On the uniform generation of modular diagrams

Fenix W.D. Huang[1] and Christian M. Reidys[2]

Center for Combinatorics, LPMC-TJKLC
Nankai University, Tianjin 300071, P.R. China

[1]fenixprotoss@gmail.com, [2]duck@santafe.edu

Submitted: May 29, 2010; Accepted: Dec 1, 2010; Published: Dec 10, 2010
Mathematics Subject Classifications: 05A18

Abstract

In this paper we present an algorithm that generates k-noncrossing, σ-modular diagrams with uniform probability. A diagram is a labeled graph of degree ≤ 1 over n vertices drawn in a horizontal line with arcs (i, j) in the upper half-plane. A k-crossing in a diagram is a set of k distinct arcs $(i_1, j_1), (i_2, j_2), \ldots, (i_k, j_k)$ with the property $i_1 < i_2 < \ldots < i_k < j_1 < j_2 < \ldots < j_k$. A diagram without any k-crossings is called a k-noncrossing diagram and a stack of length σ is a maximal sequence $((i, j), (i+1, j-1), \ldots, (i+(\sigma-1), j-(\sigma-1)))$. A diagram is σ-modular if any arc is contained in a stack of length at least σ. Our algorithm generates after $O(n^k)$ preprocessing time, k-noncrossing, σ-modular diagrams in $O(n)$ time and space complexity.

Keywords: k-noncrossing diagram, uniform generation, RSK-algorithm

1 Introduction

A ribonucleic acid (RNA) molecule is the helical configuration of a primary structure of nucleotides, **A**, **G**, **U** and **C**, together with Watson-Crick (**A**-**U**, **G**-**C**) and (**U**-**G**) base pairs (arcs). It is well-known that RNA structures exhibit cross-serial nucleotide interactions, called pseudoknots. First recognized in the turnip yellow mosaic virus in [14], they are now known to be widely conserved in functional RNA molecules.

Modular k-noncrossing diagrams represent a model of RNA pseudoknot structures [10, 11], that is RNA structures exhibiting cross-serial base pairings. The particular case of modular noncrossing diagrams, i.e. RNA secondary structures have been extensively studied [8, 12, 15, 16, 17].

A diagram is a labeled graph over the vertex set $[n] = \{1, \ldots, n\}$ with vertex degrees not greater than one. The standard representation of a diagram is derived by drawing its vertices in a horizontal line and its arcs (i, j) in the upper half-plane. A k-crossing is a set of k distinct arcs $(i_1, j_1), (i_2, j_2), \ldots, (i_k, j_k)$ with the property

$$i_1 < i_2 < \ldots < i_k < j_1 < j_2 < \ldots < j_k. \tag{1.1}$$

A diagram without any k-crossings is called a k-noncrossing diagram. Furthermore, a stack of length σ is a maximal sequence of "parallel" arcs,

$$((i, j), (i + 1, j - 1), \ldots, (i + (\sigma - 1), j - (\sigma - 1)))$$

and is also referred to as a σ-stack. A k-noncrossing diagram having only stacks of lengths one is called a core.

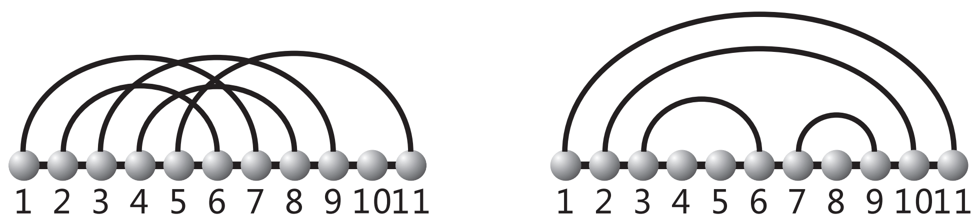

Figure 1: k-noncrossing diagrams: a 4-noncrossing diagram (left) and a 2-noncrossing diagram (right). The arcs $(2, 6), (4, 8)$ and $(5, 11)$ form a 3-crossing in the left diagram.

Biophysical structures do not exhibit any isolated bonds. That is, any arc in their diagram representation is contained in a stack of length at least two. We call a diagram, whose arcs are contained in stacks of lengths at least σ, σ-modular. Modular, k-noncrossing diagrams are likely candidates for natural molecular structures. Sequence lengths of interest for such structures range from 75–300 nucleotides.

The main result of this paper is an algorithm that generates k-noncrossing, σ-modular diagrams with uniform probability. Our construction is motivated by the ideas of [5], where a combinatorial algorithm has been presented that uniformly generates k-noncrossing diagrams in $O(n^k)$ time complexity. To be precise, we generate k-noncrossing modular diagrams "locally" having a success rate that depends on specific parameters, see Fig. 2.

The paper is organized in two sections. In Section 2 we lay the foundations for our main result by generating core diagrams with uniform probability. In Section 3 we introduce weighted cores and subsequently prove the main theorem.

2 Core diagrams

A shape λ is a set of squares arranged in left-justified rows with weakly decreasing number of boxes in each row. A Young tableau is a filling in squares in the shape with numbers, which is weakly increasing in each row and strictly increasing in each column.

An oscillating tableaux [1] is a sequence of shapes $\varnothing = \lambda^0, \lambda^1, \ldots, \lambda^n$ where λ^i, $0 < i \leq n$, is obtained from λ^{i-1} by adding or removing exactly one square. While

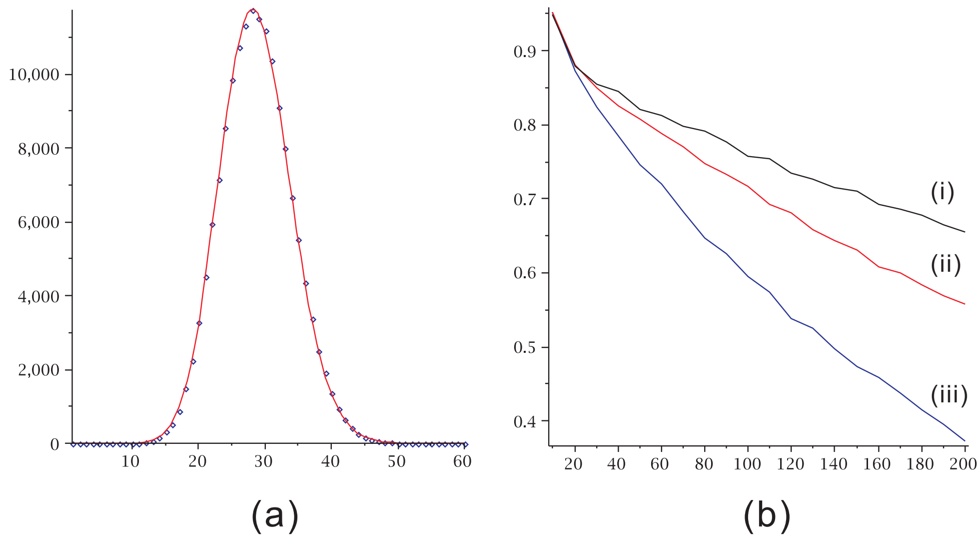

Figure 2: Uniformity and success-rate of Algorithm 2. We run Algorithm 2 for 5×10^6 times attempting to generate 3-noncrossing 2-modular diagrams over 20 vertices. $4,354,410$ of these executions generate a modular diagram. In (a) we display the frequency distribution of multiplicities (dots) and the Binomial distribution (curve). In (b) we display the success rate of Algorithm 2 as a function of n for the following classes of modular diagrams: $k = 3$, $\sigma = 2$ (i), $k = 4$, $\sigma = 2$ (ii) and $k = 5$, $\sigma = 2$ (iii).

oscillating tableaux first appeared (though not with that name) in [1], a bijection between oscillating tableaux of empty shape and matchings, i.e. diagrams without isolated points, was discovered by Stanley [2, Chapter 5] and extended by Sundaram [9] in the context of the Cauchy identity for the symplectic group. In the literature we also find the equivalent notion of "up-down tableaux" [9]. Similar concepts are vacillating tableaux [2] and generalized vacillating tableaux [3]. These tableaux sequences are in bijection with partitions [2, Theorem 5] and tangled diagrams [3, Theorem 3.6] and [4], respectively.

We introduce next the notion of $*$-tableau of shape λ^n, following [5]. A $*$-tableau is a sequence of shapes,

$$\varnothing = \lambda^0, \lambda^1, \ldots, \lambda^n,$$

such that λ^i differs from λ^{i-1} by *at most* one square, thus allowing for $\varnothing$-steps, see Fig. 3 (a).

Let us next consider the bijection between oscillating-tableaux of empty shape and diagrams without isolated points, which can be directly generalized to $*$-tableau. It is based on the Robinson-Schensted-Knuth (RSK) algorithm [7]: reading the $*$-tableau having n steps from left to right we do the following: if $\lambda^i \setminus \lambda^{i-1} = +\square$, we insert i in the new square. Otherwise if $\lambda^i \setminus \lambda^{i-1} = -\square$, we extract the unique entry j via an inverse RSK algorithm [2, 5, 3] and form an arc (j, i). By inverse RSK algorithm we mean the following: given a Young tableau Y^i of shape λ^i and a shape λ^{i+1} such that $\lambda^{i+1} \setminus \lambda^i = -\square$, there exists a unique entry j of Y^i and a Young tableau Y^{i+1} of shape λ^{i+1} such that RSK-insertion of j into Y^{i+1} recovers Y^i. Finally, in case of $\lambda^i \setminus \lambda^{i-1} = \varnothing$ we do nothing, see Fig. 3. Given a k-noncrossing diagram, we read the vertices from right to left and initialize $\lambda^n = \varnothing$. If i is a terminal of an arc, (j, i), we obtain λ^{i-1} by inserting j into λ^i via RSK insertion. If i is an isolated vertex we do nothing, and remove the square containing i when it is an origin of an arc, see Fig. 3.

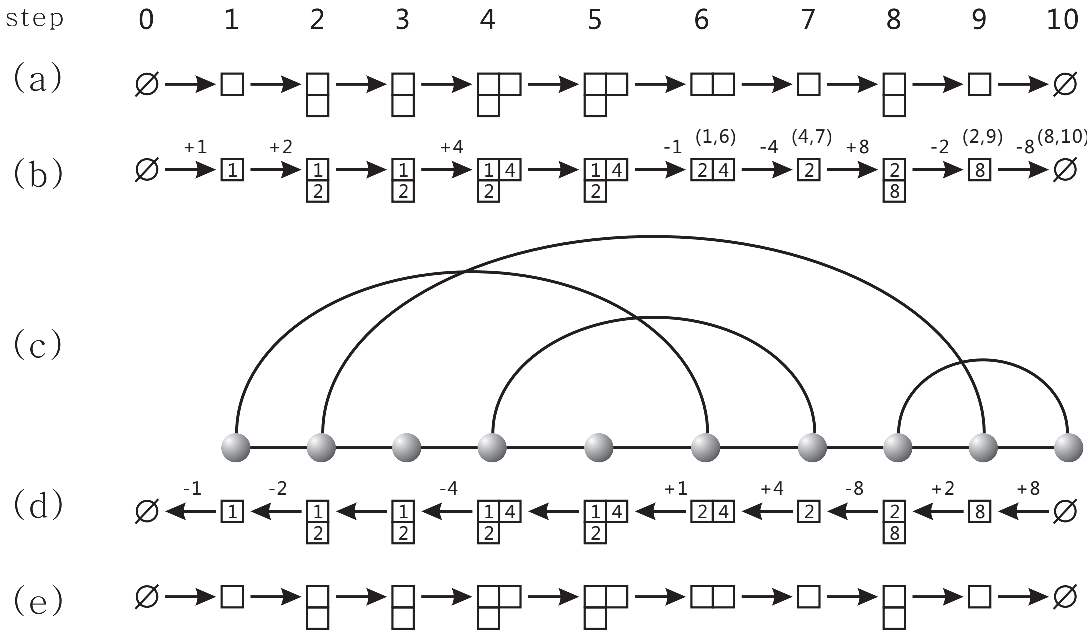

Figure 3: From $*$-tableau to diagrams and back. Reading (a) from left to right, we insert i into the new square in case of $\lambda^i \setminus \lambda^{i-1}$ being a $+\square$-step and extract the square via inverse RSK if $\lambda^i \setminus \lambda^{i-1}$ is a $-\square$-step. The extraction leads to an arc. Reading (c) from right to left, λ^{i-1} is obtained by RSK insertion of j into λ^i if i is the terminal of an arc. We do nothing if i is an isolated vertex and we remove the square with entry i in case of i being an origin of an arc.

Let $k \geq 2$ be some fixed natural number and

$$T_i(\lambda) = \{(\lambda^h)_{0 \leq h \leq i} \mid (\lambda^h)_h \text{ is a } *\text{-tableau having at most } (k-1) \text{ rows and } \lambda^i = \lambda\}.$$

Any $\vartheta \in T_i(\lambda)$ induces a unique arc-set $A(\vartheta)$. We set $A_0(\vartheta) = \varnothing$ and do the following in step h $(0 < h \leq i)$

- for a $+\square$-step, we insert h into the new square, and set $A_h(\vartheta) = A_{h-1}(\vartheta)$,
- for a $\varnothing$-step, we do nothing, and $A_h(\vartheta) = A_{h-1}(\vartheta)$,
- for a $-\square$-step, we extract the unique entry, $j(h)$, of the tableau Y^{h-1} which, if RSK-inserted into Y^h, recovers Y^{h-1} and set $A_h(\vartheta) = A_{h-1}(\vartheta) \dot\cup \{(j(h), h)\}$.

Setting $A(\vartheta) = A_i(\vartheta)$ we obtain an induced arc set $A(\vartheta)$, as well as a unique sequence of Young tableaux $Y(\vartheta) = \{Y^0 = \varnothing, Y^1, \ldots, Y^i\}$, where for $h \leq i$, Y^h is a Young tableau of shape λ^h. These extractions generate a set of arcs $(j(i), i)$, which in turn uniquely determines a diagram.

According to [2, Theorem 6], the maximal number of mutually crossing arcs in the diagram equals the maximum number of rows appearing in the shapes of its corresponding $*$-tableau. In the following all tableaux are assumed to have at most $(k-1)$ rows and accordingly, any arc-sets or diagrams are always k-noncrossing, see eq. (1.1). From now on in this paper, we fix $k \geq 2$.

Lemma 1. *Suppose $r \geq 1$ and $\vartheta_{p,q,r} \in T_i(\lambda)$ is a $*$-tableau such that*

$$(p, q), (p+1, q-1), \ldots, (p+r, q-r)$$

are stacked pairs of insertion-extraction steps. Let $f(\vartheta_{p,q,r}) \in T_i(\lambda)$ be the $*$-tableau in which all r insertion-extraction pairs $(p+1, q-1), \ldots, (p+r, q-r)$ are replaced by $2r$ $\varnothing$-steps. Then we have a correspondence between $\vartheta_{p,q,r}$ and $f(\vartheta_{p,q,r})$.

This lemma projects stacked insertion-extraction steps into a unique insertion-extraction pair and a natural number. Accordingly, it deals with many boxes of the $*$-tableaux reminiscent of the combinatorial framework of Gessel [6], where generalized paths on the Young's lattice, induced by adding or removing horizontal or vertical strips were investigated. While the latter strips [6, Chapter 4] naturally arise in the context of Pieri's rule for symmetric functions, our construction is more related to that of weights of arcs, arising in the context of ideal triangulations of marked Riemannian surfaces [13]

Proof. Let $Y(\vartheta_{p,q,r})$ denote its associated sequence of Young tableaux,

$$(Y^t)_{0 \leq t \leq i} = (Y^0 = \varnothing, Y^1, \ldots, Y^i). \tag{2.1}$$

We next construct a new sequence of Young tableaux,

$$Y(f(\vartheta_{p,q})) = \{J^0, J^1, \ldots, J^n = Y^i\}, \tag{2.2}$$

from right to left via the following algorithm

- for a $-\square$-step of the original $*$-tableau, $\vartheta_{p,q,r}$, let j be the unique entry extracted from Y^{t-1} which if RSK-inserted into Y^t recovers Y^{t-1}. If $t = q, q-1, \ldots, q-r$ we do nothing, otherwise: J^{t-1} is obtained by RSK-insertion of j into J^t,
- for a $\varnothing$-step, we do nothing,
- for a $+\square$-step, if $t = p+1, \ldots, p+r$, we do nothing, otherwise J^{t-1} is obtained by removing the square with entry t from J^t.

By construction, $J^0 = \varnothing$ and considering the induced sequence of shapes of the sequence of Young tableaux $J^0, \ldots, J^i$ we obtain a unique $*$-tableau $f(\vartheta_{p,q,r})$. By construction $f(\vartheta_{p,q,r})$ has $\varnothing$-steps at step $p+1, \ldots, p+r$ and steps $q-1, \ldots, q-r$, respectively.

Suppose we are given a $*$-tableau $\psi_{p,q,r}$ having the insertion-extraction pair (p, q) and $\varnothing$-steps at step $p+1, \ldots, p+r$ and $q-1, \ldots, q-r$, respectively together with its sequence of Young tableaux $(J^t)_{0 \leq t \leq i}$. Then we construct the sequence of Young tableaux $(Y^t)_{0 \leq t \leq i}$ initialized $Y^0 = J^0 = \varnothing$:

- for a $-\square$-step of the original $*$-tableau, $\psi_{p,q,r}$, let j be the unique entry extracted from Y^{t-1} which if RSK-inserted into Y^t recovers Y^{t-1}. Y^{t-1} is obtained by RSK-insertion of j into Y^t,

- for a $\varnothing$-step of $\psi_{p,q,r}$, if $t = q-1, \ldots, q-r$, we add a square and insert $p+1, \ldots, p+r$. If $t = p+1, \ldots, p+r$, we remove the square with the respective entry $p+1, \ldots, p+r$. Otherwise, we do nothing.

- for a $+\square$-step of $\psi_{p,q,r}$, Y^{t-1} is obtained by removing the square with entry t.

It is straightforward to verify that the above algorithm is well-defined and recovers the $*$-tableau $\vartheta_{p,q,r}$ from $f(\vartheta_{p,q,r})$, whence the lemma. See Fig. 4. $\square$

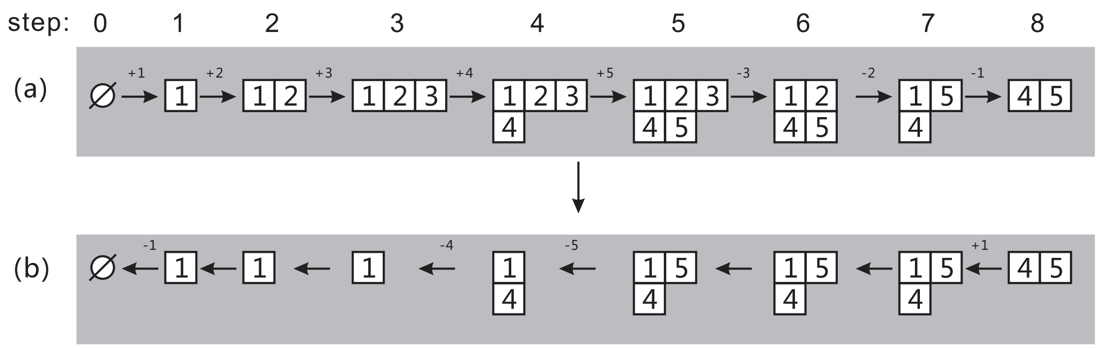

Figure 4: (a) a $*$-tableau $\vartheta_{1,8,2}$ in which $(1,8),(2,7)$ and $(3,6)$ are stacked pairs of insertion-extraction steps. (b) $f(\vartheta_{1,8,2})$ is the unique $*$-tableau derived from $\vartheta_{1,8,2}$ in which steps 2, 3, 6, and 7 are $\varnothing$-steps.

We next consider

$$T_i^c(\lambda) = \{t \in T_i(\lambda) \mid \forall a \in A(t), a \text{ is an isolated arc}\} \tag{2.3}$$

and set $t_i^c(\lambda) = |T_i^c(\lambda)|$. Given a shape λ^i, let $\lambda_{j^+}^{i-1}$ denote the shape from which λ^i is obtained by adding a square in the jth row, and $\lambda_{j^-}^{i-1}$ denote the shape from which λ^i is derived by removing a square in the jth row. Thus tracing back a shape λ^i we observe that it is either derived by

- $\lambda_{j^+}^{i-1}$ (λ^i is obtained by adding a square in the jth row of this),

- λ_0^{i-1} (λ^i is obtained by doing nothing on λ^{i-1}), or

- $\lambda_{j^-}^{i-1}$ (λ^i is obtained by removing a square in the jth row of this).

Lemma 2.

$$t_i^c(\lambda^i) = t_{i-1}^c(\lambda_0^{i-1}) + \sum_{j=1}^{k-1} t_{i-1}^c(\lambda_{j^+}^{i-1}) + \sum_{j=1}^{k-1} \sum_{p=0}^{\lfloor \frac{i-1}{2} \rfloor} (-1)^p t_{i-1-2p}^c(\lambda_{j^-}^{i-1-2p}). \tag{2.4}$$

Proof. By construction, $+\Box$-steps as well as $\varnothing$-steps do not induce new arcs. An arc α is only formed when removing a square and such an arc is potentially stacking. Let

$$G_{i-1}(\lambda_{j^-}^{i-1}) = \{(\lambda^h)_{0\le h\le i-1} \in T_{i-1}^c(\lambda_{j^-}^{i-1}) \mid \lambda^i \setminus \lambda^{i-1} = -\Box_j \text{ and } \alpha \text{ is stacking}\}.$$

Thus, for any $t \in T_{i-1}^c(\lambda_{j^-}^{i-1}) \setminus G_{i-1}(\lambda_{j^-}^{i-1})$, the $*$-tableau $(t, -\Box_j)$ is contained in $T_i^c(\lambda^i)$. We accordingly arrive at

$$T_i^c(\lambda) = T_{i-1}^c(\lambda_0^{i-1}) \dot\cup \left(\bigcup_{j=1}^{k-1} T_{i-1}^c(\lambda_{j^+}^{i-1})\right) \dot\cup \left(\bigcup_{j=1}^{k-1} [T_{i-1}^c(\lambda_{j^-}^{i-1}) \setminus G_{i-1}(\lambda_{j^-}^{i-1})]\right) \tag{2.5}$$

which implies

$$t_i^c(\lambda^i) = t_{i-1}^c(\lambda_0^{i-1}) + \sum_{j=1}^{k-1} t_{i-1}^c(\lambda_{j^+}^{i-1}) + \sum_{j=1}^{k-1} \left[t_{i-1}^c(\lambda_{j^-}^{i-1}) - g_{i-1}(\lambda_{j^-}^{i-1})\right]. \tag{2.6}$$

We next provide an interpretation of $G_{i-1}(\lambda_{j^-}^{i-1})$. Suppose the entry extracted at step i is $j(i)$. The fact that α is in a stack implies that the $(i-1)$th step is also a $-\Box$ step and that the extracted entry is $j(i)+1$. For $\vartheta \in G_{i-1}(\lambda_{j^-}^{i-1})$, we apply Lemma 1 and replace the insertion of step $j(i)+1$ and the extraction at step $(i-1)$ by respective $\varnothing$-steps, and thereby obtain the $*$-tableau $f(\vartheta)$. We then remove the two $\varnothing$-steps and obtain the unique $*$-tableau

$$\vartheta' \in T_{i-3}^c(\lambda_{j^-}^{i-3}),$$

where λ^i can be derived from $\lambda_{j^-}^{i-3}$ by removing a square in the jth row. We next claim $\vartheta' \in T_{i-3}^c(\lambda_{j^-}^{i-3}) \setminus G_{i-3}(\lambda_{j^-}^{i-3})$. Suppose $\vartheta' \in G_{i-3}(\lambda_{j^-}^{i-3})$, then ϑ contains a stack of length three, implying $\vartheta \notin G_{i-1}(\lambda_{j^-}^{i-1})$, which is impossible. Therefore, we have the bijection

$$\beta\colon G_{i-1}(\lambda_{j^-}^{i-1}) \longrightarrow T_{i-3}^c(\lambda_{j^-}^{i-3}) \setminus G_{i-3}(\lambda_{j^-}^{i-3}), \tag{2.7}$$

from which we conclude

$$g_{i-1}(\lambda_{j^-}^{i-1}) = t_{i-3}^c(\lambda_{j^-}^{i-3}) - g_{i-3}(\lambda_{j^-}^{i-3}).$$

Replacing the term $g_r(\lambda_{j^-}^r)$ and using the fact that for any shape μ, $g_1(\mu) = g_0(\mu) = 0$ holds, we arrive at

$$g_{i-1}(\lambda_{j^-}^{i-1}) = \sum_{p=1}^{\lfloor \frac{i-1}{2} \rfloor} (-1)^{p-1} t_{i-2p-1}^c(\lambda_{j^-}^{i-2p-1}).$$

This allows us to rewrite eq. (2.6) as

$$t_i^c(\lambda^i) = t_{i-1}^c(\lambda_0^{i-1}) + \sum_{j=1}^{k-1} t_{i-1}^c(\lambda_{j^+}^{i-1}) + \sum_{j=1}^{k-1} \sum_{p=0}^{\lfloor \frac{i-1}{2} \rfloor} (-1)^p t_{i-1-2p}^c(\lambda_{j^-}^{i-1-2p})$$

and the proof of the lemma is complete. $\square$

Lemma 2 allows us to compute the terms $t_i^c(\lambda)$ for arbitrary i and λ recursively via the terms $t_h^c(\lambda')$, where $h < i$ and the shapes λ' differ from λ by at most one square.

We next generate a $*$-tableau $\vartheta \in T_n^c(\lambda^n = \varnothing)$ from right to left. For this purpose we set $\mu^i = \lambda^{n-i}$ for all $0 \leq i \leq n$ and initialize $\mu^0 = \varnothing$. Suppose we have at step i the shape μ^i and consider the $T_{n-i}^c(\lambda^{n-i})$-paths starting from $\lambda^0 = \varnothing$ and ending at $\lambda^{n-i} = \mu^i$.

Corollary 1. *The transition probabilities*

$$\mathbb{P}(X^{i+1} = \mu^{i+1} \mid X^i = \mu^i) = \begin{cases} \frac{t_{n-i-1}^c(\mu^{i+1})}{t_{n-i}^c(\mu^i)} & \mu^i \setminus \mu^{i+1} = +\square_j, \varnothing \\ \frac{\sum_{p=0}^{\lfloor (n-i-1)/2 \rfloor} (-1)^p t_{n-i-2p-1}^c(\mu^{i+1})}{t_{n-i}^c(\mu^i)} & \mu^i \setminus \mu^{i+1} = -\square_j, \end{cases} \tag{2.8}$$

where $1 \leq j \leq k-1$, *induce a locally uniform Markov-process* $(X^i)_i$ *whose sampling paths are shape-sequences* $(\mu^i)_i$.

Let $\mathbf{Rand}(\mu^i)$ denote the random process of locally uniformly choosing $X^{i+1} = \mu^{i+1}$ for given $X^i = \mu^i$ using the transition probabilities given in eq. (2.8). Corollary 1 gives rise to the following algorithm:

Algorithm 1 $\mathbf{Core}(n, k)$

$m \leftarrow 0$
while $m < n$ **do**
  $\mu^{m+1} \leftarrow \mathbf{Rand}(\mu^m)$
  if $\mu^{m+1} \setminus \mu^m = +\square$ **then**
    insert $(m+1)$ in the new square
  else if $\mu^{m+1} \setminus \mu^m = -\square$ **then**
    let *pop* be the unique extracted entry of T^m which if RSK-inserted into T^{m+1} recovers T^m
    create an arc $(pop, m+1)$
    if $(pop, m+1)$ is stacking with *lastpair* **then**
      restart the process $\mathbf{Core}(n, k)$
    else
      put $(pop, m+1)$ in the arc set A
      lastpair $\leftarrow (pop, m+1)$
    end if
  end if
  $m \leftarrow m+1$
end while

The key observation now is that any core-diagram generated via the above Markov process has *uniform* probability.

Theorem 1. *Any core-diagram generated via the Markov-process* $(X^i)_i$ *(by means of the algorithm* $\mathbf{Rand}(\mu^i)$*) is generated with uniform probability.*

Proof. Suppose we are given a sequence of shapes

$$\mu^i, \mu^{i-1}, \dots, \mu^0 = \varnothing$$

Let $U_{n-i}(\mu^i)$ denote the subset of $*$-tableaux

$$\varnothing = \lambda^0, \lambda^1, \dots, \lambda^{n-i} = \mu^i$$

such that there is no stack in the induced arc set of

$$(\lambda^0, \dots, \lambda^{n-i-1}, \lambda^{n-i} = \mu^i, \mu^{i-1}, \dots, \mu^0 = \varnothing).$$

In particular, $U_n(\varnothing)$ denotes the set of all $*$-tableaux of shape $\varnothing$ having at most $(k-1)$ rows that generate only core-diagrams. Let $u_n(\varnothing) = |U_n(\varnothing)|$ denote the number of cores of length n. By construction, we have

$$U_{n-i}(\mu^i) \subseteq T^c_{n-i}(\mu^i),$$

We now condition the process $(X^i)_i$, whose transition probabilities are given by eq. (2.8), on generating cores. That is, we consider only those $*$-tableaux generated by $(X^i)_i$ that are contained in $U_n(\varnothing)$. Let this process be denoted by $(Z^i)_i$. We observe

$$\begin{aligned}
(T^c_{n-i-1}(\mu^{i+1}) \setminus G_{n-i-1}(\mu^{i+1})) \cap U_{n-i-1}(\mu^{i+1}) &= U_{n-i-1}(\mu^{i+1}) \\
T^c_{n-i}(\mu^i) \cap U_{n-i}(\mu^i) &= U_{n-i}(\mu^i) \\
T^c_{n-i-1}(\mu^{i+1}) \cap U_{n-i-1}(\mu^{i+1}) &= U_{n-i-1}(\mu^{i+1}).
\end{aligned}$$

Accordingly, using eq. (2.8), we derive for the transition probabilities

$$\mathbb{P}(Z^{i+1} \mid Z^i) = \frac{|U_{n-i-1}(\mu^{i+1})|}{|U_{n-i}(\mu^i)|}.$$

Therefore we arrive at

$$\mathbb{P}(Z^{i+1}) = \prod_{p=0}^{i} \frac{|U_{n-i-1+p}(\mu^{i+1-p})|}{|U_{n-i+p}(\mu^{i-p})|} = \frac{|U_{n-i-1}(\mu^{i+1})|}{|U_n(\mu^0 = \varnothing)|} = \frac{|U_{n-i-1}(\mu^{i+1})|}{u_n(\varnothing)}$$

and in particular

$$\mathbb{P}(Z^n = \varnothing) = \frac{|U_0(\mu^n = \varnothing)|}{|U_n(\mu^0 = \varnothing)|} = \frac{1}{u_n(\varnothing)},$$

which implies that the process $(Z^i)_i$ generates cores with uniform probability. □

3 Modular diagrams

Any σ-modular diagram can be mapped into a σ-weighted core, i.e. a diagram whose arcs have additional weights $\geq \sigma$. Suppose we have a $*$-tableau of $\varnothing$, ϑ, whose induced diagram is a σ-modular diagram. Repeated application of Lemma 1 for each respective stack

$$S = ((p,q),(p+1,q-1),\dots,(p+(s-1),q-(s-1))),$$

allows us to replace any insertion-step $p+1,\dots,p+(s-1)$ as well as any extraction-step $q-(s-1),\dots,q-1$ by $\varnothing$-steps, respectively. Removing the $2(s-1)$ $\varnothing$-steps and assigning the stack-lengths s to the *extraction* in step q, generates a $*$-tableau of $\varnothing$ with weights, θ (σ-weighted $*$-tableau).

Using the correspondence between $*$-tableau and diagrams, a σ-weighted core can therefore be represented as a sequence of shapes, θ in which, preceding each extraction step, we have the additional insertion of exactly $2(s-1)$ $\varnothing$-steps, see Fig. 5. Let $W_i^\sigma(\lambda^r)$

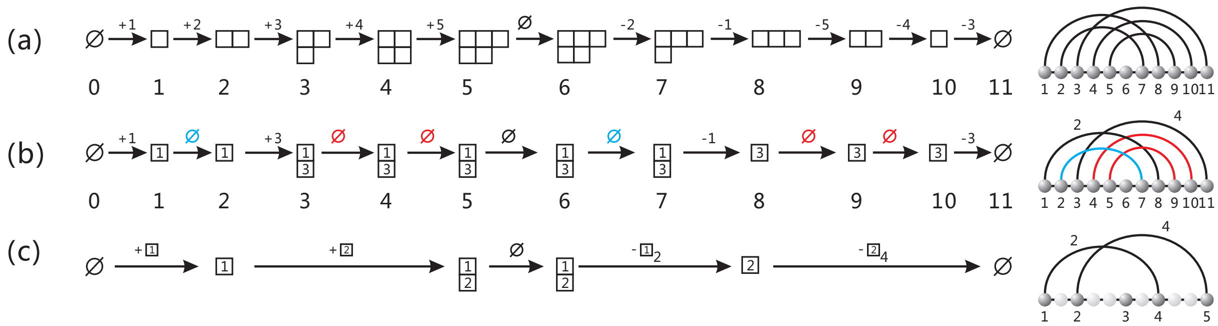

Figure 5: (a) a $*$-tableau whose induced diagram is a 2-modular diagram. (b) the $*$-tableau obtained by repeated application of Lemma 1. The red and blue removed arcs correspond the red and blue $\varnothing$-steps in the $*$-tableau, respectively. (c) the weighted $*$-tableau induced by (b) with weights 2 and 4 assigned to the two extraction steps, respectively, and its induced weighted core.

denote the set of σ-weighted $*$-tableau. Each such $\theta \in W_i^\sigma(\lambda^r)$ induces a unique $*$-tableau, $p(\theta)$, contained in $T_r^c(\lambda^r)$ and we have

$$i = r + \sum_{\ell=1}^{h\leq r/2} 2(s_\ell - 1),$$

where s_ℓ is the weight of the ℓth extraction in θ. We set $w_i^\sigma(\lambda^r) = |W_i^\sigma(\lambda^r)|$.

Lemma 3. *We have the recursion formula*

$$\begin{aligned} w_i^\sigma(\lambda^r) &= w_{i-1}^\sigma(\lambda_0^{r-1}) + \sum_{j=1}^{k-1} w_{i-1}^\sigma(\lambda_{j^+}^{r-1}) \qquad (3.1)\\ &+ \sum_{j=1}^{k-1}\sum_{s=\sigma}^{\lfloor\frac{i+1}{2}\rfloor}\sum_{\ell=1}^{\lfloor\frac{s}{\sigma}\rfloor}(-1)^{\ell-1}p(s,\ell,\sigma)w_{i-2s+1}^\sigma(\lambda_{j^-}^{r-1}), \end{aligned}$$

where $p(a,\ell,\sigma)$ denotes the number of partitions of a into ℓ blocks, $\{a_1,a_2,\dots,a_\ell\}$, such that $\forall i\leq \ell$, $a_i\geq\sigma$.

Proof. Any $*$-tableau $\theta \in W_i^\sigma(\lambda^r)$, where $i = r + \sum_{\ell=1}^{h} 2(s_\ell - 1)$, s_ℓ is the weight assigned to the ℓth extraction step in θ. We consider the weighted $*$-tableau, θ', derived from θ by removing the shape in step r. If λ^r is derived from λ^{r-1} by doing nothing, then $\theta' \in W_{i-1}^\sigma(\lambda_0^{r-1})$. Similarly, if λ^r is derived from λ^{r-1} by adding a square in the jth row, we have $\theta' \in W_{i-1}^\sigma(\lambda_{j^+}^{r-1})$. In case of λ^r being derived from λ^{r-1} via removing a square from the jth row, we are given an extraction step with associated weight s. Thus,

$$\theta' \in W^\sigma_{r-1+\sum_{\ell=1}^{h-1} 2(s_\ell-1)}(\lambda_{j^-}^{r-1}) = W^\sigma_{(i-1)-2(s-1)}(\lambda_{j^-}^{r-1}).$$

θ' determines $*$-tableau, $p(\theta') \in T_{r-1}^c(\lambda_{j^-}^{r-1}) \setminus G_{r-1}(\lambda_{j^-}^{r-1})$. Let $V^\sigma_{(i-1)-2(s-1)}(\lambda_{j^-}^{r-1})$ denote the set of weighted $*$-tableau θ_1 such that $p(\theta_1) \in G_{r-1}(\lambda_{j^-}^{r-1})$. We set $v^\sigma_{(i-1)-2(s-1)}(\lambda_{j^-}^{r-1}) = |V^\sigma_{(i-1)-2(s-1)}(\lambda_{j^-}^{r-1})|$. Note that then $\theta' \in W^\sigma_{(i-1)-2(s-1)}(\lambda_{j^-}^{r-1}) \setminus V^\sigma_{(i-1)-2(s-1)}(\lambda_{j^-}^{r-1})$, whence

$$\begin{aligned} W_i^\sigma(\lambda^r) = {} & W_{i-1}^\sigma(\lambda_0^{r-1}) \,\dot\cup \left(\bigcup_{j=1}^{k-1} W_{i-1}^\sigma(\lambda_{j^+}^{r-1})\right) \\ & \dot\cup \left(\bigcup_{j=1}^{k-1} \bigcup_{s=\sigma}^{\lfloor \frac{i+1}{2} \rfloor} [W^\sigma_{(i-1)-2(s-1)}(\lambda_{j^-}^{r-1}) \setminus V^\sigma_{(i-1)-2(s-1)}(\lambda_{j^-}^{r-1})]\right). \end{aligned}$$

We therefore derive

$$w_i^\sigma(\lambda^r) = w_{i-1}^\sigma(\lambda_0^{r-1}) + \sum_{j=1}^{k-1} w_{i-1}^\sigma(\lambda_{j^+}^{r-1}) + \sum_{j=1}^{k-1} \sum_{s=\sigma}^{\lfloor \frac{i+1}{2} \rfloor} [w^\sigma_{i-2s+1}(\lambda_{j^-}^{r-1}) - v^\sigma_{i-2s+1}(\lambda_{j^-}^{r-1})]. \quad (3.2)$$

We proceed by considering a $*$-tableau $\zeta \in V^\sigma_{(i-1)-2(s-1)}(\lambda_{j^-}^{r-1})$. By construction, $(r-1)$ is a $-\Box$-step. Suppose the induced arc of this extraction is α and the weight assigned to it is given by s'. Then $p(\zeta) \in G_{r-1}(\lambda_{j^-}^{r-1})$ and we have the bijection

$$\beta \colon G_{r-1}(\lambda_{j^-}^{r-1}) \longrightarrow T_{r-3}^c(\lambda_{j^-}^{r-3}) \setminus G_{r-3}(\lambda_{j^-}^{r-3}),$$

obtained by removing the insertion and extraction step of the extracted square in step $(r-1)$. Taking into the account weights, β gives rise to the bijection

$$\beta' \colon V^\sigma_{(i-1)-2(s-1)}(\lambda_{j^-}^{r-1}) \longrightarrow \bigcup_{s'=\sigma}^{\lfloor \frac{i-2s+1}{2} \rfloor} [W^\sigma_{(i-1)-2(s+s'-1)}(\lambda_{j^-}^{r-3}) \setminus V^\sigma_{(i-1)-2(s+s'-1)}(\lambda_{j^-}^{r-3})],$$

from which we conclude

$$v^\sigma_{(i-1)-2(s-1)}(\lambda_{j^-}^{r-1}) = \sum_{s'=\sigma}^{\lfloor \frac{i-2s+1}{2} \rfloor} [w^\sigma_{(i-1)-2(s+s'-1)}(\lambda_{j^-}^{r-3}) - v^\sigma_{(i-1)-2(s+s'-1)}(\lambda_{j^-}^{r-3})].$$

Using (a)

$$\sum_{s_1=\sigma} \cdots \sum_{s_\ell=\sigma} x_{s_1+\cdots+s_\ell} = p(s,\ell,\sigma) x_s$$

where $p(s,\ell,\sigma)$ denotes the number of partitions of s into ℓ blocks of size $\geq \sigma$, and (b) that for any shape μ, $v_1^\sigma(\mu) = v_0^\sigma(\mu) = 0$. We iterate the above formula by replacing the terms $v_r^\sigma(\lambda_{j^-}^r)$

$$\begin{aligned}
&\sum_{s=\sigma}^{\lfloor\frac{i+1}{2}\rfloor} (w_{i-2s+1}^\sigma(\lambda_{j^-}^{r-1}) - v_{i-2s+1}^\sigma(\lambda_{j^-}^{r-1}))\\
&= \sum_{s=\sigma}^{\lfloor\frac{i+1}{2}\rfloor} w_{i-2s+1}^\sigma(\lambda_{j^-}^{r-1}) - \sum_{s=\sigma}^{\lfloor\frac{i+1}{2}\rfloor}\sum_{s'=\sigma}^{\lfloor\frac{i-2s+1}{2}\rfloor} (w_{i-2s-2s'+1}^\sigma(\lambda_{j^-}^{r-3}) - v_{i-2s-2s'+1}^\sigma(\lambda_{j^-}^{r-3}))\\
&\vdots\\
&= \sum_{s_1=\sigma} \cdots \sum_{s_\ell=\sigma}^{2(s_1+\cdots+s_\ell)\leq i+1} (-1)^\ell w_{i-2(s_1+\cdots+s_\ell)+1}^\sigma(\lambda_{j^-}^{r-2\ell+1})\\
&= \sum_{s=\sigma}^{\lfloor\frac{i+1}{2}\rfloor}\sum_{\ell=1}^{\lfloor\frac{s}{\sigma}\rfloor}(-1)^{\ell-1}p(s,\ell,\sigma)w_{i-2s+1}^\sigma(\lambda_{j^-}^{r-1}),
\end{aligned}$$

whence the lemma. □

Lemma 3 allows us to compute $w_i^\sigma(\mu)$ for arbitrary i, μ inductively via the terms $w_h^\sigma(\lambda)$ and $h < i$. We next consider the generation of a $*$-tableau, ϑ, which corresponds to a σ-modular diagram. For this purpose we shall generate a weighted $*$-tableau $\theta \in W_n^\sigma(\lambda^m = \varnothing)$. Taking the sum over all weights we have $m = n - \sum_h 2(s_h - 1)$. We construct θ inductively from right to left setting $\mu^r = \lambda^{m-r}$. We initialize $\mu^0 = \varnothing$ and assign in case of $\mu^r \setminus \mu^{r+1} = -\square$ a weight to step r. Suppose we have arrived at μ^r, with the corresponding set of weights, S^r. Considering sequences of weighted $*$-tableaux contained in $W^\sigma_{n-r-\sum_{s_\ell\in S^r} 2(s_\ell-1)}(\mu^r)$, Lemma 3 implies

Corollary 2. *Let*

$$\begin{aligned}
t &= n - r - \sum_{s_\ell \in S^r} 2(s_\ell - 1)\\
z_{t-1}^\sigma(\mu_{j^-}^{r+1}, s) &= \sum_{\ell=1}^{\lfloor\frac{s}{\sigma}\rfloor}(-1)^{\ell-1}p(s,\ell,\sigma)w_{t-2s+1}^\sigma(\mu_{j^-}^{r+1}).
\end{aligned}$$

The transition probabilities

$$\begin{aligned}
&\mathbb{P}(X^{r+1} = (\mu^{r+1}, S^{r+1}) \mid X^r = (\mu^r, S^r))\\
&= \begin{cases}
\frac{w_{t-1}^\sigma(\mu^{r+1})}{w_t^\sigma(\mu^r)} & \mu^r \setminus \mu^{r+1} = +\square_j, S^{r+1} = S^r\\
\frac{w_{t-1}^\sigma(\mu^{r+1})}{w_t^\sigma(\mu^r)} & \mu^r \setminus \mu^{r+1} = \varnothing, S^{r+1} = S^r\\
\frac{z_{t-1}^\sigma(\mu^{r+1},s)}{w_t^\sigma(\mu^r)} & \mu^r \setminus \mu^{r+1} = -\square_j, S^{r+1} = S^r \cup \{s\},
\end{cases}
\end{aligned} \tag{3.3}$$

for $1 \leq j \leq k-1$, generate a locally uniform Markov-process $(X^i)_i$.

Corollary 2 represents an algorithm for constructing σ-modular diagrams. In analogy to the case of core-diagrams, if X successfully constructs a modular diagram, it generates the latter with uniform probability, see Fig. 2 (lefthand side). Consequently, the process

Algorithm 2 Canonical(n, k, σ)

$m \leftarrow 0$
while $m < n$ **do**
  $(\mu^{m+1}, size) \leftarrow$ **RandStep**(μ^m)
  if $\mu^{m+1} \setminus \mu^m = +\square$ **then**
    insert $(m+1)$ in the new square
    assign $size$ to the the new square
    $m \leftarrow m + size - 1$
  else if $\mu^{m+1} \setminus \mu^m = -\square$ **then**
    let pop be the unique extracted entry of T^m which if RSK-inserted into T^{m+1}, recovers T^m and let $size$ be the integer assigned to the extracted square
    create a stack $\{(pop, m + size), \cdots, (pop + size - 1, m + 1)\}$
    if $(pop, m + size)$ is stacking with $lastpair$ **then**
      restart the process **Canonical**(n, k, σ)
    else
      put $\{(pop, m + size), \cdots, (pop + size - 1, m + 1)\}$ in the arc set A
      $lastpair \leftarrow (pop, m + size)$
      $m \leftarrow m + size - 1$
    end if
  end if
  $m \leftarrow m + 1$
end while

$(X^i)_n$ generates random σ-modular, k-noncrossing diagram in $O(n)$ time and space complexity. According to the recursion of Lemma 3, we compute $w_i^\sigma(\lambda^i)$ for arbitrary λ^i with at most $(k-1)$ rows and all $i \leq n$ in $O(n) \times O(n^{k-1}) = O(n^k)$ time and space complexity.

Theorem 2. *Any modular diagram derived via the Markov-process $(X^r)_r$ is generated with uniform probability.*

Proof. Suppose we have a sequence of shapes

$$\mu^r, \mu^{r-1}, \ldots, \mu^0 = \varnothing$$

with weights assigned to each $\mu^{i-1} \setminus \mu^i = -\square_j$-step and set of weights, S^r. Let

$$t = n - r - \sum_{s_\ell \in S^r} 2(s_\ell - 1)$$

and $D^\sigma_{m-r}(\mu^r)$ be the set of weighted $*$-tableaux

$$\lambda^0 = \varnothing, \lambda^1, \ldots, \lambda^{m-r} = \mu^r$$

such that

$$\varnothing = \lambda^0, \dots, \lambda^{m-r-1}, \lambda^{m-r} = \mu^r, \mu^{r-1}, \dots, \mu^0 = \varnothing$$

is contained in $W_n^\sigma(\lambda^m = \varnothing)$. Summing over all weights we have $m = n - \sum_{s_h} 2(s_h - 1)$ and by construction, $D_{m-r}^\sigma(\mu^r) \subseteq W_t^\sigma(\mu^r)$. In particular

$$d_m^\sigma(\varnothing) = |D_m^\sigma(\mu^m = \varnothing)|$$

equals the number of weighted core of length m, i.e. the number of modular diagrams of length n. Suppose now we only consider sampling paths of weighed cores generated via $(X^r)_r$ (whose transition probabilities is given by eq. (3.3)) contained in $D_m^\sigma(\varnothing)$. We denote the resulting process by $(Z^r)_r$. In view of

$$D_{m-r}^\sigma(\mu^r) \subseteq W_t^\sigma(\mu^r),$$

we observe that

$$\begin{aligned} (W_{t-1}^\sigma(\mu^{r+1}) \setminus V_{t-1}^\sigma(\mu^{r+1})) \cap D_{m-r-1}^\sigma(\mu^{r+1}) &= D_{m-r-1}^\sigma(\mu^{r+1}) \\ W_t^\sigma(\mu^r) \cap D_{m-r}^\sigma(\mu^r) &= D_{m-r}^\sigma(\mu^r) \\ W_{t-1}^\sigma(\mu^{r+1}) \cap D_{m-r-1}^\sigma(\mu^{r+1}) &= D_{m-r-1}^\sigma(\mu^{r+1}). \end{aligned}$$

Therefore we have

$$\mathbb{P}(Z^{r+1} \mid Z^r) = \frac{|D_{m-r-1}^\sigma(\mu^{r+1})|}{|D_{m-r}^\sigma(\mu^r)|}$$

and consequently

$$\mathbb{P}(Z^{r+1}) = \prod_{p=0}^{r} \frac{|D_{m-r-1+p}^\sigma(\mu^{r+1-p})|}{|D_{m-r+p}^\sigma(\mu^{r-p})|} = \frac{|D_{m-r-1}^\sigma(\mu^{r+1})|}{|D_m^\sigma(\mu^0 = \varnothing)|}. \tag{3.4}$$

In particular,

$$\mathbb{P}(Z^m = \varnothing) = \frac{|D_0^\sigma(\mu^m = \varnothing)|}{|D_m^\sigma(\mu^0 = \varnothing)|} = \frac{1}{|D_m^\sigma(\mu^0 = \varnothing)|} = \frac{1}{d_m^\sigma(\varnothing)}.$$

That is, the process (Z^r) generates modular diagrams uniformly and the theorem follows. □

In Fig. 6, we showcase two paths constructed via the process **Canonical**(n, k, σ), for $n = 8, k = 3$ and $\sigma = 2$. In Fig. 7 we construct the corresponding 2-modular diagram from the red path displayed in Fig. 6.

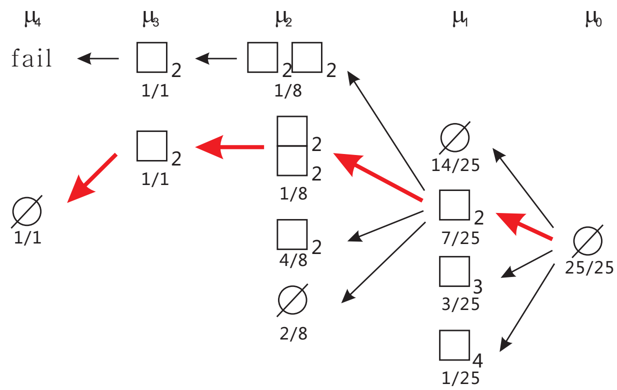

Figure 6: Building weighted $*$-tableau via the transition probabilities given in eq. (3.3). Here, the top path fails to generate a 2-modular diagram while the red path succeeds. According to Theorem 2 each such modular diagram is generated with uniform probability.

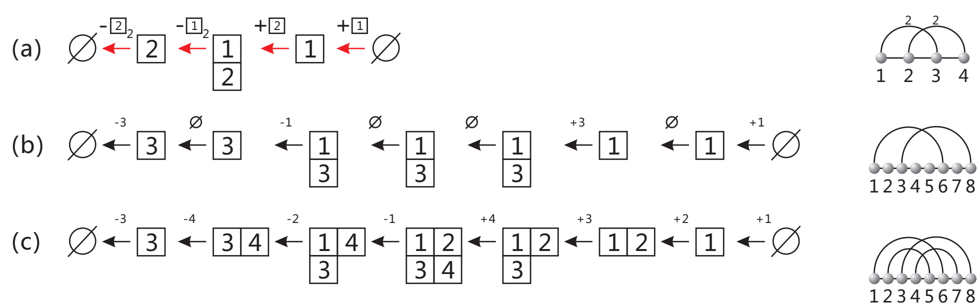

Figure 7: (a) the red path of Fig. 6. (b) the $*$-tableau derived by adding four $\varnothing$-steps in (a). (c) adding two pairs of insertion and extraction steps, which produces the 2-modular diagram.

References

[1] Berele, A., A schensted-type correspondence for the symplectic group, *J. Comb. Theor. (A)* **43** (1986), 320–328.

[2] Chen, W. Y. C., Deng, E. Y. P., Du, R. R. X., Stanley, R. P. and Yan, C. H., Crossings and nestings of matchings and partitions, *Trans. Am. Math. Soc.* **359** (2007), 1555–1575.

[3] Chen, W. Y. C., Qin, J. and Reidys, C. M., Crossings and nestings in tangled diagrams, *Elec. J. Comb.* **15** (2008), #R86.

[4] Chen, W. Y. C., Qin, J., Reidys, C. M. and Zeilberger, D., Efficient counting and asymptotocs of k-noncrossing tangled-diagrams, *Elec. J. Comb.* **16** (2009), #37.

[5] Chen, W. Y. C., Han, H. S. W. and Reidys, C. M., Random k-noncrossing RNA Structures, *Proc. Natl. Acad. Sci, USA* **106(52)** (2009), 22061–22066.

[6] Gessel, I. M., Counting Paths in Young's Lattice, *J. Stat. Plann. Infer.* **34(1)** (1993), 125–134.

[7] Knuth, D. E.,Permutations, matrices, and generalized Young tableaux, *Pacific J. Math.* **34** (1970), 709–727.

[8] Hofacker, I. L., Schuster, P. and Stadler, P. F., Combinatorics of RNA secondary structures, *Discr. Appl. Math.* **88** (1998), 207–237.

[9] Sundaram, S., The Cauchy identity for $Sp(2n)$, *J. Comb. Theor. (A)* **53** (1990), 209–238.

[10] Jin, E. Y., Qin, J. and Reidys, C. M., Combinatorics of RNA structures with pseudoknots, *Bull. Math. Biol.* **70(1)** (2008), 45–67.

[11] Jin, E. Y. and Reidys, C. M., Combinatorial design of pseudoknot RNA, *Adv. Appl. Math.* **42** (2008), 135–151.

[12] Penner, R. C. and Waterman, M. S., Spaces of RNA secondary structures, *Adv. Math.* **101** (1993), 31–49.

[13] Penner, R. C., The structure ad singularities of quotient arc complexes, *J. Topo.* **1** (2008), 527-550.

[14] Rietveld, K., Van Poelgeest, R., Pleij, C. W., Van Boom, J. H. and Bosch, L., The tRNA-like structure at the $3'$ terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA, *Nucl. Acid. Res.* **10** (1982), 1929–1946.

[15] Waterman, M. S. and Smith, T. F., RNA secondary structure: a complete mathematical analysis, *Math. Biol.* **42** (1978), 257–266.

[16] Waterman, M. S., Combinatorics of RNA hairpins and cloverleaves, *Stud. Appl. Math.* **60** (1979), 91–96.

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