A bijective method for saturated extended 2-regular simple stacks in the Nussinov-Jacobson energy model

Qianghui Guo\textsuperscript{a}, Yinglie Jin\textsuperscript{a}, Lisa H. Sun\textsuperscript{b}, Mingxing Weng\textsuperscript{a}
\textsuperscript{a}School of Mathematical Sciences, LPMC, Nankai University, Tianjin 300071, P. R. China
\textsuperscript{b}Center for Combinatorics, LPMC, Nankai University, Tianjin 300071, P. R. China
Emails: guo@nankai.edu.cn, yljin@nankai.edu.cn, sunhui@nankai.edu.cn

Abstract. In combinatorics, RNA secondary structures can be seen as noncrossing diagrams in which each arc length is at least two and the degree of each vertex is at most one, which are also called 2-regular simple stacks. In the Nussinov-Jacobson model, in which each base pair contributes energy $-1$, the saturated structures are those in which no more arc can be added without violating the structure definition. In this paper, we consider the extended 2-regular simple stacks, in which each arc length is at least two, the degree of each internal vertex is at most one, and the degree of the two terminal vertices is bounded by two. We construct a bijection between the saturated extended 2-regular simple stacks and the forests of small trees, which are rooted trees with height one and also called meadows in graph theory. Following this bijection, we obtain an explicit formula for the number of the saturated extended 2-regular simple stacks. Furthermore, the results for saturated 2-regular simple stacks due to Clote, the optimal extended 2-regular simple stacks due to Guo, Jin, Sun, and Xu, and the $k$-saturated extended 2-regular simple stacks are derived as simple consequences.

Keywords: combinatorial enumeration, RNA secondary structure, meadow, extended 2-regular simple stacks, saturated and optimal structures, explicit formulas

1 Introduction

Ribonucleic acid (RNA) is known to have an essential regulatory role in the cell, and different RNAs have different roles. The function of RNA is determined by RNA tertiary structure, which significantly depends on the secondary structure [1]. Determining the minimum free energy configuration for a given RNA sequence has thus become a widely studied problem, and the enumeration of secondary structure is a natural problem. Following Waterman [14], a secondary structure $S$ for a given RNA sequence $s = s_1 s_2 \cdots s_n$ is a set of ordered pairs $(i, j)$, such that $1 \leq i < j \leq n$ and the following satisfied: (a) If $(i, j), (i, k) \in S$, then $j = k$; if $(i, j), (k, j) \in S$, then $i = k$; (b) If $(i, j) \in S$, then $|j - i| \geq 2$; (c) If $(i, j), (k, l) \in S$, then it is not the case that $i < k < j < l$. 

1
In combinatorics, we present RNA secondary structure by a diagram, in which vertices 1, 2, ..., n are increasing ordered along the horizontal axis, and the base pairs are shown as arcs. We say any two arcs \((i, j)\) and \((k, l)\) form a nesting, if \(i < k < l < j\), and crossing if \(i < k < j < l\). A noncrossing diagram is called a stack, and a nonnesting diagram is called a queue. Following [3,10], a structure (stack or queue) with arc length at least \(m\) is called \(m\)-regular; a structure with the degree of each vertex bounded by one is called simple. Thereby, an RNA secondary structure is actually a 2-regular simple stack. Furthermore, an \(m\)-regular simple stack with the degrees of the two terminal vertices bounded by two is called an extended \(m\)-regular simple stack.

Schmitt and Waterman [14] obtained an explicit formula for the number of RNA secondary structures of length \(n\) with \(k\) base pairs by establishing a one-to-one correspondence between secondary structures and linear trees. Nebel [12] derived the generating functions of \(m\)-regular RNA secondary structures from generating functions for binary tree structures with Horton-Strahler number. Reidys systematically developed an entirely novel approach for the study of RNA secondary structures with pseudoknot in his monograph [13].

Most of the RNA prediction algorithms are based on free energy minimization [16,18]. The optimal structures are those with the lowest free energy, and the saturated structure, introduced by Zuker [17], is formally defined as the secondary structure in which no base pairs can be added without violating the definition of secondary structure. Clote et al [4,7,15] delves into the combinatorial problem related to the number of saturated RNA secondary structures. Following Zuker [17], Clote [4] defined \(k\)-saturated secondary structure as the secondary structure, which is saturated and additionally contains \(k\) fewer base pairs than the maximum possible number of base pairs, and provided recurrence relations to compute the number of \(k\)-saturated secondary structures.

The enumeration of various RNA secondary structures has been studied extensively. However, few explicit formulas have been found except for the result presented by Schmitt and Waterman [14], and most results are in the form of generating function equation, recurrence relation, and asymptotic formula [3,8,10]. Inspired by the work due to Schmitt and Waterman [14], Chen [2] and Clote [4], we construct an algorithm to map saturated extended 2-Regular simple stacks to meadows. By enumerating the meadows, we get the explicit formulas for the number of saturated extended 2-regular simple stacks.

2 The bijective algorithm

This section will present the algorithm to decompose extended 2-regular simple stacks to meadows, which will be the primary tool for the investigations presented subsequently.
First, we introduce some of the terms in graph theory about trees. The fiber of \( v \) is the set of children of \( v \). A linear tree is a rooted tree together with a linear ordering on the fiber of each inner vertex in the tree. A rooted tree of height one will be called a small tree. We adopt the term 'meadow' for a forest of small trees, which has already been used in graph theory. Since the small tree has only one inner vertex, i.e. root, the set of leaves of the small tree is called the fiber of the small tree. The inner vertex whose children are all leaves is called outmost inner vertex.

Next, we present two crucial bijections. Let \( R_{n,k} \) be the secondary structures on \([n]\) which have exactly \( k \) arcs, and let \( T_{n,k} \) be the set of unlabeled linear trees having \( n \) vertices, \( k \) of which are inner vertices.

**Lemma 2.1.** \([14]\) For all \( n, k \geq 1 \), there exists a bijection \( \varphi : R_{n+k-2,k-1} \rightarrow T_{n,k} \).

Let \( B \) be a secondary structure on \([n+k-2]\) with \( k-1 \) arcs. Denote the set of isolated vertices by \( F \). Let \( V \) be the set \( \{(i,j) : (i,j) \in B\} \cup F \cup \{*\} \). Partially order \( V \) by letting \(*\) be maximal, and otherwise ordering by set inclusion. The Hasse diagram of this poset is a rooted tree, with root \(*\), having a total of \( n \) vertices, \( k \) of which are inner vertices. The linear order of the set \( F \) of terminal vertices gives this tree a linear structure. For example, Figure 2.1 shows a secondary structure in \( R_{21,6} \) and the corresponding unlabeled linear tree in \( T_{16,7} \).

Let \( \mathcal{LT}_{n,k} \) be the set of all labeled linear trees on \( n \) \((n>1)\) vertices with \( k \) inner vertices and \( \mathcal{M}_{n,k} \) the set of meadows of \( k \) small trees with roots less than \( n-k+1 \) on \( n \) vertices. In the meadow of \( k \) small trees on \( n \) vertices, we mark all the vertices greater than \( n-k+1 \) by the symbol \(*\). Then the roots of the meadows in \( \mathcal{M}_{n,k} \) are all unstarred.

**Lemma 2.2.** For all \( n, k \geq 1 \), there is a bijection \( \psi : \mathcal{LT}_{n,k} \rightarrow \mathcal{M}_{n+k-1,k} \).

**Proof.** I first give the procedure to construct a labeled linear tree on \( n \) \((n>1)\) vertices with \( k \) inner vertices from a meadow \( M \) of \( k \) small trees on on \( \{1,2,\cdots,n+k-1\} \) with root less than \( n \).

(1) Find the tree \( T \) in \( M \) with the largest root such that there is no marked vertex in \( T \). Let \( i \) be the root of \( T \).

(2) Find the tree \( T^* \) in \( M \) that contains \( (n+1)^* \), then replace \( (n+1)^* \) with \( T \) in \( T^* \).

(3) Repeat the above procedure for vertices \( n+2,\cdots,n+k-1 \).

The following is the procedure to decompose a labeled linear tree into small trees.
Figure 2.2: A labeled linear tree and its meadow decomposition

Table 2.1: Correspondences between secondary structures, rooted trees and meadows

| secondary structure          | rooted tree          | meadow                        |
|-----------------------------|----------------------|-------------------------------|
| $k$ arcs                    | $k + 1$ inner vertices | $k + 1$ small trees          |
| $n - 2k$ isolated vertices  | $n - 2k$ leaves       | $n - 2k$ unstarred leaves     |
| $p$ hairpins                | $p$ outmost inner vertices | $p$ small trees with no starred vertices |
| arc of length $m$           | outmost inner vertex with degree $m - 1$ | small trees with $m - 1$ unstarred leaves |
| $b$ visible vertices        | $b$ leaves in the fiber of root |

(1) Find the largest inner vertex $i$ such that all children of $i$ are leaves. Then we obtain a small tree with root $i$ and leaf set fiber of $i$.

(2) Remove the fiber of $i$ and relabel the vertex $i$ by $n + 1$.

(3) Repeat the above procedure and relabel the encountered roots of trees subsequently by $n + 1, \ldots, n + k - 1$.

Figure 2.2 shows an instance for the merge and decompose algorithm in Lemma 2.2.

A hairpin loop in secondary structure $S$ is given by base pair $(i, j) \in S$, such that $i + 1, \ldots, j - 1$ are unpaired in $S$. Call a vertex $i \in [n]$ visible if it is not covered by any arc. By observing bijections $\varphi$ and $\psi$, we can easily find the corresponding relationships between secondary structures, rooted trees and meadows, see Table 2.1.

An vertex $v$ is incident with an arc $(i, j)$ if $v \in \{i, j\}$; then $(i, j)$ is an edge at $v$. For a vertices set $V$, arc $(i, j)$ is at $V$ if there exists $v \in V$ such that $v$ incident with $(i, j)$. Let $S$ be a stack and $V$ be a subset of vertices of $S$. We call the component of $S$ containing $V$ and
As shown in Table 2.2, the arc patterns with respect to \( \{1, n\} \) split \([n]\) into disjoint intervals, on which substructures can be constructed. To meet the restrictions of saturated extended 2-regular simple stack, the substructures can be classified into the following seven types:

- \( T_1 \): an arbitrary saturated simple stack;
- \( T_2 \): an arbitrary nonempty saturated simple stack;
- \( T_3 \): an arbitrary saturated simple stack with no visible vertex;
- \( T_4 \): an arbitrary nonempty saturated simple stack with no visible vertex;
- \( T_5 \): \( T_4 \) or an isolated vertex.
- \( T_6 \): \( T_3 \) followed by an isolated vertex, or just \( T_3 \);
- \( T_7 \): \( T_3 \) followed by an isolated vertex, or just \( T_4 \);

Note that, the substructures of types \( T_1, T_3 \) can be empty. \( T'_6 \) and \( T'_7 \) stand for the reverse structures of \( T_6 \) and \( T_7 \), respectively.
In order to enumerate the number of stacks showed in Table 2.2, we introduce the meadow decompose algorithm $\Phi$ for extended simple stacks. Let $S$ be an arbitrary extended simple stack, $V$ be a subset of vertices of $S$, and $A$ be the arc pattern w.r.t. $V$ of $S$. Assume the number of arcs of $A$ is $k_1$.

1. For any vertex $v \in V$, denote the degree of $v$ in $A$ by $\text{deg}_A v$. If $\text{deg}_A v = 0$, delete $v$ in $S$; if $\text{deg}_A v = 1$, do nothing; if $\text{deg}_A v = 2$, add a new vertex $u$ to the left of $v$ and bond one of the two arcs at $v$ to $u$ such that no crossing occurs;

2. Map the stack obtained in step (1) into an unlabeled linear tree with $k_1 + 1$ determined inner vertices by $\varphi$.

3. Assume the linear tree obtained by step (3) have $n$ vertices. Label the $k_1 + 1$ determined inner vertices by $[k_1 + 1]$ in the order from top to bottom and left to right, and label the remaining vertices by $[k_1 + 2, n]$ in $(n - k_1 - 1)!$ ways.

4. Decompose the labeled linear tree into meadows by $\psi$.

Let $S$ be an arbitrary saturated RNA secondary structure, and $M$ be an arbitrary meadow in $\Phi(S)$. By observing bijections $\varphi, \psi$, and algorithm $\Phi$, we note that the fiber of any small trees in $M$ contains at most two adjacent unstarred vertices and call such fiber as $s$-fiber. Corresponding to the substructures of type $T_1, T_2, T_3, T_4$, the fiber of small trees of meadows can also be categorized into four types:

- $T_1$: an arbitrary $s$-fiber;
- $T_2$: an arbitrary nonempty $s$-fiber;
- $T_3$: an arbitrary $s$-fiber with no unstarred vertices;
- $T_4$: an arbitrary nonempty $s$-fiber with no unstarred vertices;

Applying the meadow decompose algorithm $\Phi$ to the stack in Figure 1.1 after step (1) we will get the simple stack shows in Figure 2.1. And the labeling linear tree in Figure 2.2 is one of the labeling linear trees we get after step (3).

3 **Enumeration of the saturated extended 2-regular simple stacks with given arc pattern**

In this section, we will study the general counting formula for extended 2-regular simple stacks with given arc pattern, and use this formula to derive the explicit counting formulas for 2-regular simple stacks and extended 2-regular simple stacks.

Let $[m, n]$ (resp. $[m, n]^*$) denote the interval $\{i, i + 1, \cdots, j\}$ (resp. $[i^*, (i + 1)^*, \cdots, j^*]$). Specially, we abbreviate $[1, n]$ as $[n]$. Note that the interval is allowed to be empty.

A **partition** of a finite set $S$ is a collection $\pi = \{B_1, B_2, \cdots, B_k\}$ of subsets $B_i \subseteq S$ such that $B_i \neq \emptyset$, $B_i \cap B_j = \emptyset$ for $i \neq j$, and $B_1 \cup B_2 \cup \cdots \cup B_k = S$. We call $B_i$ a **block** of
π. If elements of each block of π are ordered linearly, we shall call π a composition of S. An ordered partition (resp. ordered composition) is a partition (resp. composition) in which the blocks are linearly ordered. If a partition (resp. ordered partition, composition, ordered composition) π contains exactly k blocks, we call π a k-partition (resp. k-ordered partition, k-composition, k-ordered composition).

**Lemma 3.1.** The number of k-compositions of \([n]\) that each block contains at most two elements is \(\frac{n!}{k!(n-k)}\).

**Proof.** To obtain a k-ordered composition of \([n]\) that each block contains at most two elements, we may linearly order \([n]\) in \(n!\) ways and then divide the sequence into k linearly ordered nonempty compartments such that each compartment contains at most two elements. The ways to split the sequence can be easily obtained from the generating function

\[
[x^n](x + x^2)^k = [x^{n-k}] \sum_{i=0}^{k} \binom{k}{i} x^i = \binom{k}{n-k}.
\]

And the proof follows from dividing by the number of permutations of k compartments. ■

**Lemma 3.2.** The number of k-composition of \([s] \cup [k]^*\) that each block contains one and only one starred element is \(si(2k+s-1)\).

**Proof.** We can linearly order \([s] \cup [k]^*\) and \(k-1\) vertical bars, such that starred elements and vertical bars alternate, and then break the sequence at each vertical bar. To this end we linearly order \([s]\) and \([k]^*\) in \(s!\) and \(k!\) ways, respectively, and choose \(s\) positions from \((2k+s-1)\) ways to put unstarred elements. Alternatively, starred elements and vertical bars are placed in the rest \(2k-1\) positions. Dividing by the number of permutations of k blocks and lemma follows. Dividing by the number of permutations of k blocks and lemma follows. ■

**Lemma 3.3.** For all \(l \geq 1, r \geq 0, u \geq 0, v \geq 0, r + v \geq 1\). Denote the number of \((r+u+v)\)-ordered compositions of \([l] \cup [l+1,l+r]^*\) with the following three properties by \(C(l,r,u,v)\).

(a) There can be no more than two unstarred elements in each block, and two unstarred elements in the same block must be adjacent.

(b) First \(u\) blocks contain no unstarred elements.

(c) Element \((l+r)^*\) is in the first \(u + v\) blocks.

We have

\[
C(l,r,u,v) = l!r! \sum_{t=1}^{\min(l,r+v)} \binom{t}{l-t} \binom{r + v - 1}{t - 1} \binom{2t + r}{t - u - v} f(t, r, u, v), \tag{3.1}
\]

where

\[
f(t, r, u, v) = \frac{ut + (u + v)(t + r + v)}{t(2t + r)}.
\]
Proof. Assume that \( l \) unstarred vertices are distributed over exactly \( t \) blocks, and these \( t \) blocks contain \( s \) starred vertices, then \[ \frac{t}{2} \leq t \leq \min\{l, r + v\}, 0 \leq s \leq t - u - v. \] We denote \( r + u + v \) as \( p \) for convenience. When \( 1 \leq t \leq p - 1 \), we construct the ordered composition in five steps.

1. Construct a \( t \)-compositions \( C_1 \) of \( l \) unstarred vertices such that each block contains at most two vertices. According to Lemma 3.1 there are \( \frac{l!}{t!(l-t)!} \) ways.

2. Sort \( r \) starred vertices. The number of ways of sorting will be discussed along with the block ordering in step (5).

3. Construct a \( t \)-compositions \( C_2 \) of the first \( s \) starred vertices and blocks of \( C_1 \), such that each block contains exactly one block of \( C_1 \). Note that \( s \) starred vertices are already linearly ordered. By Lemma 3.2 there are \( \binom{2t+s-1}{s} \) ways.

4. Construct a \( (p-t) \)-composition of the latter \( r - s \) starred vertices. The number of ways to do this is \( \frac{1}{(p-t)!} \binom{r-s-1}{p-t-1} \).

5. Now we order the \( p \) blocks constructed in step (1), (3), (4) such that properties (b), (c) hold. Categorize the discussion by the position of \( (l+r)^* \) in the sort of step (2).

- If \( (l+r)^* \) is setted in the first \( s \) position, there are \( s(r-1)! \) ways. According to step (3), the block containing \( (l+r)^* \) contains unstarred vertices. Therefore we have \( vu!(r + v - 1)!(p-t)! \) ways to order the blocks.

- Else we have \( (r-s)(r-1)! \) ways to sort. In this case, the block containing \( (l+r)^* \) contains no unstarred vertices. If we place the block containing \( (l+r)^* \) in the first \( u \) positions, there are \( u(u-1)!(r + v)! \binom{p-t-1}{u-1} \) ways. Otherwise, the position of block containing \( (l+r)^* \) should be between \( u + 1 \) and \( u + v \), then the number of ways is \( vu!(r + v - 1)!(p-t-1) \).

When \( t = p \), all of the blocks contains unstarred elements and all \( r \) starred elements are contained in \( t \) blocks. Then we can similarly derive that the number of ordered composition is

\[
\binom{0}{u} \frac{l!}{(r+v)! (l-r-v)!} \binom{r+v}{l-r-v} \binom{3r+2v-1}{r} v(r+v-1)!. \]

We therefore have

\[
C(l, r, u, v) = \frac{l!r!v}{(r+v)} \binom{0}{u} \binom{r+v}{l-r-v} \binom{3r+2v-1}{r} + l!r! \sum_{t=1}^{\min\{l, r+v-1\}} \binom{t}{l-t} \binom{r+v-1}{t-1} \sum_{s=0}^{t-u-v} \binom{2t+s-1}{s} \binom{r-s-1}{p-t-1} g(s, t, r, u, v). \]

where

\[
g(s, t, r, u, v) = \frac{sv}{tr} + \frac{(r-s)[u(r+v) + v(r + v - t)]}{(p-t)tr}. \]

The summation with respect to \( s \) can be eliminated by the combinatorial identity in [11] (page 8) and theorem follows.
Theorem 3.1. Let $V$ be a given subset of $[n]$ and $A$ be an arc pattern w.r.t $V$ which splits $[n]$ into $I_1, I_2, J_1, J_2$ disjoint intervals of type $T_1, T_2, T_3$ and $T_4$, respectively. Denote $k_1$ as the number of arcs in $A$ and $d := \sum_{v \in V}(\deg v - 1)$. Let $P(n, k; k_1, d, I_1, I_2, J_1, J_2)$ denote the number of saturated stacks on $[n]$ with $k$ arcs whose arc pattern w.r.t $V$ is $A$. Then we have

$$P(n, k; k_1, d, I_1, I_2, J_1, J_2) = \sum_{i=0}^{l_1} \sum_{j=0}^{l_2} \binom{l_1}{i} \binom{l_2}{j} C(l, r, j, l_2 + i, l! r!), \quad (3.2)$$

where $l = n + d - 2k$, $r = k - k_1$, and $C(l, r, u, v) -$ see Lemma [3.3].

Proof. Note that the total number of vertices in the intervals is $n + d - 2k$ since the number of vertices of degree 2 minus the number of isolated vertices equals $d$. For the trivial case $k = k_1$, it is obviously that interval of type $T_4$ can not exist and interval of type $T_3$ must be empty. We suppose there are $i$ nonempty intervals in $I_1$ intervals of type $T_1$, so that $n + d - 2k$ vertices are distributed in $I_2 + i$ intervals and these intervals can contain only one or two isolated vertices. Choose $l - I_2 - i$ intervals from $I_2 + i$ intervals to place two vertices. Therefore

$$P(n, k_1; k_1, d, I_1, I_2, J_1, J_2) = \begin{cases} 0, & J_2 > 0 \\ \frac{1}{i} \sum_{i=0}^{l_1} \binom{l_1}{i} \binom{l_2}{j} C(l, r, J_2 + j, I_2 + i, l! r!), & J_2 = 0 \end{cases}, \quad (3.3)$$

The general case of $k > k_1$ is discussed as follow. By the decompose algorithm $\Phi$, the linear trees obtained after step (3) is of $n + d - k + 1$ vertices, with $k + 1$ inner vertices, $k_1 + 1$ vertices whose labels have been fixed and $n + d - k - k_1 = l + r$ vertices with random labels. After step (4) we get the meadows on $[n - 1 + d] \cap [n - k + 2 + d, n + d + 1]$ containing $k + 1$ small trees and satisfying the following properties.

(a) The vertices $1, 2, \cdots, k_1 + 1$ are roots of the small trees, and the positions of vertices $(n + d + 1)^*, \cdots, (n + d - k_1 + 2)^*$ are determined.

(b) Ignoring the determined vertices $(n + d + 1)^*, \cdots, (n + d - k_1 + 2)^*$. The fiber types of small trees with roots in $[k_1 + 1]$ are determined, where there are $I_1, I_2, J_1, J_2$ vertices of type $T_1, T_2, T_3$, and $T_4$, respectively.

(c) Vertex $(n + d - k_1 + 1)^*$ is a leaf of the small tree with roots $1, 2, \cdots, k_1 + 1$.

We take three steps to construct the meadows with these properties.

1. Assume there are $i, j$ nonempty fibers in the $I_1, J_1$ fibers of type $T_1$ and $T_3$, respectively. Choose these fibers in $\binom{l_1}{i} \binom{l_2}{j}$ ways.

2. Select the remaining $k - k_1$ root label from $[k_1 + 2, n - k + 1 + d]$ in $\binom{l + r}{k_1}$ ways.

3. Ignoring vertices $(n + d + 1)^*, \cdots, (n + d - k_1 + 2)^*$, there are $n + d - 2k$ unstarred vertices and $k - k_1$ starred vertices distributed in $r + i + j + I_2 + J_2$ s-fibers. Construct a $(r + i' + j')$-ordered composition of these vertices in $C(l, r, j', i')$ ways and allocate to the roots.
It derives that
\[(l + r)! P(n, k; d, I_1, I_2, J_1, J_2) = \sum_{i=0}^{I_1} \sum_{j=0}^{J_1} \binom{I_1}{i} \binom{J_1}{j} \binom{l + r}{r} C(l, r, j', i'). \] (3.4)
and the Equation 3.2 follows. It is easy to prove that Equation 3.2 equals 0 when \(k < k_1\) and reduces to Equation 3.3 when \(k = k_1\). Theorem therefore holds for all integer \(k\). □

Let \(LO(n, k)\) (resp. \(MO(n, k)\)) denote the number of all saturated 2-regular simple stacks (resp. with no visible vertex). Let \(ELO(n, k)\) (resp. \(EMO(n, k)\)) denote the number of all extended saturated 2-regular simple stacks (resp. with no visible vertex).

Corollary 3.1. For any \(n \geq 1, 0 \leq k \leq \left\lfloor \frac{n-1}{2} \right\rfloor\), we have

\[ LO(n, k) = \frac{1}{k + 1} \sum_{t=1}^{T} \binom{k + 1}{t} \binom{t}{n - 2k - t} \binom{2t + k - 1}{t - 1}. \]

For any \(n \geq 3, 1 \leq k \leq \left\lfloor \frac{n-1}{2} \right\rfloor\), we have

\[ MO(n, k) = \frac{1}{k} \sum_{t=1}^{T} \binom{k}{t} \binom{t}{n - 2k - t} \binom{2t + k}{t - 1}, \]

where \(T = \min\{n - 2k, k + 1\}\).

Proof. When \(n \geq 1\), a saturated 2-regular simple stacks is just a structure of type \(T_2\). Similarly, when \(n \geq 3\), a saturated 2-regular simple stacks with no visible vertices is just a structure of type \(T_4\). Take \(V = \emptyset\) and empty arc pattern \(A_e\) in Theorem 3.1, we immediately get

\[ LO(n, k) = P(n, k; 0, 0, 0, 1, 0, 0) = \frac{1}{k + 1} \sum_{t=1}^{T} \binom{k + 1}{t} \binom{t}{n - 2k - t} \binom{2t + k - 1}{t - 1}, \]
\[ MO(n, k) = P(n, k; 0, 0, 0, 0, 0, 1) = \frac{1}{k} \sum_{t=1}^{T} \binom{k}{t} \binom{t}{n - 2k - t} \binom{2t + k}{t - 1}. \]

Furthermore, the number of optimal 2-regular simple stacks of length \(n\) is given by the following formula.

Corollary 3.2. [4] For any \(k \geq 1\), we have

\[ LO(2k + 1, k) = 1, \]
\[ LO(2k, k - 1) = \binom{k + 1}{2}. \]
where $P$ vertex then the length of the structure becomes $T$ get the conclusion. For the case of $T$, can be presented as follows.

\[ \text{Theorem 3.2. For any } n \geq 6, k \geq 3, \text{ we have} \]
\[ ELO(n, k) = \sum_{t=1}^{k+1} \binom{k + 1}{t} \binom{2t + k}{t - 1} \sum_{i=1}^{4} \binom{n - 2k - 2 - t + i}{t - i} P_1(k, t), \]

where
\[ P_1(k, t) = \frac{2(k - t + 1)}{(k + 1)_{(k + 2t)}} [k^3 + (4t - 2)k^2 + (2t^2 - 4t + 1)k - 2t^3 - 4t^2 + 4t]. \]
\[ P_2(k, t) = \frac{k - t + 1}{(k + 1)_{(k + 2t)}} [7k^5 + (36t - 36)k^4 + (41t^2 - 137t + 61)k^3 + (-28t^3 - 105t^2 + 169t - 36)k^2 + (-38t^4 + 62t^3 + 96t^2 - 96t + 4)k + 12t^2 + 76t^4 - 96t^3 - 28t^2 + 36t]. \]
\[ P_3(k, t) = \frac{2(k - t + 1)_{(k + 2t)}}{(k + 1)_{(k + 2t)}} [k^4 + (7t - 6)k^3 + (15t^2 - 32t + 13)k^2 + (7t^3 - 44t^2 + 45t - 12)k - 4t^4 - 18t^3 + 48t^2 - 30t + 4]. \]
\[ P_4(k, t) = \frac{(t - 1)(k + t + 1)}{(k + 1)_{(k + 2t)}} [4k^6 + (16t - 42)k^5 + (8t^2 - 106t + 160)k^4 + (-24t^3 + 26t^2 + 204t - 270)k^3 + (-10t^4 + 132t^3 - 244t^2 - 50t + 196)k^2 + (13t^5 - 8t^4 - 217t^3 + 468t^2 - 208t - 48)k - 2t^6 - 20t^5 + 66t^4 + 68t^3 - 304t^2 + 192t]. \]

**Proof.** Denote the number of the structures with $n$ vertices and $k$ arcs of arc pattern $A_i$ in Table 2.2 by $s_i(n, k)$, and the number of such structures with no visible vertices by $\tilde{s}_i(n, k)$. Obviously, $A_i$ and $A_i^t (i = 1, 2, 3)$ are symmetric, we thus consider only $A_i$.

Set $V_0 = \{1, n\}$. To avoid lengthy formulas, we replace $n - 2k - t$ with $h$ in the following proof. Theorem 3.1 shows that the number of stacks of arc pattern $A_i, 1 \leq i \leq 6$ w.r.t. $V_0$ can be presented as follows.

For arc pattern $A_1$, there are 1 interval for each type of $T_1$ and $T_2$. We have two cases of the third interval with type $T_0$. For the case of $T_3$, we can use the Theorem 3.1 directly to get the conclusion. For the case of $T_3$ followed by an isolated vertex, we delete the isolated vertex then the length of the structure becomes $n - 1$. This process is clearly reversible.

\[ s_1(n, k) = P(n, k; 2, 0, 1, 1, 1, 0) + P(n - 1, k; 2, 0, 1, 1, 1, 0) \]
\[ = \sum_{t=1}^{k+1} \left[ \binom{t}{h} + \binom{t}{h - 1} \right] \sum_{i=0}^{1} \sum_{j=0}^{1} \binom{k + i - 2}{t - 1} \binom{2t + k - 2}{t - i - j - 1} f(t, k - 2, j, 1 + i). \]

where $f(t, r, u, v)$ defines in Lemma 3.3.

| $k$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|-----|---|---|---|---|---|---|---|---|---|
| 1   | 1 | 2 | 7 | 9 | 8 | 6 | 2 |
| 2   | 1 | 2 | 5 | 3 |

Table 3.1: $ELO(n, k)$ (left), $EMO(n, k)$ (right)
For arc pattern \( A_2 \), there is 1 interval of type \( T_2 \). We use the same technique used in arc pattern \( A_1 \) to deal with interval of type \( T_6 \).

\[
s_2(n, k) = P(n, k; 2, 1, 0, 1, 1, 0) + P(n - 1, k; 2, 1, 0, 1, 1, 0)
\]

\[
= \frac{1}{k - 1} \sum_{t=1}^{k+1} \binom{k - 1}{t} \left( \frac{2t + k - 1}{t - 1} \right) \left( \frac{t}{h + 1} + \frac{t}{h} \right).
\]

For arc pattern \( A_3 \), there is 1 interval for each type of \( T_1, T_2, \) and \( T_3 \). For the forth interval, if it is just a \( T_4 \), we can just apply Theorem 3.1 to get the result. Otherwise, there is \( T_3 \) followed by an isolated vertex. By deleting the isolated vertex, we obtain a structure with \( n - 1 \) vertices.

\[
s_3(n, k) = P(n, k; 3, 1, 1, 1, 1, 1) + P(n - 1, k; 3, 1, 1, 1, 2, 0)
\]

\[
= \sum_{t=1}^{k+1} \sum_{i=0}^2 \binom{k + i - 3}{t - 1} \left[ \binom{t}{h + 1} \binom{2t + k - 3}{t - i - j - 2} f(t, k - 3, 1 + j, 1 + i)
\]

\[
+ \binom{t}{h} \binom{2t + k - 3}{t - i - j - 1} f(t, k - 3, 1 + j, 1 + i) \right].
\]

In the same way, we have

\[
s_4(n, k) = P(n, k; 1, 0, 0, 0, 0, 1) = \sum_{t=1}^{n-2k} \binom{n - 2k - 1}{t - 1} \binom{t}{h} \left( \frac{2t + k - 1}{t - 1} \right).
\]

\[
s_5(n, k) = P(n, k; 3, 2, 1, 2, 0, 0) = \sum_{t=1}^{k+1} \sum_{i=0}^{2i} \frac{2 + i}{t} \binom{k - 2 + i}{t - 1} \binom{2t + k - 4}{t - 2 - i}.
\]

\[
s_6(n, k) = P(n, k; 4, 2, 3, 2, 0, 0) = \sum_{t=1}^{k+1} \sum_{i=0}^{3i} \frac{2 + i}{t} \binom{3 + i}{t - 1} \binom{2t + k - 5}{t - 2 - i}.
\]

Then substituting into

\[
ELO(n, k) = 2(s_1(n, k) + s_2(n, k) + s_3(n, k)) + s_4(n, k) + s_5(n, k) + s_6(n, k).
\]

Simplify the formula by Maple command, which completes the proof.

\[\blacksquare\]

**Corollary 3.3.** When \( k \geq 3 \), the number of optimal extended 2-regular simple stacks is

\[
ELO(2k, k) = 2k - 3,
\]

\[
ELO(2k + 1, k) = \frac{2}{3} k^3 - \frac{5}{3} k + 5.
\]

By Matlab, some of the \( ELO(n, k) \) values are listed in Table 3.2.

**Theorem 3.3.** For any \( n \geq 6, k \geq 3 \), we have

\[
EMO(n, k) = \sum_{t=1}^{k+1} \binom{k}{t} \left( \frac{2t + k}{t - 1} \right) \sum_{i=1}^{3} \binom{n - 2k - 1 - t + i}{t} Q_t(k, t),
\]

12
$\begin{array}{cccccccccccccccc}
\hline
k & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \\
\hline
1 & 1 &   &   &   &   &   &   &   &   &   &   &   &   &   \\
2 & 2 & 7 & 9 & 8 & 6 & 2 &   &   &   &   &   &   &   &   \\
3 & 3 & 18 & 46 & 73 & 82 & 70 & 40 & 10 &   &   &   &   &   &   \\
4 & 5 & 41 & 162 & 395 & 666 & 834 & 799 & 563 & 251 &   &   &   &   &   \\
5 & 7 & 80 & 444 & 1534 & 3667 & 6449 & 8690 &   &   &   &   &   &   &   \\
6 & 9 & 139 & 1026 & 4728 & 15151 &   &   &   &   &   &   &   &   &   \\
7 & 11 & 222 & 2099 &   &   &   &   &   &   &   &   &   &   &   \\
8 & 13 &   &   &   &   &   &   &   &   &   &   &   &   &   \\
\hline
\text{sum} & 1 & 2 & 7 & 12 & 26 & 57 & 116 & 251 & 545 & 1159 & 2517 & 5503 & 11962 & 26204 \\
\hline
k & 17 & 18 & 19 & 20 & 21 & 22 & 23 & 24 \\
\hline
4 & 50 &   &   &   &   &   &   &   &   &   &   &   &   &   \\
5 & 9146 & 7403 & 4312 & 1570 & 260 &   &   &   &   &   &   &   &   &   \\
6 & 35820 & 64919 & 92557 & 105168 & 94660 & 65265 & 32109 & 9875 &   &   &   &   &   &   \\
7 & 12362 & 50796 & 154746 & 1003604 & 1214930 & 1191281 &   &   &   &   &   &   &   &   \\
8 & 333 & 3921 & 28613 & 145817 & 371629 & 6100976 &   &   &   &   &   &   &   &   \\
9 & 15 & 476 & 6827 & 60299 & 371629 & 1708309 & 6100976 &   &   &   &   &   &   &   \\
10 & 17 & 655 & 11239 & 117860 & 862174 &   &   &   &   &   &   &   &   &   \\
11 & 19 & 874 & 17676 &   &   &   &   &   &   &   &   &   &   &   \\
12 & 21 &   &   &   &   &   &   &   &   &   &   &   &   &   \\
\hline
\text{sum} & 57711 & 127054 & 280704 & 622425 & 1381923 & 3074897 & 6858928 & 15323958 &   &   &   &   &   &   \\
\end{array}$

Table 3.2: $ELO(n, k), 3 \leq n \leq 24$

where

$Q_1(k, t) = \frac{k-t}{(k+2t)^3}[(k-21)k^3 + (50t^2 - 105t + 35)k^2 + (18t^3 - 138t^2 + 112t - 12)k - 12t^4 - 60t^3 + 112t^2 - 36t - 4]$.

$Q_2(k, t) = \frac{2(k-t)}{(k+2t)^3}[(k-6)k^3 + (15t^2 - 32t + 13)k^2 + (7t^3 - 44t^2 + 45t - 12)k - 4t^4 - 18t^3 + 48t^2 - 30t + 4]$.

$Q_3(k, t) = \frac{t-1}{(k+2t)^3}[4k^3 + (16t-26)k^2 + (14t^2 - 78t + 56)k + (-8t^3 - 58t^2 + 144t - 46)k^2 + (-8t^4 + 14t^3 + 90t^2 - 122t + 12)k + 3t^5 + 22t^4 - 17t^3 - 84t^2 + 72t]$.

**Proof.** As Table 2.2 shows, the structures of arc pattern $A_1, A_1'$ inevitably contain visible vertices, while the structure of arc pattern $A_2, A_2', A_3, A_3', A_4, A_5$ are all saturated 2-regular simple stacks with no visible vertex. Therefore for $i = 2, 3, 4, 5$, $\bar{s}_i(n) = s_i(n)$.

The structure with arc pattern $A_6$ contain no visible vertex if and only if the third interval contains a structure of type $T_3$. Thereby

$\bar{s}_6(n, k) = P(n, k; 4, 2, 2, 2, 1, 0)$

$= \sum_{t=1}^{k+1} \binom{n-2k+2-t}{t} \sum_{i=0}^{1} \sum_{j=0}^{1} \binom{2}{i} \binom{k+i-3}{t-i-j-2} f(t, k-4, j, i+2) \cdot (2t+k-4)$

substituting into

$ELO(n, k) = 2(s_2(n, k) + s_3(n, k)) + s_4(n, k) + s_5(n, k) + \bar{s}_6(n, k)$.

And the theorem can be obtained by simplifying it with the Maple command.
| $k$ | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | 20  | 21  | 22  | 23  | 24  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1   | 1   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 2   | 2   | 5   | 3   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 3   |     | 3   | 16  | 32  | 31  | 12  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 4   |     |     | 5   | 38  | 133 | 263 | 306 | 198 | 55  |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 5   |     |     |     | 7   | 76  | 390 | 1196| 2371| 3093| 2584|     |     |     |     |     |     |     |     |     |     |     |     |
| 6   |     |     |     |     | 9   | 134 | 935 | 3978| 11363|     |     |     |     |     |     |     |     |     |     |     |     |
| 7   |     |     |     |     |     | 11  | 216 | 1957|     |     |     |     |     |     |     |     |     |     |     |     |     |
| 8   |     |     |     |     |     |     | 13  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| sum | 1   | 2   | 5   | 6   | 16  | 32  | 31  | 12  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

| $k$ | 17  | 18  | 19  | 20  | 21  | 22  | 23  | 24  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 5   | 1260| 273 |     |     |     |     |     |     |
| 6   | 22689| 31998| 31417| 20520| 8035| 1428|     |     |
| 7   | 10870| 41244| 112568| 226462| 338135| 371646| 293026| 157179|     |     |
| 8   | 326  | 3712 | 25886| 124288| 435777| 1151940| 2333286| 3639555|     |     |
| 9   | 15   | 468  | 6533 | 55637 | 327207 | 1416835 | 4686114 |     |     |
| 10  | 17   |     | 646  | 10840 | 110407 | 776809 |     |     |
| 11  | 19   |     | 864  | 17150 |     |     |     |     |
| 12  |     |     | 21   |     |     |     |     |     |
| sum | 35145| 77242| 170339| 377820| 838230| 1863080| 4154418| 9276828|     |     |

Table 3.3: $EMO(n,k), 3 \leq n \leq 24$

**Corollary 3.4.** When $k \geq 3$, the number of optimal extended 2-regular simple stacks with no visible vertex is

$$EMO(2k, k) = 2k - 3,$$

$$EMO(2k + 1, k) = \frac{2}{3}k^3 - \frac{8}{3}k + 6.$$ 

By Matlab, some of the $EMO(n,k)$ values are given by Table 3.2.

**References**

[1] Banerjee, A.R., Jaeger, J.A. and Turner, D.H. 1993. Thermal unfolding of a group I ribozyme: the low-temperature transition is primarily disruption of tertiary structure. Biochemistry. 32, 153–163.

[2] Chen, W.Y., 1990. A general bijective algorithm for trees. Proceedings of the National Academy of Sciences, 87(24), 9635–9639.

[3] Chen, W.Y.C., Guo, Q.-H., Sun, L.H. and Wang, J. 2014. Zigzag stacks and $m$-regular linear stacks. J. Comput. Biol. 21 (12), 915–935.

[4] Clote, P. 2006. Combinatorics of saturated secondary structures of RNA. J. Comput. Biol. 13, 1640–1657.
[5] Clote, P., Kranakis, E., Krizanc, D. and Salvy, B. 2009. Asymptotics of canonical and saturated RNA secondary structures. J. Bioinform. Comput. Biol. 7, 869-893.

[6] Clote, P., Kranakis, E., Krizanc, D. and Stacho, L. 2007. Asymptotic expected number of base pairs in optimal secondary structure for random RNA using the Nussinov-Jacobson energy model. Discrete Appl. Math. 155, 759–787.

[7] Fusy, E. and Clote, P. 2014. Combinatorics of locally optimal RNA secondary structures. J. Math. Biol. 68, 341–375.

[8] Guo, Q.-H. and Sun, L.H. 2018. Combinatorics of Contacts in Protein Contact Maps. Bull. Math. Biol. 80, 385–403.

[9] Guo, Q.-H., Sun, L.H. and Wang, J. 2016. Enumeration of Extended m-Regular Linear Stacks. J. Comput. Biol. 23, 943–956.

[10] Guo, Q.-H., Sun, L.H. and Wang, J. 2017. Regular Simple Queues of Protein Contact Maps. Bull. Math. Biol. 79, 21–35.

[11] Riordan, J. 1968. Combinatorial identities. Wiley.

[12] Nebel, M.E. 2002. Combinatorial properties of RNA secondary structures. J. Comput. Biol. 9, 541–573.

[13] Reidys, R. 2011. Combinatorial Computational Biology of RNA Pseudoknots and Neutral Networks. Springer-Verlag, New York.

[14] Schmitt, W.R. and Waterman, M. 1994. Linear trees and RNA secondary structure. Discrete Appl. Math. 51, 317–323.

[15] Waldispühl, J. and Clote, P., 2007. Computing the partition function and sampling for saturated secondary structures of RNA, with respect to the Turner energy model. Journal of Computational Biology, 14(2), 190–215.

[16] Waterman, M. and Smith, T. 1978. RNA secondary structure: a complete mathematical analysis. Math. Biosci. 42, 257–266.

[17] Zuker, M., 1986. RNA folding prediction: The continued need for interaction between biologists and mathematicians. Lectures on Mathematics in the Life Sciences. 17, 87–124.

[18] Zuker, M., Jaeger, J.A. and Turner, D.H., 1991. A comparison of optimal and suboptimal RNA secondary structures predicted by free energy minimization with structures determined by phylogenetic comparison. Nucleic acids research, 19(10), 2707-2714.