge \sigma - 1 \right\} \right],$$

given by $f_{k,\sigma}(\delta) = (c(\delta), (a_j)_{1 \le j \le h - b})$. We can conclude that $f_{k,\sigma}$ is well-defined and a bijection. We proceed computing the multiplicities of the resulting core-structures [13]:

$$(3.11) \qquad \left| \left\{ (a_j)_{1 \le j \le b} \,\Big|\, \sum_{j=1}^{h-b} a_j = b; \ a_j \ge \sigma - 1 \right\} \right| = \binom{b + (2 - \sigma)(h - b) - 1}{h - b - 1}.$$

Eq. (3.11) and eq. (3.10) imply

$$(3.12) \qquad \mathsf{T}_{k,\sigma}^{[4]}(n, h) = \sum_{b=\sigma-1}^{h-1} \binom{b + (2 - \sigma)(h - b) - 1}{h - b - 1} \mathsf{C}_k^*(n - 2b, h - b),$$

whence eq. (3.3). Next we consider the map

$$(3.13) \qquad c^* : T_k^*(n, h) \to \dot{\bigcup}_{0 \le b \le h - 1} C_k^*(n - 2b, h - b), \quad \delta \mapsto c^*(\delta)$$

Indeed, c^* is well defined, since any diagram in $T_k^*(n, h)$ can be mapped into a core structure without 1- and β-arcs, i.e. into an element of $C_k^*(n', h')$. That gives rise to

$$(3.14) \qquad \mathsf{T}_k^*(n, h) = \sum_{b=0}^{h-1} \binom{h - 1}{b} \mathsf{C}_k^*(n - 2b, h - b)$$

and via Möbius-inversion formula we obtain eq. (3.4). It is straightforward to show there are $\lambda(n, j_1, j_2, j_3) = \binom{n-j_1-2j_2-3j_3}{j_1,j_2,j_3,n-2j_1-3j_2-4j_3}$ ways to select j_1 1-arcs ,j_2 β_2-arcs and j_3 β_3-arcs over $[n]$. Since removing j_1 1-arcs, j_2 β_2-arcs and j_3 β_3-arcs removes $2j_1 + 3j_2 + 4j_3$ vertices, the number of configurations of at least j_1 1-arcs, j_2 β_2-arcs and j_3 β_3-arcs is given by $\lambda(n, j_1, j_2, j_3) f_k(n - 2j_1 - 3j_2 - 4j_3, n - 2h - j_2 - 2j_3)$. Via inclusion-exclusion principle, we arrive at

$$\mathsf{T}_k^*(n, h) = \sum_{0 \leq j_1 + j_2 + j_3 \leq h} (-1)^{j_1 + j_2 + j_3} \lambda(n, j_1, j_2, j_3) f_k(n - 2j_1 - 3j_2 - 4j_3, n - 2h - j_2 - 2j_3) \, ,$$

whence Theorem 3. $\qquad\square$

The following functional identity, relating the bivariate generating functions of $\mathsf{T}_{k,\sigma}^{[4]}(n, h)$ and $\mathsf{C}_k^*(n, h)$, is instrumental for proving our main result in the next section, Theorem 4.

Lemma 1. [13] *Let $k, \sigma \in \mathbb{N}$, $k \geq 2$ and let u, x be indeterminants. Suppose we have*

$$(3.15) \quad \forall\, h \geq 1, \quad \mathsf{A}_{k,\sigma}(n, h) = \sum_{b=\sigma-1}^{h-1} \binom{b + (2 - \sigma)(h - b) - 1}{h - b - 1} \mathsf{B}_k(n - 2b, h - b) \text{ and } \mathsf{A}_{k,\sigma}(n, 0) = 1 \, .$$

Then we have the functional relation

$$(3.16) \quad \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathsf{A}_{k,\sigma}(n, h) u^h x^n \;=\; \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathsf{B}_k(n, h) \left(\frac{u \cdot (ux^2)^{\sigma - 1}}{1 - ux^2} \right)^h x^n \, .$$

According to Lemma 1 eq. (3.14) and eq. (3.3) we obtain the two functional identities

$$(3.17) \quad \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathsf{T}_k^*(n, h) u^h x^n = \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathsf{C}_k^*(n, h) \left(\frac{u}{1 - ux^2} \right)^h x^n$$

$$(3.18) \quad \sum_{n \geq 0} \mathsf{T}_{k,\sigma}^{[4]}(n) x^n = \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathsf{C}_k^*(n, h) \left(\frac{(x^2)^{\sigma - 1}}{1 - x^2} \right)^h x^n \quad \text{for } \sigma \geq 3 \, .$$

4. Asymptotic Enumeration

In this section we study the asymptotics of $\langle k, 4, \sigma \rangle$-structures, where $\sigma \geq 3$. We are particularly interested in deriving simple formulas that can be used for assessing the complexity of prediction

algorithms for k-noncrossing RNA structures. In order to state Theorem 4 below we introduce

$$(4.1) \qquad w_0(x) \;=\; \frac{x^{2\sigma-2}}{1 - x^2 + x^{2\sigma}}$$

$$(4.2) \qquad v(x) \;=\; 1 - x + w(x)x^2 + w(x)x^3 + w(x)x^4$$

$$(4.3) \qquad v_0(x) \;=\; 1 - x + w_0(x)x^2 + w_0(x)x^3 + w_0(x)x^4.$$

Theorem 4. *Let $k, \sigma \in \mathbb{N}$, $k, \sigma \geq 3$, x be an indeterminate and ρ_k the dominant, positive real singularity of $\sum_{n\geq 0} f_k(2n,0)z^{2n}$. Then $\mathbf{T}^{[4]}_{k,\sigma}(x)$, the generating function of $\langle k, 4, \sigma \rangle$-structures is given by*

$$(4.4) \qquad \mathbf{T}^{[4]}_{k,\sigma}(x) = \frac{1}{v_0(x)} \sum_{n\geq 0} f_k(2n,0) \left(\frac{\sqrt{w_0(x)}\,x}{v_0(x)} \right)^{2n}.$$

Furthermore

$$(4.5) \qquad \mathsf{T}^{[4]}_{k,\sigma}(n) \sim c_k\, n^{-(k-1)^2 - \frac{k-1}{2}} \left(\frac{1}{\gamma^{[4]}_{k,\sigma}} \right)^n, \quad for \quad k = 3, 4, \ldots, 9$$

holds, where $\gamma^{[4]}_{k,\sigma}$ is the positive real dominant singularity of $\mathbf{T}^{[4]}_{k,\sigma}(x)$ and the minimal positive real solution of the equation $\frac{\sqrt{w_0(x)}\,x}{v_0(x)} = \rho_k$ and $f_k(2n,0) \sim n^{-(k-1)^2 - \frac{k-1}{2}} \left(\frac{1}{\rho_k} \right)^{2n}$ (eq. (2.3)).

Proof. In the following we will use the notation w_0 instead of $w_0(x)$, eq. (4.1). The first step derives a functional equation relating the bivariate generating functions of $T^*_k(n,h)$ and $f_k(2h',0)$. For this purpose we use eq. (3.5).

Claim 1.

$$(4.6) \qquad \sum_{n\geq 0} \sum_{h\leq \frac{n}{2}} \mathsf{T}^*_k(n,h) w^h x^n = \frac{1}{v(x)} \sum_{n\geq 0} f_k(2n,0) \left(\frac{\sqrt{w}\,x}{v(x)} \right)^{2n}.$$

Set $\varphi_m(w) = \sum_{h \leq \frac{m}{2}} \binom{m}{2h} f_k(2h, 0) w^h$. In order to prove Claim 1 we compute

$$\sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \mathsf{T}_k^*(n, h) w^h x^n$$

$$= \sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \sum_{0 \leq j_1 + j_2 + j_3 \leq h} (-1)^{j_1 + j_2 + j_3} \lambda(n, j_1, j_2, j_3) f_k(n - 2j_1 - 3j_2 - 4j_3, n - 2h - j_2 - 2j_3) w^h x^n$$

$$= \sum_{n \geq 0} \sum_{j_1 + j_2 + j_3 \leq \frac{n}{2}} (-1)^{j_1 + j_2 + j_3} \lambda(n, j_1, j_2, j_3) x^n$$

$$\times \sum_{h \geq j_1 + j_2 + j_3} f_k(n - 2j_1 - 3j_2 - 4j_3, n - 2h - j_2 - 2j_3) w^h$$

$$= \sum_{n \geq 0} \sum_{j_1 + j_2 + j_3 \leq \frac{n}{2}} (-1)^{j_1 + j_2 + j_3} \lambda(n, j_1, j_2, j_3) x^n$$

$$\times \sum_{h \geq j_1 + j_2 + j_3} \binom{n - 2j_1 - 3j_2 - 4j_3}{n - 2h - j_2 - 2j_3} f_k(2(h - j_1 - j_2 - j_3), 0) w^h$$

$$= \sum_{n \geq 0} \sum_{j_1 + j_2 + j_3 \leq \frac{n}{2}} (-1)^{j_1 + j_2 + j_3} \lambda(n, j_1, j_2, j_3) w^{j_1 + j_2 + j_3} \varphi_{n - 2j_1 - 3j_2 - 4j_3}(w) x^n .$$

We interchange the summation over $j_1 + j_2 + j_3$ and n and arrive at

$$\sum_{j_1 + j_2 + j_3 \geq 0} \sum_{n \geq 2j_1 + 3j_2 + 4j_3} (-1)^{j_1 + j_2 + j_3} \binom{n - j_1 - 2j_2 - 3j_3}{j_1, j_2, j_3, n - 2j_1 - 3j_2 - 4j_3} w^{j_1 + j_2 + j_3} \varphi_{n - 2j_1 - 3j_2 - 4j_3}(w) x^n$$

$$= \sum_{j_1 + j_2 + j_3 \geq 0} \frac{(-w)^{j_1 + j_2 + j_3}}{j_1! j_2! j_3!} \sum_{n \geq 2j_1 + 3j_2 + 4j_3} \frac{(n - j_1 - 2j_2 - 3j_3)!}{(n - 2j_1 - 3j_2 - 4j_3)!} \varphi_{n - 2j_1 - 3j_2 - 4j_3}(w) x^n .$$

Setting $m = n - 2j_1 - 3j_2 - 4j_3$ this becomes

$$= \sum_{j_1 + j_2 + j_3 \geq 0} \frac{(-w)^{j_1 + j_2 + j_3}}{j_1! j_2! j_3!} x^{2j_1 + 3j_2 + 4j_3} \sum_{m \geq 0} \frac{(m + j_1 + j_2 + j_3)!}{m!} \varphi_m(w) x^m$$

$$= \sum_{m \geq 0} \left[\sum_{j_1 + j_2 + j_3 \geq 0} \binom{m + j_1 + j_2 + j_3}{m, j_1, j_2, j_3} (-wx^2)^{j_1} (-wx^3)^{j_2} (-wx^4)^{j_3} \right] \varphi_m(w) x^m$$

$$= \sum_{m \geq 0} \varphi_m(w) x^m \left(\frac{1}{1 + wx^2 + wx^3 + wx^4} \right)^{m+1}$$

$$= \frac{1}{1 + wx^2 + wx^3 + wx^4} \sum_{m \geq 0} \varphi_m(w) \left(\frac{x}{1 + wx^2 + wx^3 + wx^4} \right)^m .$$

Next we compute

$$
\begin{aligned}
\sum_{m \geq 0} \varphi_m(w) y^m &= \int_0^\infty \sum_{m \geq 0} \varphi_m(w) \frac{(xy)^m}{m!} e^{-x} dx \\
&= \int_0^\infty \sum_{m \geq 0} \sum_{h \leq \frac{m}{2}} \binom{m}{2h} f_k(2h, 0) w^h \frac{(xy)^m}{m!} e^{-x} dx \\
&= \int_0^\infty \sum_{m \geq 0} \sum_{h \leq \frac{m}{2}} f_k(2h, 0) w^h \frac{(xy)^{2h}}{(2h)!} \frac{(xy)^{m-2h}}{(m-2h)!} e^{-x} dx \\
&= \int_0^\infty \sum_{h \geq 0} f_k(2h, 0) \frac{(\sqrt{w}\, xy)^{2h}}{(2h)!} \sum_{m \geq 2h} \frac{(xy)^{m-2h}}{(m-2h)!} e^{-x} dx \\
&= \sum_{n \geq 0} f_k(2n, 0) \frac{(\sqrt{w} y)^{2n}}{(2n)!} \int_0^\infty e^{-(1-y)x} x^{2n} dx \\
&= \sum_{n \geq 0} f_k(2n, 0) \frac{(\sqrt{w} y)^{2n}}{(2n)!} \frac{\int_0^\infty e^{-(1-y)x}((1-y)x)^{2n} d((1-y)x)}{(1-y)^{2n+1}} \\
&= \frac{1}{1-y} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{\sqrt{w} y}{1-y} \right)^{2n}.
\end{aligned}
$$

Therefore the bivariate generating function can be written as

$$
\sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} \mathsf{T}_k^*(n, h) w^h x^n = \frac{1}{v(x)} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{\sqrt{w}\, x}{v(x)} \right)^{2n},
$$

whence Claim 1. In view of eq. (3.17) and Claim1 we arrive at

$$
(4.7) \qquad \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} \mathsf{C}_k^*(n, h) \left(\frac{w}{1 - wx^2} \right)^h x^n = \frac{1}{v(x)} \sum_{n \geq 0} f_k(2n, 0) \left(\frac{\sqrt{w}\, x}{v(x)} \right)^{2n}.
$$

By definition of $w_0 = w_0(x)$ have

$$
(4.8) \qquad \frac{(x^2)^{\sigma-1}}{1 - x^2} = \frac{w_0}{1 - w_0 x^2}.
$$

According to eq.(3.18), eq.(4.8) and eq.(4.7) this allows us to derive

$$
\begin{aligned}
\mathbf{T}_{k,\sigma}^{[4]}(x) &= \sum_{n\geq 0}\sum_{0\leq h\leq \frac{n}{2}} \mathsf{C}_k^*(n,h)\left(\frac{(x^2)^{\sigma-1}}{1-x^2}\right)^h x^n \\
&= \sum_{n\geq 0}\sum_{0\leq h\leq \frac{n}{2}} \mathsf{C}_k^*(n,h)\left(\frac{w_0}{1-w_0 x^2}\right)^h x^n \\
&= \frac{1}{v_0(x)}\sum_{n\geq 0} f_k(2n,0)\left(\frac{\sqrt{w_0}x}{v_0(x)}\right)^{2n} ,
\end{aligned}
$$

whence (4.4). Let $\mathbf{V}_k(x) = \sum_{n\geq 0} f_k(2n,0)\left(\frac{\sqrt{w_0}\,x}{v_0(x)}\right)^{2n}$.

Claim 2. The unique, minimal, positive, real solution of

$$
(4.9) \qquad\qquad \vartheta_\sigma(x) = \frac{\sqrt{w_0}x}{v_0(x)} = \rho_k , \quad for \quad k = 3,4,\ldots,9
$$

denoted by $\gamma_{k,\sigma}^{[4]}$ is the unique dominant singularity of $\mathbf{T}_{k,\sigma}^{[4]}(x)$.

Clearly, a dominant singularity of $\frac{1}{v_0(x)}\mathbf{V}_k(x)$ is either a singularity of $\mathbf{V}_k(x)$ or $\frac{1}{v_0(x)}$. Suppose there exists some singularity $\zeta \in \mathbb{C}$ which is a pole of $\frac{1}{v_0(x)}$. By construction $\zeta \neq 0$ and ζ is necessarily a non-finite singularity of $\mathbf{V}_k(x)$. If $|\zeta| \leq \gamma_{k,\sigma}^{[4]}$, then we arrive at the contradiction

$$
|\mathbf{V}_k(\zeta)| > |\mathbf{V}_k(\gamma_{k,\sigma}^{[4]})| \geq \mathbf{V}_k(|\zeta|)
$$

since $\mathbf{V}_k(\zeta)$ is not finite and $\mathbf{V}_k(\gamma_{k,\sigma}^{[4]}) = \sum_{n\geq 0} f_k(2n,0)\rho_k^{2n} < \infty$. Therefore all dominant singularities of $\mathbf{T}_{k,\sigma}^{[4]}(x)$ are singularities of $\mathbf{V}_k(x)$. According to Pringsheim's Theorem [23], $\mathbf{T}_{k,\sigma}^{[4]}(x)$ has a dominant positive real singularity which by construction equals $\gamma_{k,\sigma}^{[4]}$ being the minimal positive real solution of eq. (4.9). To prove this, we use that for $3 \leq k \leq 9$, the generating function $\mathbf{F}_k(x)$ has only the two dominant singularities $\pm\rho_k$, see Section 2, Tab. 5. Furthermore we verify that for $3 \leq k \leq 9$, $\gamma_{k,\sigma}^{[4]}$, has strictly smaller modulus than all solutions of $\vartheta_\sigma(z) = -\rho_k$, whence Claim 2. Accordingly, Theorem 2 applies and we have

$$
(4.10) \qquad\qquad \mathsf{T}_{k,\sigma}^{[4]}(n) \sim c_k n^{-(k-1)^2 - \frac{k-1}{2}}\left(\frac{1}{\gamma_{k,\sigma}^{[4]}}\right)^n \quad \text{for some constant } c_k
$$

completing the proof of Theorem 4. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\Box$

References

[1] *Mapping RNA form and function* Science, **2** (2005).

[2] André. D., 1887. Solution directed du probleème, résolu par M. Bertrand. C. R. Acad. Sci. Paris, 105, 436–437.

[3] T. Akutsu, *Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots* Discr. Appl. Math. **104**(2000), 45-62.

[4] W.Y.C. Chen, E.Y.P. Deng, R.R.X. Du, R.P. Stanley and C.H. Yan, *Crossings and Nestings of Matchings and Partitions* Trans. Amer. Math. Soc. **359** (2007), No. 4, 1555–1575.

[5] W.Y.C. Chen, J. Qin and C.M. Reidys, *Crossings and Nestings of tangled-diagrams*, arXiv:0710.4053v2

[6] P. Flajolet and R. Sedgewick, *Analytic combinatorics*, (2007).

[7] Gessel. I. M., Zeilberger. D., 1992. Random walk in a Weyl chamber, Proc. Amer. Math. Soc. 115 27–31.

[8] D.J. Grabiner and P. Magyar, *Random walks in Weyl chambers and the decomposition of tensor powers* Discr. Appl. Math. **2**(1993),239-260.

[9] H.S.W. Han and C.M.Reidys, *Pseudoknot RNA structure with arc-length* ≥ 4. J. Comp. Biol., to appear.

[10] Huang, F.W.D., Peng W.W.J., Reidys, C.M. 2008. *Folding RNA pseudoknot structures*, in preparation

[11] E.Y. Jin, J. Qin and C.M. Reidys, *Combinatorics of RNA structures with pseudoknots* Bull. Math. Biol. **70**(2008), 45-67, PMID: 17896159.

[12] E.Y. Jin and C.M. Reidys, *Asymptotic enumberation of RNA structures with pseudoknots* Bull. Math. Biol. (2007), DOI 10.1007/s11538-007-9265-2.

[13] E.Y. Jin and C.M. Reidys, *RNA-LEGO: Combinatorial Design of Pseudoknot RNA.* Adv. Appl. Math., to appear.

[14] E.Y. Jin and C.M. Reidys, *RNA pseudoknot structures with arc-length* ≥ 3 *and stack-length* $\geq \sigma$. submitted.

[15] E.Y. Jin, C.M. Reidys and R. Wang, *Asymptotic analysis k-noncrossing matchings*, 2008, arXiv:0803.0848

[16] D.A.M. Konings and R.R. Gutell, *A comparison of thermodynamic foldings with comparatively derived structures of 16S and 16S-like rRNAs* RNA. **1**(1995),559-574.

[17] A. Loria and T. Pan, *Domain structure of the ribozyme from eubacterial ribonuclease P* RNA. **2**(1996), 551-563.

[18] R. Lyngso and C. Pedersen, *Pseudoknots in RNA secondary structures* Physics of Biological Systems: From Molecules to Species Sci. (1996).

[19] A.M. Odlyzko, *Asymptotic enumeration methods*, Handbook of combinatorics Vol. 2, 1021-1231, (1995).

[20] N. Parkin, M. Chamorro and H.E. Varmus, *An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA.*J. Proc Natl Acad Sci USA,**89**(1991)713-717.

[21] E. Rivas and S. Eddy, *A Dynamic Programming Algorithm for RNA structure prediction inclusing pseudoknots* J. Mol. Biol. **285**(1999), 2053-2068.

[22] R.P. Stanley, Differentiably finite power series. Europ. J. Combinatorics. **1**, (1980), 175-188.

[23] E.C. Titchmarsh, *The theory of functions* Oxford University Press, London, 1939.

[24] C. Tuerk, S. MacDougal and L. Gold, *RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase.* Proc. Natl. Acad. Sci. USA, **89**(1992)6988-6992.

[25] Y. Uemura, A. Hasegawa, S. Kobayashi and T. Yokomori, *Tree adjoining grammars for RNA structure prediction* Theoret. Comput. Sci. **210**(1999), 277-303.

[26] W. Wasow, *Asymptotic expansions for ordinary differential equations, Dover*(1987) . *A reprint of the John Wiley edition*(1965).

[27] M.S. Waterman, *Combinatorics of RNA hairpins and cloverleafs* Stud. Appl. Math. **60**(1979), 91-96.

[28] M.S. Waterman, *Secondary structure of single - stranded nucleic acids* Adv. Math.I (suppl.) **1**(1978), 167-212.

[29] J.A. Howell, T.F. Smith and M.S. Waterman, *Computation of generating functions for biological molecules* SIAM J. Appl. Math. **39**(1980), 119-133.

[30] M.S. Waterman and T.F. Smith, *Rapid dynamic programming algorithms for RNA secondary structure.* Adv. Appl. Math. **7**(1986), 455-464.

[31] D. Zeilberger, *A Holonomic systems approach to special functions identities.* J. of Computational and Applied Math. **32** (1990), 321-368.

CENTER FOR COMBINATORICS, LPMC-TJKLC, NANKAI UNIVERSITY, TIANJIN 300071, P.R. CHINA, PHONE: *86-22-2350-6800, FAX: *86-22-2350-9272

E-mail address: `reidys@nankai.edu.cn`