

## Stacks in canonical RNA pseudoknot structures

Hillary S.W. Han<sup>a</sup>, Christian M. Reidys<sup>a,b,\*</sup>

<sup>a</sup> Center for Combinatorics, LPMC-TJKLC, Nankai University, Tianjin 300071, PR China

<sup>b</sup> College of Life Sciences, Nankai University, Tianjin 300071, PR China

### ARTICLE INFO

#### Article history:

Received 6 August 2008

Accepted 19 December 2008

Available online 5 February 2009

#### Keywords:

*k*-Non-crossing RNA structure  
Pseudoknot  
Generating function  
Singularity analysis  
Central limit theorem  
Stack

### ABSTRACT

In this paper we study the distribution of stacks/loops in *k*-non-crossing,  $\tau$ -canonical RNA pseudoknot structures ( $\langle k, \tau \rangle$ -structures). Here, an RNA structure is called *k*-non-crossing if it has no more than *k* – 1 mutually crossing arcs and  $\tau$ -canonical if each arc is contained in a stack of length at least  $\tau$ . Based on the ordinary generating function of  $\langle k, \tau \rangle$ -structures [G. Ma, C.M. Reidys, Canonical RNA pseudoknot structures, *J. Comput. Biol.* 15 (10) (2008) 1257] we derive the bivariate generating function  $\mathbf{T}_{k,\tau}(x, u) = \sum_{n \geq 0} \sum_{0 \leq t \leq \frac{n}{2}} T_{k,\tau}(n, t) u^t x^n$ , where  $T_{k,\tau}(n, t)$  is the number of  $\langle k, \tau \rangle$ -structures having exactly *t* stacks and study its singularities. We show that for a specific parametrization of the variable *u*,  $\mathbf{T}_{k,\tau}(x, u)$  exhibits a unique, dominant singularity. The particular shift of this singularity parametrized by *u* implies a central limit theorem for the distribution of stack-numbers. Our results are of importance for understanding the ‘language’ of minimum-free energy RNA pseudoknot structures, generated by computer folding algorithms.

© 2009 Published by Elsevier Inc.

### 1. Introduction

In this paper we compute the bivariate generating function

$$\mathbf{T}_{k,\tau}(x, u) = \sum_{n \geq 0} \sum_{0 \leq t \leq \frac{n}{2}} T_{k,\tau}(n, t) u^t x^n, \quad (1)$$

where  $T_{k,\tau}(n, t)$  is the number of *k*-non-crossing,  $\tau$ -canonical structures having exactly *t* stacks. Furthermore we prove the central limit theorem

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{X_{k,\tau}^n - \mu_{k,\tau} n}{\sqrt{n \sigma_{k,\tau}^2}} < x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \quad (2)$$

where  $X_{k,\tau}^n$  is a random variable with distribution

$$\mathbb{P}(X_{k,\tau}^n = t) = T_{k,\tau}(n, t) / T_{k,\tau}(n), \quad (3)$$

see Fig. 1 and Table 1. Our results allow to compute the stack-numbers of a random *k*-non-crossing,  $\tau$ -canonical RNA structure. Since the latter are obtained via the reflection principle, it is non-trivial and currently not known how to construct such structures with uniform probability in polynomial time.

An RNA molecule is a sequence of the four nucleotides **A**, **G**, **U** and **C** together with the Watson–Crick (**A–U**, **G–C**) and **U–G** base pairing rules. RNA molecules form ‘helical’ structures by pairing

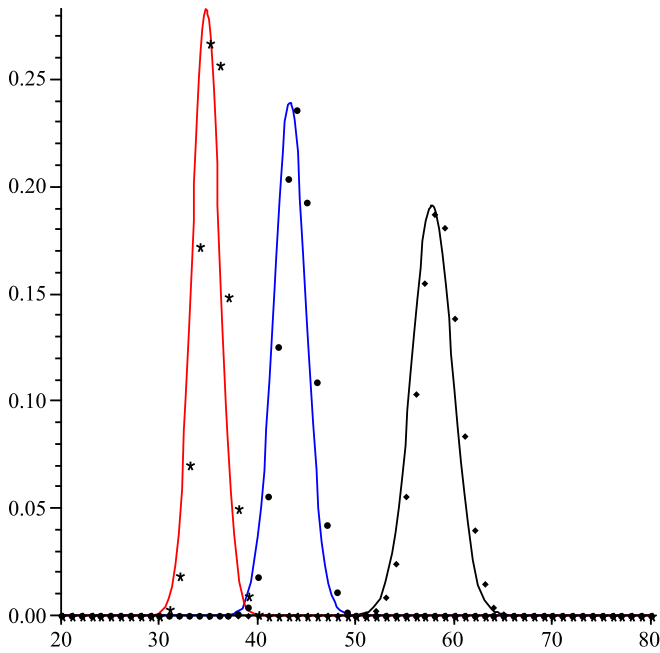
nucleotides and thereby lowering their minimum free energy (mfe). The biochemistry of these nucleotide-pairings favors parallel stacking of bonds due to mesomeric effects. The resulting three-dimensional configuration of the nucleotides is the RNA tertiary structure which determines the functionality of the molecule.

In this paper we study the distribution of stacks in *k*-non-crossing,  $\tau$ -canonical RNA structures. We represent RNA structures as diagrams and identify Watson–Crick (**A–U**, **G–C**) and **U–G** base pairings with arcs drawn in the upper halfplane, ignoring the bonds of the primary sequence. A diagram is a graph over the vertex set  $[n] = \{1, \dots, n\}$  in which each vertex has degree less or equal to one. It is represented by drawing the vertices in a horizontal line and its arcs  $(i, j)$ , where  $i < j$ , in the upper half-plane, see Fig. 2. The vertices and arcs correspond to nucleotides and Watson–Crick (**A–U**, **G–C**) and **U–G** base pairs, respectively. Diagrams have the three key parameters *k*,  $\lambda$  and  $\tau$ . Here *k* – 1 is the maximum number of mutually crossing arcs,  $\lambda$  is the minimum arc-length and  $\tau$  the minimum length of a stack. A  $\lambda$ -arc is an arc  $(i, j)$ , where  $j - i = \lambda$  and a stack of length  $\tau$  is a sequence of ‘parallel’ arcs:  $((i, j), (i + 1, j - 1), \dots, (i + (\tau - 1), j - (\tau - 1)))$ .

We call a *k*-non-crossing diagram with arc-length  $\lambda \geq 4$  and stack-length  $\tau \geq 3$  a *k*-non-crossing,  $\tau$ -canonical RNA structure ( $\langle k, \tau \rangle$ -structure). We denote the number of  $\langle k, \tau \rangle$ -structures and those with exactly *t* stacks by  $T_{k,\tau}(n)$  and  $T_{k,\tau}(n, t)$ , respectively. We furthermore note that, by construction, the numbers of stacks and loops in RNA structures coincide. For *k* = 2 this recovers the well known RNA secondary structures [12,28–30]. Of course, the interesting cases are  $k \geq 3$ , where we allow for crossings, i.e. the RNA pseudoknot structures. However, to the best of our

\* Corresponding author. Address: College of Life Sciences, Nankai University, Tianjin 300071, PR China. Tel.: +86 22 2350 6800; fax: +86 22 2350 9272.

E-mail address: [reidys@nankai.edu.cn](mailto:reidys@nankai.edu.cn) (C.M. Reidys).

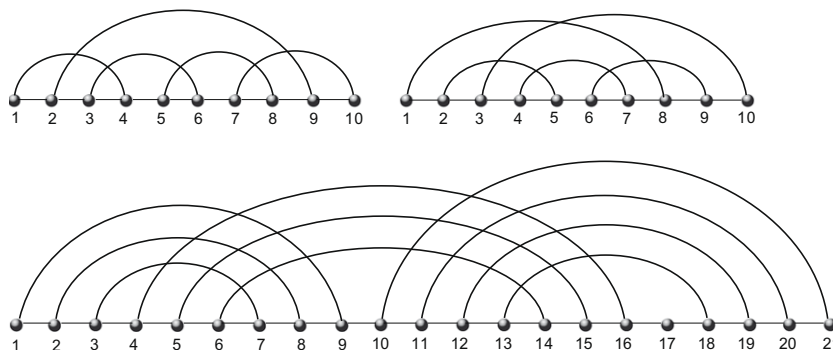


**Fig. 1.** Central limit distributions for the stack-numbers in 3-non-crossing  $\tau$ -canonical RNA structures for  $n = 500$ : we display the limit distribution of Theorem 4.2 for  $\tau = 3$  (black),  $\tau = 4$  (blue) and  $\tau = 5$  (red). We also display the exact enumeration results obtained via Theorem 3.1, Eq. (10) represented by  $\diamond$ ,  $\bullet$  and  $*$ , respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

knowledge, our results are – even for RNA secondary structures new. In Fig. 3 we give two representations of a typical  $\langle 3, 3 \rangle$  RNA pseudoknot structure.

**Table 1**  
The mean  $\mu_{k,\tau}$  and variance  $\sigma_{k,\tau}^2$  of the normal limit distributions of the random variable  $X_{k,\tau}^n$ , for different  $k$  and  $\tau$ .

	$\mu_{k,\tau}$	$\sigma_{k,\tau}^2$	$\mu_{k,\tau}$	$\sigma_{k,\tau}^2$	$\mu_{k,\tau}$	$\sigma_{k,\tau}^2$
	$k = 2$		$k = 3$		$k = 4$	
$\tau = 3$	0.090323	0.0189975	0.115473	0.0086760	0.123509	0.0076977
$\tau = 4$	0.071677	0.0131316	0.086554	0.0055685	0.091737	0.0049917
$\tau = 5$	0.059591	0.0098165	0.069467	0.0039688	0.073166	0.0035769
$\tau = 6$	0.051092	0.0077233	0.058149	0.0026885	0.060964	0.0027313
$\tau = 7$	0.044774	0.0062991	0.050083	0.0017584	0.052319	0.0021788
	$k = 5$		$k = 6$		$k = 7$	
$\tau = 3$	0.128157	0.0070210	0.131353	0.0065187	0.133748	0.0061254
$\tau = 4$	0.094768	0.0020037	0.119551	0.0080515	0.098461	0.0040797
$\tau = 5$	0.075345	0.0033114	0.076864	0.0031162	0.078016	0.0029639
$\tau = 6$	0.062629	0.0025364	0.063794	0.0023936	0.064680	0.0022823
$\tau = 7$	0.053648	0.0020277	0.054580	0.0019171	0.055291	0.0018310



**Fig. 2.**  $k$ -non-crossing diagrams: we display the only two non-planar 3-non-crossing diagrams over 10 vertices (upper) and a 3-non-crossing, 3-canonical diagram with arc-length  $\lambda \geq 4$  (lower).

The results of this paper are relevant for the understanding of minimum free energy (mfe) RNA pseudoknot structures. Computer folding algorithms, like for instance, the dynamic programming routine of Rivas and Eddy [23], can generate, at least for  $n \leq 120$ , a subclass of mfe-pseudoknot structures, reasonably fast. Recently, Huang et al. presented the combinatorial fold cross [14] which generates the above  $\langle k, \tau \rangle$ -structures. Cross can fold mfe-structures for  $n = 100$  with an average folding time of 180s. Of course, there are a number of different folding algorithms employing the dynamic programming [22,5,27,1,19], the maximum matching [10,3,26], the iterated loop matching [24] paradigm, as well as the sampling of RNA structures via the Markov-chain Monte-Carlo method [21].

In light of the existence of pseudoknot folding algorithms, it is now a natural question to compare mfe-structures, generated by one of the above folds with random structures for  $n = 100$ , say. The question then becomes how to obtain the statistics of loops in random RNA pseudoknot structures? While it is well known how to generate a random RNA secondary structure, the particular construction hinges on the fact that the latter can be build inductively [7]. Although  $\langle k, \tau \rangle$ -structures are  $D$ -finite, no inductive recursion exists [16,17] and it is at present time not known how to generate them with uniform probability. In other words, there is no way – other than by exhaustive generation – to derive the distribution of loops in  $\langle k, \tau \rangle$ -structures, an impossible task for  $n = 100$ . Theorem 4.2 proved Section 4 now shows that the distribution of loops is asymptotically normal, also specifying the key parameters of the distribution see Table. 1. In [15] our results have been employed and the authors observe significantly lower numbers of loops in mfe- $\langle k, \tau \rangle$ -structures than in random  $\langle k, \tau \rangle$ -structures.

There is, however, another connection between the results on stack/loop distributions and the complexity of mfe-folding algorithms. The keyword here is that of ‘parameterized complexity’

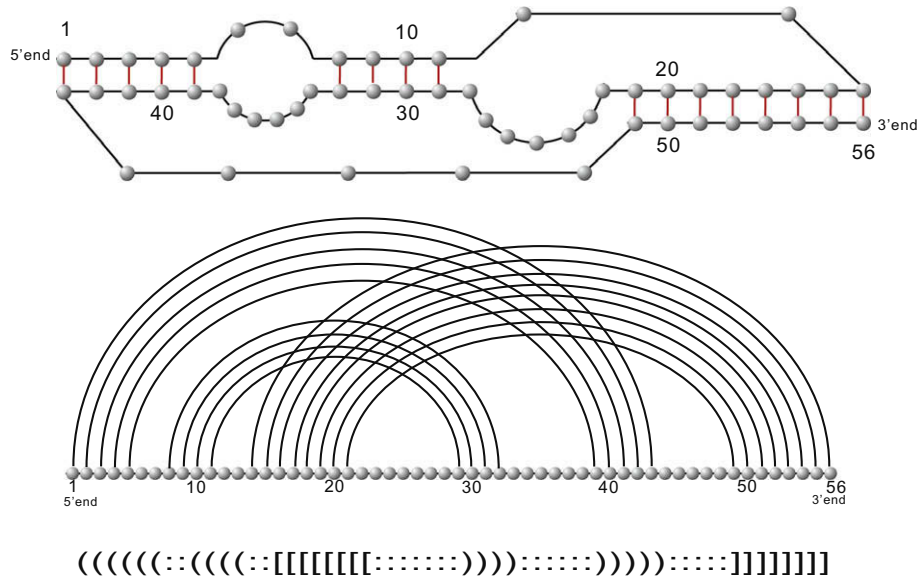


Fig. 3. A 3-non-crossing 3-canonical RNA pseudoknot structure over 56 nucleotides drawn as a planar graph (upper) and as a 3-non-crossing diagram (lower).

[6]. Analysis of cross [14] shows that its most time consuming component is the generation of certain ‘skeleta-trees’. The nodes of these trees are all non-inductive arc-configurations in a potential  $\langle k, \tau \rangle$ -structure for a given input sequence. Combinatorial properties of these trees, for instance their ‘bandwidth’, are intimately connected to the complexity of the algorithm. Therefore detailed knowledge about stacks/loops in RNA pseudoknot structures will translate into important information about the structure of these trees.

Our findings are therefore of relevance not only for understanding the ‘language of RNA’, via relating mfe- and random structures. They also play a role in the design of efficient folding routines.

### 2. Some basic facts

In the following, we shall identify RNA pseudoknot structures with  $k$ -non-crossing diagrams with minimum arc-length  $\geq 4$  and stack-length  $\tau \geq 3$ . From now on we will always assume that any structure has a minimum arc-length greater than or equal to four [13,20]. Let  $T_{k,\tau}(n)$  ( $\mathbb{T}_{k,\tau}(n)$ ) denote the set (number) of  $\langle k, \tau \rangle$ -structures over  $[n]$ . Furthermore, let  $T_{k,\tau}(n, t)$  and  $\mathbb{T}_{k,\tau}(n, t)$  denote the set (number) of  $\langle k, \tau \rangle$ -structures having exactly  $t$  stacks. In addition, let  $T_{k,\tau}(n, h)$  and  $\mathbb{T}_{k,\tau}(n, h)$  denote the set (number) of  $\langle k, \tau \rangle$ -structures having exactly  $h$  arcs. The generating function,  $\mathbf{T}_{k,\tau}(x) = \sum_{n \geq 0} \mathbb{T}_{k,\tau}(n) x^n$ , where  $k \geq 2, \tau \geq 3$  of  $\langle k, \tau \rangle$ -structures has been obtained in [20]. The function is closely related to  $\mathbf{F}_k(x) = \sum_n f_k(2n) x^{2n}$ , the generating function of  $k$ -non-crossing matchings [4]. Beyond functional equations implied by the reflection-principle [11], the following asymptotic formula [18] will be of importance

$$\forall k \in \mathbb{N}; f_k(2n) \sim c_k n^{-((k-1)^2 + (k-1)/2)} (2(k-1))^{2n}, \quad c_k > 0. \quad (4)$$

$\mathbf{T}_{k,\tau}(x)$  is derived in [20] as follows:

**Theorem 2.1.** Let  $k \geq 2, \tau \geq 3, x$  be an indeterminate and  $\rho_k$  be the dominant, positive real singularity of  $\mathbf{F}_k(z)$ . Then  $\mathbf{T}_{k,\tau}(x)$  is given by

$$\mathbf{T}_{k,\tau}(x) = \frac{1}{v_0(x)} \mathbf{F}_k \left( \frac{\sqrt{w_0(x)x}}{v_0(x)} \right), \quad (5)$$

where  $w_0(x) = \frac{x^{2\tau-2}}{1-x^2+x^{2\tau}}$  and  $v_0(x) = 1 - x + w_0(x)x^2 + w_0(x)x^3 + w_0(x)x^4$ . Furthermore,

$$\mathbb{T}_{k,\tau}(n) \sim c_k n^{-(k-1)^2 - \frac{k-1}{2}} \left( \frac{1}{\gamma_{k,\tau}} \right)^n, \quad \text{for } k = 2, 3, \dots, 9 \quad (6)$$

holds, where  $\gamma_{k,\tau}$  is the minimal positive real solution of the equation  $\frac{\sqrt{w_0(x)x}}{v_0(x)} = \rho_k$ .

In addition we shall make use of the following basic theorem on limit distributions:

**Theorem 2.2 (Lévy-Cramér).** Let  $\{\xi_n\}$  be a sequence of random variables and let  $\{\varphi_n(x)\}$  and  $\{F_n(x)\}$  be the corresponding sequences of characteristic and distribution functions. If there exists a function  $\varphi(t)$ , such that  $\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t)$  uniformly over an arbitrary finite interval enclosing the origin, then there exists a random variable  $\xi$  with distribution function  $F(x)$  such that

$$F_n(x) \Rightarrow F(x)$$

uniformly over any finite or infinite interval of continuity of  $F(x)$ .

### 3. Combinatorics of stacks

The objective of this section is to compute the bivariate generating function

$$\mathbf{T}_{k,\tau}(x, u) = \sum_{n \geq 0} \sum_{0 \leq t \leq \frac{n}{2}} \mathbb{T}_{k,\tau}(n, t) u^t x^n. \quad (7)$$

For this purpose let us recall the concept of  $k$ -non-crossing cores developed in [17]. A  $k$ -non-crossing core is a  $k$ -non-crossing diagram in which all stacks have size one. We denote the set and the number of cores over  $[n]$  by  $C_k(n)$  and  $c_k(n)$ , respectively. Furthermore, let  $C_k(n, t)$  and  $c_k(n, t)$  denote the set and number of cores having exactly  $t$  stacks. We consider the arc-sets

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

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

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

$$C_k^*(n, t) = \{\zeta \mid \zeta \in C_k(n, t); \zeta \text{ contains no } \beta\text{-arc}\} \quad (8)$$

$$T_{k,1}^*(n, h) = \{\zeta \mid \zeta \in T_{k,1}(n, h); \zeta \text{ contains no } \beta\text{-arc}\}, \quad (9)$$

where  $T_{k,1}(n, h)$  denotes the set of  $k$ -non-crossing diagrams without 1-arcs having exactly  $h$  arcs. We set  $C_k^*(n, 0) = T_{k,1}^*(n, 0) = 1$  for  $0 \leq n$ .

In our first theorem we shall show that the number of all  $(k, \tau)$ -structures with exactly  $t$  stacks is a sum of the number of  $C_k^*(n, t)$ -diagrams with positive integer coefficients. Let  $f_k(m, \ell)$  denote the number of  $k$ -non-crossing partial matchings having exactly  $\ell$  isolated points.

**Theorem 3.1.** *Suppose we have  $k, t, \tau \in \mathbb{N}$ ,  $k \geq 2$ ,  $\tau \geq 3$  and  $1 \leq t \leq \lfloor n/2 \rfloor$ . Then*

$$T_{k,\tau}(n, t) = \sum_{h \geq \tau} \binom{(1-\tau)t+h-1}{t-1} C_k^*(n+2t-2h, t) \quad (10)$$

$$C_k^*(n, h) = \sum_{b=0}^{h-1} (-1)^{h-b-1} \binom{h-1}{b} T_{k,1}^*(n-2h+2b+2, b+1) \quad (11)$$

for  $h \geq 1$ .

Furthermore,  $T_{k,1}^*(n, h)$  satisfies

$$T_{k,1}^*(n, h) = \sum_{0 \leq j_1+j_2+j_3 \leq h} (-1)^{j_1+j_2+j_3} \lambda(n, j_1, j_2, j_3) f_k(n-2j_1-3j_2-4j_3, n-2h-j_2-2j_3) \quad (12)$$

where

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

**Proof.** First, there exists a mapping from  $(k, \tau)$ -structures with  $t$  stacks into  $C_k^*(n-2(h-t), t)$ :

$$c : T_{k,\tau}(n, t) \rightarrow \bigcup_{\tau \leq h \leq \lfloor \frac{n}{2} \rfloor} C_k^*(n-2(h-t), t), \quad \zeta \mapsto c(\zeta). \quad (13)$$

The core,  $c(\zeta)$ , is obtained in two steps. First we map arcs and isolated vertices as follows:

$$\forall \ell \geq \tau - 1; \quad ((i-\ell, j+\ell), \dots, (i, j)) \mapsto (i, j) \quad \text{and } j \mapsto j \text{ if } j \text{ is isolated.} \quad (14)$$

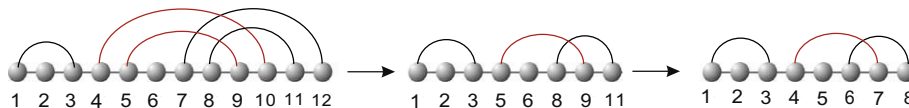
Second we relabel the vertices of the resulting diagram from left to right in increasing order. That is we replace each stack by a single arc, keep isolated vertices and then relabel, see Fig. 4. It is straightforward to verify that  $c : T_{k,\tau}(n, t) \rightarrow \bigcup_{\tau \leq h \leq \lfloor \frac{n}{2} \rfloor} C_k^*(n-2(h-t), t)$  is well-defined and surjective. Taking into account that each stack has its specific length gives rise to consider

$$f_{k,\tau} : T_{k,\tau}(n, t) \rightarrow \bigcup_{\tau \leq h \leq \lfloor \frac{n}{2} \rfloor} \left( C_k^*(n+2t-2h, t) \times \left\{ (\alpha_i)_{1 \leq i \leq t} \mid \sum_{i=1}^t \alpha_i = h-t, \alpha_i \geq \tau-1 \right\} \right), \quad (15)$$

given by  $f_{k,\tau}(\zeta) = (c(\zeta), (\alpha_i)_{1 \leq i \leq t})$ . By construction  $f_{k,\tau}$  is well-defined and a bijection: obviously we can reconstruct the original  $T_{k,\tau}(n, t)$ -element from its core and the sequence of its stack-multiplicities (labeling the stacks from left to right). Computing the multiplicities of the resulting cores we derive

$$\left| \left\{ (\alpha_i)_{1 \leq i \leq t} \mid \sum_{i=1}^t \alpha_i = h-t; \alpha_i \geq \tau-1 \right\} \right| = \binom{(1-\tau)t+h-1}{t-1}. \quad (16)$$

We can therefore conclude



**Fig. 4.** The mapping  $c : T_{k,\tau}(n, t) \rightarrow \bigcup_{\tau \leq h \leq \lfloor \frac{n}{2} \rfloor} C_k^*(n-2(h-t), t)$  is obtained in two steps: first contraction of the stacks and second relabeling the resulting diagram.

$$T_{k,\tau}(n, t) = \sum_{h \geq \tau} \binom{(1-\tau)t+h-1}{t-1} C_k^*(n+2t-2h, t), \quad (17)$$

whence Eq. (10). Next we switch and consider arcs instead of stacks, using the parameter  $h$ . Again, collapsing the stack induces the mapping

$$c^* : T_{k,1}^*(n, h) \rightarrow \bigcup_{0 \leq b \leq h-1} C_k^*(n-2b, h-b) \quad \zeta \mapsto c^*(\zeta). \quad (18)$$

Indeed,  $c^*$  is well defined, since a  $T_{k,1}^*(n, h)$ -diagram can be mapped into a core structure without 1- and  $\beta$ -arcs, i.e. into an element of  $C_k^*(n', h')$ . It is straightforward to show that  $c^*$  is surjective. Labeling the  $h-b$  stacks of  $\zeta \in T_{k,1}^*(n, h)$  from left to right and keeping track of multiplicities gives rise to the map

$$m_k : T_{k,1}^*(n, h) \rightarrow \bigcup_{0 \leq b \leq h-1} \left[ C_k^*(n-2b, h-b) \times \left\{ (a_i)_{1 \leq i \leq h-b} \mid \sum_{i=1}^{h-b} a_i = b, a_i \geq 0 \right\} \right] \quad (19)$$

given by  $m_k(\zeta) = (c^*(\zeta), (a_i)_{1 \leq i \leq h-b})$ . The mapping  $m_k$  is well-defined and a bijection. Clearly,

$$\left| \left\{ (a_i)_{1 \leq i \leq h-b} \mid \sum_{i=1}^{h-b} a_i = b; a_i \geq 0 \right\} \right| = \binom{h-1}{b} \quad (20)$$

and Eqs. (19) and (20) imply

$$T_{k,1}^*(n, h) = \sum_{b=0}^{h-1} \binom{h-1}{b} C_k^*(n-2b, h-b). \quad (21)$$

Via Möbius-inversion Eq. (11) follows. It is straightforward to show that 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,  $j_2$   $\beta_2$ -arcs and  $j_3$   $\beta_3$ -arcs over  $[n]$ . Since removing  $j_1$  1-arcs,  $j_2$   $\beta_2$ -arcs and  $j_3$   $\beta_3$ -arcs removes  $2j_1+3j_2+4j_3$  vertices, the number of configurations of at least  $j_1$  1-arcs,  $j_2$   $\beta_2$ -arcs and  $j_3$   $\beta_3$ -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 the inclusion-exclusion principle, we arrive at

$$T_{k,1}^*(n, h) = \sum_{0 \leq j_1+j_2+j_3 \leq h} (-1)^{j_1+j_2+j_3} \lambda(n, j_1, j_2, j_3) f_k(n-2j_1-3j_2-4j_3, n-2h-j_2-2j_3), \quad (22)$$

whence Theorem 3.1.  $\square$

Of course, Theorem 3.1 implies a functional relation for  $T_{k,\tau}(x, u)$ :

**Lemma 3.2.** *Let  $k \geq 2$ ,  $\tau \geq 3$  and let  $u, x$  be indeterminats. Then we have the functional relation*

$$\sum_{n \geq 0} \sum_{0 \leq t \leq \lfloor \frac{n}{2} \rfloor} T_{k,\tau}(n, t) u^t x^n = \sum_{n \geq 0} \sum_{0 \leq t \leq \lfloor \frac{n}{2} \rfloor} C_k^*(n, t) \left( \frac{ux^{2(\tau-1)}}{1-x^2} \right)^t x^n \quad (23)$$

and in particular, for  $u = 1$

$$\sum_{n \geq 0} T_{k,\tau}(n) x^n = \sum_{n \geq 0} \sum_{0 \leq t \leq \lfloor \frac{n}{2} \rfloor} C_k^*(n, t) \left( \frac{x^{2(\tau-1)}}{1-x^2} \right)^t x^n. \tag{24}$$

**Proof.** We set

$$\sum_{t \geq 1} \left[ \sum_{n \geq 2t\tau} C_k^*(n, t) x^n \right] u^t = \sum_{t \geq 1} \varphi_t(x) u^t$$

and proceed by deducing the functional equation for  $T_{k,\tau}(x, u)$  via **Theorem 3.1**. For this purpose we note that for  $t = 0$  the Binomial coefficient  $\binom{(1-\tau)t+h-1}{t-1}$  is zero, while the term  $T_{k,\tau}(n, 0) = 1$  for  $n \geq 1$ . Clearly,  $T_{k,\tau}(n, 0) = 1$  counts for each  $n \geq 1$  the structure consisting only of isolated vertices. We accordingly have to extend the identity of **Theorem 3.1**

$$T_{k,\tau}(n, t) = \sum_{h \geq t\tau} \binom{\lfloor n/2 \rfloor}{t-1} \binom{(1-\tau)t+h-1}{t-1} C_k^*(n+2t-2h, t)$$

to the case  $t = 0, n \geq 1$  which gives rise to the term  $\sum_{n \geq 1} x^n = \frac{x}{1-x}$ . Accordingly, we derive

$$\sum_{n \geq 0} \sum_{0 \leq t \leq \lfloor \frac{n}{2} \rfloor} T_{k,\tau}(n, t) u^t x^n = \sum_{n \geq 2} \sum_{1 \leq t \leq \lfloor \frac{n}{2} \rfloor} \sum_{h \geq t\tau} C_k^*(n+2t-2h, t) \binom{\lfloor n/2 \rfloor}{t-1} \binom{(1-\tau)t+h-1}{t-1} u^t x^n + \sum_{n \geq 0} x^n. \tag{25}$$

We rewrite the right-hand side of Eq. (25)

$$\begin{aligned} &= \sum_{t \geq 1} \sum_{h \geq t\tau} \sum_{n \geq 2t\tau} C_k^*(n+2t-2h, t) x^{n+2t-2h} \binom{(1-\tau)t+h-1}{t-1} u^t x^{2h-2t} \\ &\quad + \frac{1}{1-x} \\ &= \sum_{t \geq 1} \sum_{h \geq t\tau} \varphi_t(x) \binom{(1-\tau)t+h-1}{t-1} u^t x^{2h-2t} + \frac{1}{1-x}. \end{aligned}$$

Rearranging the terms of the summation we obtain

$$\begin{aligned} &= \sum_{t \geq 1} \sum_{h \geq t\tau} \varphi_t(x) \binom{t-t\tau+h-1}{h-t\tau} x^{2h} \left( \frac{u}{x^2} \right)^t + \frac{1}{1-x} \\ &= \sum_{t \geq 1} \sum_{h \geq t\tau} \varphi_t(x) \binom{h-t\tau+(t-1)}{h-t\tau} (x^2)^{h-t\tau} (ux^{2(\tau-1)})^t + \frac{1}{1-x}. \end{aligned}$$

Using  $\sum_n \binom{n+k}{n} x^n = \frac{1}{(1-x)^{k+1}}$ , we can transform the summation over  $h$  and derive

$$\begin{aligned} \sum_{n \geq 0} \sum_{0 \leq t \leq \lfloor \frac{n}{2} \rfloor} T_{k,\tau}(n, t) u^t x^n &= \sum_{t \geq 1} \varphi_t(x) \left( \frac{1}{1-x^2} \right)^t (ux^{2(\tau-1)})^t + \frac{1}{1-x} \\ &= \sum_{n \geq 0} \sum_{0 \leq t \leq \lfloor \frac{n}{2} \rfloor} C_k^*(n, t) \left( \frac{ux^{2(\tau-1)}}{1-x^2} \right)^t x^n \end{aligned}$$

and the proof of **Lemma 3.2** is complete.  $\square$

We next consider a functional equation for  $\sum_{n,h} T_{k,\tau}^*(n, h) u^h x^n$  proved in [17].

**Lemma 3.3.** Let  $k, \tau \in \mathbb{N}, k \geq 2$  and let  $u, x$  be indeterminats. Suppose we have

$$\forall h \geq 1, A_{k,\tau}(n, h) = \sum_{b=\tau-1}^{h-1} \binom{b+(2-\tau)(h-b)-1}{h-b-1} \times B_k(n-2b, h-b) \text{ and } A_{k,\tau}(n, 0) = 1. \tag{26}$$

Then we have the functional relation

$$\sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} A_{k,\tau}(n, h) u^h x^n = \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} B_k(n, h) \left( \frac{u \cdot (ux^2)^{\tau-1}}{1-ux^2} \right)^h x^n. \tag{27}$$

Combining **Lemma 3.3** and  $T_{k,1}^*(n, h) = \sum_{b=0}^{h-1} \binom{h-1}{b} C_k^*(n-2b, h-b)$  we arrive at

$$\sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} T_{k,1}^*(n, h) u^h x^n = \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} C_k^*(n, h) \left( \frac{u}{1-ux^2} \right)^h x^n. \tag{28}$$

We shall make use of an additional relation for  $\sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} T_{k,1}^*(n, h) u^h x^n$ , proved in [20]:

**Lemma 3.4.** Let  $k \in \mathbb{N}, k \geq 2$ , and  $x, w$  be indeterminats. Then we have the functional relation

$$\sum_{n \geq 0} \sum_{h \leq \frac{n}{2}} T_{k,1}^*(n, h) w^h x^n = \frac{1}{v(x)} \sum_{n \geq 0} f_k(2n) \left( \frac{\sqrt{wx}}{v(x)} \right)^{2n}, \tag{29}$$

where  $v(x) = 1-x+wx^2+wx^3+wx^4$  and  $f_k(2n)$  is the number of  $k$ -non-crossing matchings over  $2n$  vertices.

We are now in the position to present the main result of this section:

**Theorem 3.5.** Let  $k \geq 2, \tau \geq 3$  and suppose  $u, x$  are indeterminats. Then we have the identity of formal power series

$$\sum_{n \geq 0} \sum_{0 \leq t \leq \frac{n}{2}} T_{k,\tau}(n, t) u^t x^n = \frac{1}{v(x)} \sum_{n \geq 0} f_k(2n) \left( \frac{\sqrt{u_0 x}}{v(x)} \right)^{2n}, \tag{30}$$

where  $v(x)$  and  $u_0 = u_0(x, u)$  are given by

$$u_0 = \frac{ux^{2(\tau-1)}}{ux^{2\tau} - x^2 + 1} \tag{31}$$

$$v(x) = 1-x+u_0x^2+u_0x^3+u_0x^4. \tag{32}$$

In particular, we can consider Eq. (30) as a relation between analytic functions, valid for  $u = e^s$  and  $|s| < \epsilon$  for  $\epsilon$  sufficiently small and  $|x| \leq 1/2$ .

**Proof.** We interpret the bivariate generating function

$$\sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} C_k^*(n, h) y^h x^n$$

in two different ways: first, via **Lemma 3.2** we have

$$\sum_{n \geq 0} \sum_{0 \leq t \leq \lfloor \frac{n}{2} \rfloor} T_{k,\tau}(n, t) u^t x^n = \sum_{n \geq 0} \sum_{0 \leq t \leq \lfloor \frac{n}{2} \rfloor} C_k^*(n, t) \left( \frac{ux^{2(\tau-1)}}{1-x^2} \right)^t x^n \tag{33}$$

and second according to Eq. (28):

$$\sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} T_{k,1}^*(n, h) u_0^h x^n = \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} C_k^*(n, h) \left( \frac{u_0}{1-u_0x^2} \right)^h x^n. \tag{34}$$

Note that, by definition of cores, we may replace the index ‘ $h$ ’ by ‘ $t$ ’ in the right hand side of Eq. (34). The key observation is the relation between the terms  $\frac{ux^{2(\tau-1)}}{1-x^2}$  and  $\frac{u_0}{1-u_0x^2}$ . Using the Ansatz

$$\frac{ux^{2(\tau-1)}}{1-x^2} = \frac{u_0}{1-u_0x^2}, \tag{35}$$

we obtain the unique solution  $u_0 = \frac{ux^{2(\tau-1)}}{1-x^2+ux^{2\tau}}$ . Accordingly, we conclude

$$\begin{aligned} \sum_{n \geq 0} \sum_{0 \leq t \leq \lfloor \frac{n}{2} \rfloor} T_{k,\tau}(n, t) u^t x^n &= \sum_{n \geq 0} \sum_{0 \leq t \leq \lfloor \frac{n}{2} \rfloor} C_k^*(n, t) \left( \frac{ux^{2(\tau-1)}}{1-x^2} \right)^t x^n \\ &= \sum_{n \geq 0} \sum_{0 \leq h \leq \frac{n}{2}} T_{k,1}^*(n, h) u_0^h x^n \\ &= \frac{1}{v(x)} \sum_{n \geq 0} f_k(2n) \left( \frac{\sqrt{u_0 x}}{v(x)} \right)^{2n}, \end{aligned}$$

where  $v(x) = 1 - x + u_0x^2 + u_0x^3 + u_0x^4$ , and  $u_0 = \frac{ux^{2(\tau-1)}}{ux^{2\tau} - x^2 + 1}$ . In order to consider Eq. (30) as a relation between analytic functions we need to satisfy  $ux^{2\tau} - x^2 + 1 \neq 0$ . Suppose  $u = e^s$ ,  $|s| < \epsilon$  for  $\epsilon$  sufficiently small. From  $|x| \leq 1/2$  and the continuity (in  $s$ ) of the roots of the family of polynomials

$$\omega_s(X) = e^s X^{2\tau} - X^2 + 1, \quad |s| < \epsilon$$

we conclude  $\omega_s(x) \neq 0$ . Therefore Eq. (30) holds for  $u = e^s$ ,  $|s| < \epsilon$  and sufficiently small  $\epsilon$  and  $|x| \leq 1/2$ .  $\square$

**4. The central limit theorem**

Suppose  $\epsilon > 0$ ,  $2 \leq k$  and  $u$  is parametrized as  $u = e^s$ , where  $|s| < \epsilon$ . We set

$$\varphi_{n,k,\tau}(s) = \sum_{t \leq \frac{n}{2}} T_{k,\tau}(n, t) e^{ts} \tag{36}$$

$$\mathbf{U}_k(z, s) = \sum_{n \geq 0} \varphi_{n,k,\tau}(s) z^n. \tag{37}$$

We will use Theorem 3.5, which relates the generating functions  $\mathbf{U}_k(z, s)$  and  $\mathbf{F}_k(z) = \sum_n f(2n) z^{2n}$  in order to study the singularities of  $\mathbf{U}_k(z, s)$ .

**Theorem 4.1.** *Suppose  $\epsilon > 0$ ,  $2 \leq k \leq 9$ ,  $3 \leq \tau \leq 7$  and  $u = e^s$ , where  $|s| < \epsilon$ . Then the following assertions hold:*

- (a) *For sufficiently small  $\epsilon$ ,  $\mathbf{U}_k(z, s)$  has the unique singularity,  $\gamma_{k,\tau}(s)$ , which is analytic in  $s$  and the unique minimal real positive solution of*

$$\frac{\sqrt{u_0(s)z}}{1 - z + u_0(s)z^2 + u_0(s)z^3 + u_0(s)z^4} - \rho_k = 0. \tag{38}$$

- (b) *The coefficients of  $\mathbf{U}_k(z, s)$  are asymptotically given by*

$$[z^n] \mathbf{U}_k(z, s) = A(s) (1 - O(n^{-1})) n^{-((k-1)^2 + \frac{k-1}{2})} \left( \frac{1}{\gamma_{k,\tau}(s)} \right)^n, \quad A(s) \in \mathbb{C}, \tag{39}$$

uniformly in  $s$  in a neighborhood of 0. In particular, the subexponential factors of the coefficients of  $\mathbf{U}_k(z, s)$  coincide with those of  $\mathbf{F}_k(z)$  and are independent of  $s$ .

**Proof.** According to Theorem 3.5 we have for  $|s| < \epsilon$

$$\mathbf{U}_k(z, s) = \frac{1}{v(z, s)} \sum_{n \geq 0} f_k(2n) \left( \frac{\sqrt{u_0(s)z}}{v(z, s)} \right)^{2n} \tag{40}$$

where  $v(z, s) = 1 - z + u_0(s)z^2 + u_0(s)z^3 + u_0(s)z^4$  and  $u_0(s) = \frac{e^{s2(\tau-1)}}{e^{s2\tau} - z^2 + 1}$ . We set

$$\psi_\tau(z, s) = \frac{\sqrt{u_0(s)z}}{1 - z + u_0(s)z^2 + u_0(s)z^3 + u_0(s)z^4} \tag{41}$$

$$\mathbf{W}_k(z, s) = \sum_{n \geq 0} f_k(2n) \left( \frac{\sqrt{u_0(s)z}}{v(z, s)} \right)^{2n}. \tag{42}$$

Our first objective is to prove the existence of the unique singularity,  $\gamma_{k,\tau}(s)$ . For this purpose we consider

$$F(z, s) = \psi_\tau(z, s) - \rho_k. \tag{43}$$

For  $s = 0$  and  $2 \leq k \leq 9$ ,  $3 \leq \tau \leq 7$  it is straightforward to verify that a unique minimal real solution,  $\omega$ , exists. We observe that for  $|s| < \epsilon$  the following holds: (i)  $F(\omega, 0) = 0$ , (ii)  $F_z(\omega, 0) \neq 0$  and (iii) the partial derivatives  $F_z(z, s)$  and  $F_s(z, s)$  are continuous. According to the analytic implicit function theorem [9], there exists in a sufficiently

small neighborhood of 0, a unique analytic function  $\gamma_{k,\tau}(s)$ , that satisfies

$$\forall s; |s| < \epsilon; \quad F(\gamma_{k,\tau}(s), s) = 0 \quad \text{and} \quad \gamma_{k,\tau}(0) = \omega.$$

*Claim 1.* For  $\epsilon$  sufficiently small and  $|s| < \epsilon$ ,  $\gamma_{k,\tau}(s)$  is a dominant singularity of  $\mathbf{U}_k(z, s)$ .

Let  $\zeta(s)$  be a dominant singularity of  $\mathbf{U}_k(z, s)$ . It is straightforward to prove that  $\zeta(0)$  is necessarily a singularity of  $\mathbf{W}_k(z, 0)$ . We proceed by applying an continuity argument. For  $\epsilon$  sufficiently small and  $|s| < \epsilon$  the singularities of  $v(z, s)^{-1}$  and  $\gamma_{k,\tau}(s)$  are both continuous in  $s$ . Therefore we can conclude from our observation for  $s = 0$  that, for sufficiently small  $\epsilon$ , all singularities of  $v(z, s)^{-1}$  have modulus strictly larger than  $\gamma_{k,\tau}(s)$ . Accordingly, for sufficiently small  $\epsilon$ ,  $\gamma_{k,\tau}(s)$  is a dominant singularity of  $\mathbf{U}_k(z, s)$ , whence Claim 1.

*Claim 2.*  $\gamma_{k,\tau}(s)$  is unique.

Functional relations arising from the reflection principle [11] imply that the generating function  $\mathbf{F}_k(z)$  is  $D$ -finite [25]. Accordingly there exists some  $e \in \mathbb{N}$  for which  $\mathbf{F}_k(z)$  satisfies an ODE of the form

$$q_{0,k}(z) \frac{d^e}{dz^e} \mathbf{F}_k(z) + q_{1,k}(z) \frac{d^{e-1}}{dz^{e-1}} \mathbf{F}_k(z) + q_{e,k}(z) \mathbf{F}_k(z) = 0, \tag{44}$$

where  $q_{i,k}(z)$  are polynomials. Any dominant singularity of  $\mathbf{F}_k(z)$  is contained in the set of roots of  $q_{0,k}(z)$  [25] and via direct computation we can verify that  $\gamma_{k,\tau}(0)$  is the unique solution with minimal modulus of all solutions of

$$\psi_\tau(z, 0) = |\rho_k|. \tag{45}$$

The analytic implicit function theorem applied to Eq. (43) now implies locally the existence of an unique analytic function  $\gamma_{k,\tau}(s)$  solving  $\psi_\tau(z, s) = \rho_k$ . Using continuity we conclude from the fact that  $\gamma_{k,\tau}(0)$  is the unique solution with minimal modulus of all solutions of  $\psi_\tau(z, 0) = |\rho_k|$ , that for  $\epsilon$  sufficiently small,  $\gamma_{k,\tau}(s)$  is the unique solution with minimal modulus of all solutions of  $\psi_\tau(z, s) = |\rho_k|$ . This implies Claim 2 and (a) follows.

We finally prove (b). According to Eq. (4) we have

$$f_k(2n) \sim c_k n^{-((k-1)^2 + (k-1)/2)} (2(k-1))^{2n}$$

for some  $c_k > 0$  and

$$\mathbf{F}_k(z) = \begin{cases} O((z - \rho_k)^{(k-1)^2 + (k-1)/2 - 1} \ln(z - \rho_k)) & \text{for } k \text{ odd, } z \rightarrow \rho_k \\ O((z - \rho_k)^{(k-1)^2 + (k-1)/2 - 1}) & \text{for } k \text{ even, } z \rightarrow \rho_k, \end{cases}$$

in accordance with basic structure theorems for singularities of solutions of Eq. (44) [9], p. 499. Let  $Q_{\gamma_{k,\tau}(s)}(z, s)$  denote the singular expansion of  $\mathbf{U}_k(z, s)$  at  $\gamma_{k,\tau}(s)$ . We have shown in Claim 1 that  $\psi_\tau(z, s)$  does not induce any dominant singularities and is regular at  $\rho_k$ . Let  $Q_{\rho_k}(z)$  denote the singular expansion of  $\mathbf{F}_k(z)$  at the dominant singularity  $\rho_k$ , i.e.  $F_k(z) = O(Q_{\rho_k}(z))$  for  $z \rightarrow \rho_k$ . The singular expansion of the compositum,  $\mathbf{F}_k(\psi_\tau(z, s))$ , i.e.  $Q_{\gamma_{k,\tau}(s)}(z, s)$ , is derived by substituting the Taylor-expansion of  $\psi_\tau(z, s)$  into  $Q_{\rho_k}(z)$  and we observe

$$Q_{\gamma_{k,\tau}(s)}(z, s) = Q_{\rho_k}(\psi_\tau(\zeta_k(s), s)) = O(Q_{\gamma_{k,\tau}(s)}(z)). \tag{46}$$

Indeed, Eq. (46) follows immediately substituting  $\psi_\tau(z, s) - \psi_\tau(\gamma_{k,\tau}(s), s)$  for  $z - \rho_k$  which does not change the singular expansion. Since  $\psi_\tau(z, s)$  is continuous in  $s$  the singular expansion is uniform with respect to the parameter  $s$  and the uniformity lemma of singularity analysis [8,9] implies

$$[z^n] \mathbf{U}_k(z, s) = A(s)(1 - O(1/n))n^{-((k-1)^2+(k-1)/2)} \left( \frac{1}{\gamma_{k,\tau}(s)} \right)^n \quad (47)$$

for some  $A(s) \in \mathbb{C}$ , uniformly in  $s$  in a neighborhood of 0. Therefore the asymptotic expansion is uniform in  $s$  and Eq. (39) follows. The proof shows in addition that the subexponential factors of the coefficients of  $\mathbf{U}_k(z, s)$  coincide with those of  $\mathbf{F}_k(z)$  and are independent of  $s$ .  $\square$

We now consider the random variable  $X_{k,\tau}^n$  having the distribution

$$\mathbb{P}(X_{k,\tau}^n = t) = T_{k,\tau}(n, t) / T_{k,\tau}(n) \quad (48)$$

where  $t = 0, 1, \dots, \lfloor n/2 \rfloor$ . We shall show in the following theorem that the distribution of  $X_{k,\tau}^n$  is determined by the shift of the singularity parametrized by  $s$ . In other words, Theorem 4.1 induces the below central limit theorem. The proof-idea of Theorem 4.2 is due to Bender, where it appeared in slightly different context [2]. It follows from analyzing the characteristic function, using the Lévy–Cramér Theorem of Section 2.

**Theorem 4.2.** For any  $2 \leq k \leq 9$  and  $3 \leq \tau \leq 7$  there exist a pair  $(\mu_{k,\tau}, \sigma_{k,\tau})$  such that the normalized random variable

$$Y_{k,\tau}^n = \frac{X_{k,\tau}^n - \mu_{k,\tau}n}{\sqrt{n\sigma_{k,\tau}^2}} \quad (49)$$

has asymptotically normal distribution with parameter  $(0, 1)$ . That is we have

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{X_{k,\tau}^n - \mu_{k,\tau}n}{\sqrt{n\sigma_{k,\tau}^2}} < x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}c^2} dc, \quad (50)$$

where  $\mu_{k,\tau}$  and  $\sigma_{k,\tau}^2$  are given by

$$\mu_{k,\tau} = -\frac{\gamma'_{k,\tau}(0)}{\gamma_{k,\tau}(0)}, \quad \sigma_{k,\tau}^2 = \left( \frac{\gamma'_{k,\tau}(0)}{\gamma_{k,\tau}(0)} \right)^2 - \frac{\gamma''_{k,\tau}(0)}{\gamma_{k,\tau}(0)}. \quad (51)$$

**Proof.** Suppose we are given the random variable (r.v.)  $\xi_n$  with mean  $\mu_n$  and variance  $\sigma_n^2$ . We consider the rescaled r.v.  $\eta_n = (\xi_n - \mu_n)\sigma_n^{-1}$  and the characteristic function of  $\eta_n$ :

$$f_{\eta_n}(c) = \mathbb{E}[e^{ic\eta_n}] = \mathbb{E}\left[ e^{ic\frac{\xi_n - \mu_n}{\sigma_n}} \right] e^{-\frac{i\mu_n c}{\sigma_n}}. \quad (52)$$

Writing  $X_n$  instead of  $X_{k,\tau}^n$  we derive for  $\xi_n = X_n$ , substituting for the term  $\mathbb{E}[e^{ic\eta_n}]$

$$f_{X_n}(c) = \left( \sum_{t=0}^{n/2} \frac{T_{k,\tau}(n, t)}{T_{k,\tau}(n)} e^{ic\frac{t}{\sigma_n}} \right) e^{-\frac{i\mu_n c}{\sigma_n}}. \quad (53)$$

In view of  $\varphi_{n,k,\tau}(s) = \sum_{t \leq n/2} T_{k,\tau}(n, t) e^{ts}$  we interpret

$$\begin{aligned} \varphi_{n,k,\tau}(0) &= \sum_{t \leq n/2} T_{k,\tau}(n, t) \quad \text{and} \quad \varphi_{n,k,\tau}(ic/(\sigma_n)) \\ &= \sum_{t \leq n/2} T_{k,\tau}(n, t) e^{t(ic/(\sigma_n))}. \end{aligned}$$

Writing  $\varphi_n$  instead of  $\varphi_{n,k,\tau}$ , we accordingly obtain

$$f_{X_n}(c) = \frac{1}{\varphi_n(0)} \varphi_n \left( \frac{ic}{\sigma_n} \right) e^{-\frac{i\mu_n c}{\sigma_n}}. \quad (54)$$

Now we have to provide the interpretation of  $\varphi_n(0)$  and  $\varphi_n(ic/(\sigma_n))$ . This is facilitated via Theorem 4.1:

$$[z^n] \mathbf{U}_k(z, s) = K(s)\theta_k(n)(\gamma_{k,\tau}(s))^{-n}(1 - O(1/n)) \quad (55)$$

for some  $K(s) \in \mathbb{C}$ ,

uniformly in  $s$  and where  $\theta_k(n)$  is some subexponential factor, independent of  $s$  (we showed that the singular expansion remains invariant when substituting  $\psi_\tau(z, s)$  for  $z$ ). Therefore

$$f_{X_n}(c) \sim \frac{K(\frac{ic}{\sigma_n})}{K(0)} \left[ \frac{\gamma_{k,\tau}(\frac{ic}{\sigma_n})}{\gamma_{k,\tau}(0)} \right]^{-n} e^{-\frac{i\mu_n c}{\sigma_n}}, \quad (56)$$

uniformly in  $c$ , where  $c$  is contained in an arbitrary bounded interval. Taking the logarithm we obtain

$$\ln f_{X_n}(c) \sim \ln \frac{K(\frac{ic}{\sigma_n})}{K(0)} - n \ln \frac{\gamma_{k,\tau}(\frac{ic}{\sigma_n})}{\gamma_{k,\tau}(0)} - i \frac{\mu_n}{\sigma_n} c. \quad (57)$$

Expanding  $g(s) = \ln(\gamma_{k,\tau}(s)/\gamma_{k,\tau}(0))$  in its Taylor series at  $s = 0$  (note that  $g(0) = 0$  holds) yields

$$\ln \frac{\gamma_{k,\tau}(\frac{ic}{\sigma_n})}{\gamma_{k,\tau}(0)} = \frac{\gamma'_{k,\tau}(0)}{\gamma_{k,\tau}(0)} \frac{ic}{\sigma_n} - \left[ \frac{\gamma''_{k,\tau}(0)}{\gamma_{k,\tau}(0)} - \left( \frac{\gamma'_{k,\tau}(0)}{\gamma_{k,\tau}(0)} \right)^2 \right] \frac{c^2}{2\sigma_n^2} + O\left( \left( \frac{ic}{\sigma_n} \right)^3 \right) \quad (58)$$

and  $\ln f_{X_n}(c)$  becomes asymptotically

$$\begin{aligned} \ln \frac{K(\frac{ic}{\sigma_n})}{K(0)} - n \left\{ \frac{\gamma'_{k,\tau}(0)}{\gamma_{k,\tau}(0)} \frac{ic}{\sigma_n} - \frac{1}{2} \left[ \frac{\gamma''_{k,\tau}(0)}{\gamma_{k,\tau}(0)} - \left( \frac{\gamma'_{k,\tau}(0)}{\gamma_{k,\tau}(0)} \right)^2 \right] \frac{c^2}{\sigma_n^2} \right. \\ \left. + O\left( \left( \frac{ic}{\sigma_n} \right)^3 \right) \right\} - \frac{i\mu_n c}{\sigma_n}. \end{aligned} \quad (59)$$

$\mathbf{U}_k(z, s)$  is analytic in  $s$  where  $s$  is contained in a disc of radius  $\epsilon$  around 0 and therefore in particular continuous in  $s$  for  $|s| < \epsilon$ . In view of Eq. (59) we set

$$\mu = -\frac{\gamma'_{k,\tau}(0)}{\gamma_{k,\tau}(0)}, \quad \sigma^2 = \left( \frac{\gamma'_{k,\tau}(0)}{\gamma_{k,\tau}(0)} \right)^2 - \frac{\gamma''_{k,\tau}(0)}{\gamma_{k,\tau}(0)}$$

Setting  $\mu_n = n\mu$  and  $\sigma_n^2 = n\sigma^2$  we can conclude from Eq. (55) for fixed  $c \in ]-\infty, \infty[$

$$\lim_{n \rightarrow \infty} (\ln K((ic)/(\sigma_n)) - \ln K(0)) = 0 \quad (60)$$

and Eq. (59) becomes

$$\ln f_{X_n}(c) \sim -c^2/2 + O\left( (ic/(\sigma_n))^3 \right) \quad (61)$$

with uniform error term for  $c$  from any bounded interval. This is equivalent to  $\lim_{n \rightarrow \infty} f_{X_n}(c) = \exp(-c^2/2)$  uniformly in  $c$ . The Lévy–Cramér theorem of Section 2 implies now Eq. (50) and the proof of Theorem 4.2 is complete.  $\square$

### Acknowledgments

We are grateful to Gang Ma for his comments. This work was supported by the 973 Project, the PCSIRT Project of the Ministry of Education, the Ministry of Science and Technology, and the National Science Foundation of China.

### References

- [1] T. Akutsu, Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots, *Discr. Appl. Math.* 104 (2000) 45.
- [2] E.A. Bender, Central and local limit theorems applied to asymptotic enumeration, *J. Comb. Theory A* 15 (1973) 91111.
- [3] R. Cary, G. Stormo, Graph-theoretic approach to RNA modeling using comparative data, *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 3 (1995) 75.
- [4] W.Y.C. Chen, E.Y.P. Deng, R.R.X. Du, R.P. Stanley, C.H. Yan, Crossings and nestings of matchings and partitions, *Trans. Am. Math. Soc.* 359 (4) (2007) 1555.
- [5] R.M. Dirks, N.A. Pierce, An algorithm for computing nucleic acid base-pairing probabilities including pseudoknots, *J. Comput. Chem.* 25 (10) (2004) 1295.
- [6] R.G. Downey, M.R. Fellows, *Parameterized Complexity*, Springer, 1999.

- [7] P. Ducheon, P. Flajolet, G. Louchard, G. Schaeffer, Boltzman samplers for the random generation of combinatorial structures, *Comb. Probab. Comput.* 13 (4–5) (2004) 577.
- [8] P. Flajolet, J.A. Fill, N. Kapur, Singularity analysis, Hadamard products, and tree recurrences, *J. Comput. Appl. Math.* 174 (2005) 271.
- [9] P. Flajolet, R. Sedgewick, *Analytic Combinatorics*, Cambridge University, 2008.
- [10] H.N. Gabow, An efficient implementation of Edmonds' algorithm for maximum matching on graphs, *J. Assoc. Comput. Mach.* 23 (1976) 221.
- [11] D.J. Grabiner, P. Magyar, Random walks in Weyl chambers and the decomposition of tensor powers, *Discr. Appl. Math.* 2 (1993) 239.
- [12] J.A. Howell, T.F. Smith, M.S. Waterman, Computation of generating functions for biological molecules, *SIAM J. Appl. Math.* 39 (1980) 119.
- [13] H.S.W. Han, C.M. Reidys, Pseudoknot RNA structures with arc-length  $\geq 4$ , *J. Comput. Biol.* 15 (9) (2008) 1195.
- [14] F.W.D. Huang, W.W.J. Peng, C.M. Reidys, Folding 3-noncrossing RNA pseudoknot structures. Available from: <<http://www.combinatorics.cn/cbpc//>>, 2008.
- [15] F.W.D. Huang, L.Y.M. Li, C.M. Reidys, Sequence–structure relations of pseudoknot RNA, *BMC Bioinform.* 10 (Suppl. 1) (2009) S39, doi:10.1186/1471-2105-10-S1-S39.
- [16] E.Y. Jin, J. Qin, C.M. Reidys, Combinatorics of RNA structures with pseudoknots, *Bull. Math. Biol.* 70 (1) (2008) 45.
- [17] E.Y. Jin, C.M. Reidys, RNA-LEGO: combinatorial design of pseudoknot RNA, *Adv. Appl. Math.* 42 (2009) 135–151.
- [18] E.Y. Jin, C.M. Reidys, R.R. Wang, Asymptotic analysis of k-noncrossing matchings. Available from: 0803.0848, 2008.
- [19] R.B. Lyngsø, C.N.S. Pedersen, RNA pseudoknot prediction in energy-based models, *J. Comput. Biol.* 7 (2000) 409.
- [20] G. Ma, C.M. Reidys, Canonical RNA pseudoknot structures, *J. Comput. Biol.* 15 (10) (2008) 1257.
- [21] D. Metzler, M.E. Nebel, Predicting RNA secondary structures with pseudoknots by mcmc sampling, *J. Math. Biol.* 56 (1–2) (2008) 161.
- [22] J. Reeder, R. Giegerich, Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics, *BMC Bioinform.* 5 (104) (2004), doi:10.1186/1471-2105-5-104.
- [23] E. Rivas, S. Eddy, A dynamic programming algorithm for RNA structure prediction including pseudoknots, *J. Mol. Biol.* 285 (1999) 2053.
- [24] J. Ruan, G. Stormo, W. Zhang, An iterated loop matching approach to the prediction, *Bioinformatics* 20 (2004) 58.
- [25] R. Stanley, Differentiably finite power series, *Eur. J. Comb.* 1 (1980) 175.
- [26] J. Tabaska, R. Cary, H. Gabow, G. Stormo, An RNA folding method capable of identifying pseudoknots and base triples, *Bioinformatics* 14 (1998) 691.
- [27] Y. Uemura, A. Hasegawa, S. Kobayashi, T. Yokomori, Tree adjoining grammars for RNA structure prediction, *Theor. Comput. Sci.* 210 (1999) 277.
- [28] M.S. Waterman, Secondary structure of single-stranded nucleic acids, *Adv. Math.* 11 (Suppl.) (1978) 167.
- [29] M.S. Waterman, Combinatorics of RNA hairpins and cloverleaves, *Stud. Appl. Math.* 60 (1979) 91.
- [30] M.S. Waterman, W.R. Schmitt, Linear trees and RNA secondary structure, *Discr. Appl. Math.* 51 (1994) 317.