A secondary structure is a (planar, labeled) graph on the vertex set \([n]\) having two kind of edges: the segments \([i, i + 1]\), for \(1 \leq i \leq n - 1\), and arcs in the upper half-plane connecting some vertices \(i, j\), where \(j - i > l\), for some fixed integer \(l\). Any two arcs must be totally disjoint. We establish connections between secondary structures and some well known combinatorial families, such as lattice paths, matchings and restricted permutations. Then we give some applications and connections with polygon dissections and polyominoes, using earlier enumerative results on secondary structures to provide explicit formulas and asymptotics for enumerating sequences of those families.

**Keywords:** secondary structure; Motzkin path; Dyck path; polygon dissection; restricted permutation

**MSC:** 05A15; 05A16; 05A20; 05B50; 05C30; 05C70; 33C45; 92D20

### 1.1 Introduction

Many interesting and important molecules belong to the class of linear polymers. That means they are long chains built from simpler building blocks called monomers. For example, proteins are chains of amino-acids. DNA molecules are linear polymers too, as well as closely related molecules of RNA. They together make the class of nucleic
acids, and their building blocks are called nucleotides. The nucleic acids are essential for coding, transferring and retrieving genetic information, and also in directing cell metabolism. In this paper we will look at RNA molecules.

RNA molecules are single-stranded and their nucleotides belong to one of four types. Those four type of nucleotides differ by only one part, called base. There are four different types of bases, denoted by letters A, C, G, and U. As we are not interested in details of their structure, we usually identify nucleotides with bases and use the same four letters for denoting nucleotides.

Nucleotides are polar molecules with two differing ends, usually denoted by 5′ and 3′. The 5′ end of one nucleotide readily binds to the 3′ end of another nucleotide by a phosphorus bond (or a p-bond). In that way, nucleotides form chains in which each of them (except the two terminal ones) is connected to exactly two neighbors via p-bonds. Those p-bonds form the backbone of the molecule. The sequence of nucleotides (or bases), read from the terminal nucleotide with free 5′ end, is the primary structure of the molecule. The total number of all possible primary structures of a molecule with n nucleotides is equal to $4^n$.

However, certain pairs of bases, such as C and G, A and U, and G and U, have an affinity to each other that enables them to bind via hydrogen bonds or h-bonds. Those bonds, when formed, cause folding of the molecular backbone into three-dimensional configurations that minimize potential energy. Some of the folded configurations remain planar; they are called secondary structures of an RNA molecule. Non-planar foldings are called tertiary structures. Both secondary and tertiary structures contribute to the shape of a folded molecule and thus determine its biological function. In this paper we look only at secondary structures.

The secondary structures of a given molecule are subject to certain stereo-chemical constraints. First, no base can participate in more than one h-bond. Second, a base cannot be paired by an h-bond to a base that is too close along the backbone, due to the rigidity of the backbone’s p-bonds. Third, and the most important, constraint on h-bonds is that they may not cross. It means that, if there is an h-bond pairing the bases $i$ and $j$, and an h-bond pairing the bases $k$ and $l$, then either $i < j < k < l$ or $i < k < l < j$. In biological terminology, this constraint prohibits pseudo-knots.

Besides their importance in molecular biology, secondary structures are also interesting from the mathematical point of view. We refer the reader to several papers dealing with various aspect of their mathematics [11, 15, 19, 20, 23, 24]. In particular, we will rely here on several structural and enumerative results established in [7, 8, 9].
The most suitable context for combinatorial modeling of secondary structures is the graph theory. We represent the bases by vertices, and the bonds by edges of certain graphs. The stereo-chemical constraints translate quite naturally into the graph-theoretical language. We refer the reader to [25] for all graph-theoretical terms not defined here.

Let $n$ and $l$ are integers, $n \geq 1, l \geq 0$. A secondary structure of size $n$ and rank $l$ is a labeled non-oriented graph $S$ on the vertex set $V(S) = [n] = \{1, 2, ..., n\}$ whose edge set $E(S)$ consists of two disjoint subsets, $P(S)$ and $H(S)$, satisfying the following conditions:

(a) $\{i, i + 1\} \in P(S)$, for all $1 \leq i \leq n - 1$;
(b) $\{i, j\} \in H(S)$ and $\{i, k\} \in H(S) \Rightarrow j = k$;
(c) $\{i, j\} \in H(S) \Rightarrow |i - j| > l$;
(d) $\{i, j\} \in H(S), \{k, l\} \in H(S)$ and $i < k < j \Rightarrow i < l < j$.

Obviously, the set $P(S)$ contains the edges corresponding to the $p$-bonds of the molecule’s backbone. The set $H(S)$, which may be empty, contains the edges representing the $h$-bonds. A secondary structure $S$ with $H(S) = \emptyset$ is called trivial. The number of edges in $H(S)$ is the order of $S$, and the parameter $l$ is the structure’s rank. The only secondary structure of size 1 is the graph $K_1$.

An example of a secondary structure of size 12, rank 1 and order 3 is shown on Fig.1. Note that every $h$-bond “leaps” over at least 1 base.

Some basic properties of secondary structures are stated in the following proposition.

**Proposition 1** [7] Let $S$ be a secondary structure of size $n \geq 1$ and rank $l \geq 0$. Then
(a) $S$ is connected;
(b) $S$ is simple for all $l \geq 1$;
(c) $S$ is sub-cubic, i.e. $\deg(v) \leq 3$ for all $v \in V(S)$;
(d) $S$ is outerplanar;
(e) girth$(S) \geq l + 2$;
(f) the weak dual of $S$ is a tree;
(g) $H(S)$ is a matching in $S$.

We denote the set of all secondary structures of size $n$ and rank $l$ by $S^{(l)}(n)$, while the set of all such structures of order $k$ is denoted by $S^{(l)}_k(n)$. The cardinalities of these sets, i.e., their enumerating sequences, will be denoted by $S^{(l)}(n)$ and $S^{(l)}_k(n)$, respectively. By definition, $S^{(l)}(0) = 1$, for all $l$.

There are many ways of representing secondary structures graphically. Three of them, the loop diagram, the chord diagram and the arc diagram, are shown on Fig.2 a), b) and c), respectively.

![Figure 2: The loop diagram, the chord diagram and the arc diagram representations of secondary structures](image)

1.2 Combinatorial context

In this Section we establish some connections between secondary structures and various well studied combinatorial families, such as lattice paths, matchings and permutations with forbidden patterns. We start with definitions of some important families of lattice paths and their enumerating sequences.

A Dyck path (or a “mountain path”) of length $2n$ is a lattice path in the coordinate plane $(x, y)$ from $(0, 0)$ to $(2n, 0)$ with steps $(1, 1)$ (Up) and $(1, -1)$ (Down), never falling
below the x-axis. The number of steps in a Dyck path is called its **length** and it must be even. We denote the set of all Dyck paths of length 2n by \( \mathcal{D}(n) \).

A **peak** of a Dyck path is a place where an *Up* step is immediately followed by a *Down* step. The set of all Dyck paths of length 2n with exactly \( k \) peaks (\( 1 \leq k \leq n \)) is denoted by \( \mathcal{D}_k(n) \).

It is well known that Dyck paths are enumerated by Catalan numbers, i.e. that \( |\mathcal{D}(n)| = C_n \). (Recall that the \( n \)-th **Catalan number**, \( C_n, n \geq 0 \), is defined by \( C_n = \frac{1}{n+1} \binom{2n}{n} = \frac{1}{2n+1} \binom{2n+1}{n} \). Hundreds of combinatorial families enumerated by Catalan numbers are listed in Exercise 6. 19 of [21] and in [22].)

A **Motzkin path** of length \( n \) is a lattice path in \((x, y)\) plane from \((0, 0)\) to \((n, 0)\) with steps \((1, 1)\) (*Up*), \((1, -1)\) (*Down*) and \((1, 0)\) (*Level*), never falling below the x-axis. We denote the set of all Motzkin paths of length \( n \) (i.e. with exactly \( n \) steps) by \( \mathcal{M}(n) \). The number \( M_n = |\mathcal{M}(n)| \) is \( n \)-th **Motzkin number**. By definition, \( M_0 = 1 \). A typical member of the Motzkin family \( \mathcal{M}(17) \) is shown on Fig.3.

---

**Figure 3:** A Motzkin path of length 17

A **peak** of a Motzkin path is a place where an *Up* step is immediately followed by a *Down* step. A **plateau** of length \( l \) is a sequence of \( l \) consecutive *Level* steps, immediately preceded by an *Up* step, and immediately followed by a *Down* step. A sequence of \( m \geq 0 \) consecutive *Level* steps, immediately preceded by a *Down* step, and immediately followed by an *Up* step is called a **valley** of width \( m \). If \( m = 0 \), we speak of a **trench**. A sequence of \( m \geq 1 \) consecutive *Level* steps, immediately preceded by a *Down* step, and directly followed by an *Up* step (or preceded by an *Up* step and followed by a *Down* step) is called a **terrace** of width \( m \). In our example of Fig.3 we can see
peaks at the points (5, 3) and (16, 1), a plateau of length 2 formed by the steps 11 and 12, terraces of width 1 formed by the steps 2, 4 and 8, a valley of width 2 formed by the steps 14 and 15, and a trench formed by the steps 9 and 10. We notice that a Dyck path is a Motzkin path without any Level steps.

Let us denote by $M^{(l)}(n)$ the set of all paths from $M(n)$ whose every plateau is at least $l$ steps long. For $l = 0$ we get simply $M^{(0)}(n) = M(n)$. There is a bijective correspondence between $M^{(l)}(n)$ and $S^{(l)}(n)$.

**Proposition 2** There is a bijection between $M^{(l)}(n)$ and $S^{(l)}(n)$, for all $n \geq 1$, $l \geq 0$.

This result belongs to mathematical folklore and was rediscovered several times. The proof is very simple. Indeed, take a secondary structure from $S^{(l)}(n)$ and scan it from left to right. Assign a Level step to each unpaired base, an Up step to each base in which a new $h$-bond starts, and a Down step to each base in which an already encountered $h$-bond terminates. It is obvious that the lattice path obtained in this way never falls below the $x$-axis, that it has equally many Up and Down steps (and hence terminates at $(n, 0)$), and that at least $l$ Level steps must be found between any Up and any Down step. So, this lattice path belongs to $M^{(l)}(n)$. The injectivity of the construction is obvious.

The construction is easily seen to be invertible. Start from the first step in a path from $M^{(l)}(n)$ and proceed toward the right constructing a secondary structure by assigning an unpaired base to each Level step. To an Up step assign a base in which an $h$-bond starts, and to a Down step assign a base in which the last-started $h$-bond terminates. Properties of Motzkin paths now ensure that so formed $h$-bonds will not cross and that the endpoints of any $h$-bond will be separated by at least $l$ unpaired bases. So, we obtain a valid secondary structure in $S^{(l)}(n)$, and the injectivity of the construction is, again, obvious.

It is clear from the above construction that the number of Up steps in the Motzkin path is equal to the order of the corresponding secondary structure. A secondary structure from $S^{(1)}(12)$ and the corresponding Motzkin path from $M^{(1)}(12)$ are shown in Fig.4.

It was shown in Proposition 1.2 of [8] that the correspondence can be extended also to some negative values of $l$. For example, there is a bijection between $S^{(-1)}(n)$ and $D(n + 1)$, where $S^{(-1)}(n)$ is the set of all secondary structures on $n$ bases in which an $h$-bond can form a loop, i.e., it may terminate in the same base it starts from. Of course, such mathematical objects do not correspond to biochemically realistic structures.
Let us now turn our attention to some other combinatorial interpretations of secondary structures. One of them reflects the fact that h-bonds form a matching in a secondary structure graph.

For a given graph $G$, a matching in $G$ is any set $M$ of edges of $G$ such that no two edges from $M$ have a vertex in common. From a given graph $G$ we get its $m$-th power, $G^m$, by connecting by an edge any two vertices whose distance in $G$ is at most $m$.

Now take a graph $G$ and denote by $G'$ one of its representations in the plane. A matching $M$ is non-intersecting with respect to $G'$ if edges from $M$ do not intersect in $G'$. In Fig.5 we see two representations, $G'$ and $G''$ of the same graph $K_4$, and a matching $M = \{ac, bd\}$ which is non-intersecting w.r.t $G''$, but not w.r.t $G'$.
**Theorem 3** There is a bijection between $S(l)(n)$ and the set of all matchings non-intersecting w.r.t. circular representation of $K_n - E(P_n^l)$, where $P_n$ is a path on $n$ vertices.

**Proof.** Take a copy of $K_n$ and represent it in the plane so that its vertices lie on a circle. Number them consecutively in, say, clockwise direction from 1 to $n$. Denote by $P_n$ the path consisting of the edges of the form $\{i, i+1\}$, for $1 \leq i \leq n-1$, and remove all the edges of $P_n^l$. It is obvious that every remaining edge in this graph corresponds to a possible $h$-bond in a secondary structure in a molecule whose backbone corresponds to $P_n$. Any matching in this graph satisfies the conditions (b) and (c) in the definition of secondary structure, and non-intersecting matchings satisfy also the condition (d). So we can assign a secondary structure $S$ of size $n$ and $H(S) = M$ to every non-intersecting matching $M$ in $K_n - E(P_n^l)$, and this assignment is obviously injective.

On the other hand, take a secondary structure $S$ from $S(l)(n)$. From Proposition 1 we know that $H(S)$ is a matching, and by definition of secondary structure, this matching must be non-intersecting w.r.t circular representation of $K_n$. As no $h$-bond can connect bases whose distance is less than $l+1$, all permitted $h$-bonds must be edges in $K_n - E(P_n^l)$, where edges of $P_n$ correspond to the molecule’s backbone. □

We conclude this section by pointing to a connection between secondary structures and pattern avoidance in permutations. A **pattern** is a permutation $\sigma \in \Sigma_k$, and a permutation $\pi \in \Sigma_n$ **avoids** $\sigma$ if there is no subsequence in $\pi$ whose members are in the same relative order as the members of $\sigma$. It is well known that the number of permutations from $\Sigma_n$ avoiding $\sigma \in \Sigma_3$ is $C_n$, for all patterns $\sigma \in \Sigma_3$ [6].

The notion of pattern avoidance was generalized in [1], by allowing the requirement that two letters adjacent in a pattern must be adjacent in the permutation. An example of a generalized pattern is $1-32$, where an $1-32$ subword of a permutation $\pi = a_1 \ldots a_n$ is any subword $a_i a_j a_{j+1}$ such that $i < j$ and $a_i < a_{j+1} < a_j$. Generalized pattern avoidance is treated in more detail in [4], where it is shown that the permutations from $\Sigma_n$ that avoid both $1-23$ and $13-2$ are enumerated by the Motzkin numbers.

**Proposition 4** There is a bijection between $S(l)(n)$ and the set of all permutations from $\Sigma_n$ that avoid $\{1-23, 13-2, ij\}$, where $j \leq i+1$, for all $l \geq 0$.

**Proof.** The claim follows by closer inspection of the bijection from the proof of Proposition 24 in [4]. It is easy to see that any pattern $i, i+k$ in a permutation $\sigma \in \Sigma_n$ avoiding
\{1 - 23, 13 - 2\} generates a plateau of length \(k - 1\) in the corresponding Motzkin path.

1.3 Decomposable secondary structures

We now make a digression and consider a special class of secondary structures. A secondary structure is decomposable if there is a base whose removal leaves two valid secondary structures. (One, or even both of them may be empty.) In other words, a secondary structure is decomposable if there is a horizontal step on the \(x\)-axis in the corresponding Motzkin path. We would like to know how many of all secondary structures of given rank and size belong to the class of decomposable structures.

Denote by \(D^{(l)}(n)\) and \(U^{(l)}(n)\) the sets of all decomposable and undecomposable secondary structures of rank \(l\) and size \(n\), respectively. Put \(D^{(l)}(n) = |D^{(l)}(n)|\), \(U^{(l)}(n) = |U^{(l)}(n)|\).

**Theorem 5** \(U^{(l)}(n) = S^{(l)}(n) - U^{(l)}(n - 1) - \ldots - U^{(l)}(n - l - 1)\), for \(n, l \geq 0\).

**Proof.** We first exhibit a bijection between \(D^{(l)}(n - 1)\) and a certain subset \(P_n\) of \(S^{(l)}(n)\). The claim will then follow by analyzing the structure of \(P_n\). We describe the bijection in terms of Motzkin paths with plateaus of length at least \(l\). Take a path \(Q\) from \(D^{(l)}(n - 1)\) and consider its last (the rightmost) horizontal step on the \(x\)-axis. The sub-path on the right of this step we leave unchanged, and the rest we transform as follows. We replace the rightmost horizontal step by a Down step, we insert a Up step in front of the first (the leftmost) step, and we increase the altitudes of all intervening steps by one. In this way we obtain a path \(P\) from \(U^{(l)}(n)\). Put \(P = \psi(Q)\). It is obvious that \(\psi\) is an injection. Denote by \(P_n\) the image of the set \(D^{(l)}(n - 1)\) by the mapping \(\psi\). Obviously, \(|P_n| = |D^{(l)}(n - 1)|\), and there is a function \(\varphi : P_n \rightarrow D^{(l)}(n - 1)\) such that \(\varphi = \psi^{-1}\). The function \(\varphi\) acts as follows. For a given path \(P \in P_n\), consider its first return to the \(x\)-axis. Then transform the (elevated) sub-path on the left hand side of this return in the following way. Omit the first step (which was an Up step), replace the last step (a Down step) by a horizontal step on the \(x\)-axis, and translate all the steps in between one unit down. The sub-path of \(P\) on the right hand side of its first return to the \(x\)-axis leave unchanged. The path \(\varphi(P)\) is obviously an element of \(D^{(l)}(n - 1)\).

Now take a path from \(U^{(l)}(n)\). For any such path \(S\) we have \(\varphi(S) \in D^{(l)}(n - 1)\), so \(U^{(l)}(n)\) must be a subset of \(P_n\). Are there any elements in \(P_n - U^{(l)}(n)\)? The answer is yes, because the function \(\psi\) can produce a path whose plateau length will be less than \(l\).
But the only way to get such a path is to start from a path in $D^{(l)}(n-1)$ beginning with a sequence of $m$ consecutive horizontal steps on the $x$-axis, where $1 \leq m \leq l$. There are $U^{(l)}(n-1-m)$ such paths for any given $m$. So, the total of $\sum_{m=1}^{l} U^{(l)}(n-1-m)$ paths from $D^{(l)}(n-1)$ will be mapped into elements of $P_n - U^{(l)}(n)$. Now we have

$$D^{(l)}(n-1) = U^{(l)}(n) + \sum_{m=1}^{l} U^{(l)}(n-1-m).$$

From $D^{(l)}(n-1) = S^{(l)}(n-1) - U^{(l)}(n-1)$ we get the claim. ■

**Corollary 6** There are equally many Motzkin paths of length $n$ without horizontal steps on the $x$-axis and Motzkin paths of length $n-1$ with at least one horizontal step on the $x$-axis.

This Corollary also gives a solution of Problem 10816 in [5], p.652.

### 1.4 Sequences $S^{(l)}(n)$ and $S^{(l)}_k(n)$

In this section we summarize the main results of reference [8]. In particular, we show how to obtain formulas for the number of secondary structures of given size, rank and order.

Some insight can be gained with little effort. It is clear that there is no secondary structure of order $k > \frac{n-l}{2}$, since each $h$-bond consumes 2 bases, and at least $l$ bases must remain unpaired. Using the same arguments we can conclude that the structure of the maximal order is unique in a molecule whose size $n$ has the same parity as the rank $l$. Also, for the molecules of size $n < l + 2$, only the trivial structure is possible. So, the numbers $S^{(l)}_k(n)$ will form a triangle in a table whose rows are indexed by $n$ and columns by $k$. All non-zero elements of this table will be below the line $n - 2k = l$, except $l$ 1’s, counting the trivial secondary structures in molecules of size $n < l$.

Before we describe the solution of the problem of determining $S^{(l)}_k(n)$, some combinatorial preliminaries are in order.

First recall that Dyck paths are lattice paths from $(0,0)$ to $(2n,0)$ using only the steps $(1,1)$ and $(1,-1)$, never falling below the $x$-axis. The set of all Dyck paths of length $2n$ we denote by $D(n)$, and their number is $C_n$, the $n$-th Catalan number.
The **Narayana numbers** $N(n, k)$ are defined for integers $n, k \geq 1$ by

$$N(n, k) = \frac{1}{n} \binom{n}{k} \binom{n}{k - 1} = \frac{1}{k} \binom{n}{k - 1} \binom{n - 1}{k - 1},$$

with the initial value $N(0, 0) := 1$ and the boundary values $N(n, 0) = 0, N(n, 1) = 1$ for $n \geq 1$.

It is well known that the Narayana numbers $N(n, k)$ enumerate Dyck paths of length $2n$ with exactly $k$ peaks. This fact is instrumental in obtaining explicit formulas for $S_k^{(l)}(n)$.

**Theorem 7** [8]

$$S_k^{(l)}(n) = \sum_{p=1}^{k} \frac{1}{k} \binom{k}{p} \binom{k}{p - 1} \binom{n - lp}{2k}$$

$$S_k^{(l)}(n) = \sum_{p=1}^{k} N(k, p) \binom{n - lp}{2k}$$

for all $n, p, l \geq 0$.

**Proof.** Consider a Dyck path $P$ on $2k$ steps with exactly $p$ peaks. We know that there are $N(k, p)$ such paths. By inserting $l$ horizontal steps between every two steps forming a peak, we get a path $P'$ in $M^{(l)}(2k + lp)$, the set of all Motzkin paths on $2k + lp$ steps with plateaus of length at least $l$. There are exactly $N(k, p)$ such paths. Now take additional $m$ horizontal steps and distribute them at will in the path $P'$. From every path $P' \in M^{(l)}(2k + lp)$ we can get $\binom{2k+m}{m}$ paths in $M^{(l)}(2k + lp + m)$. Denoting the total number of steps by $n$, we get $N(k,p) \binom{n-lp}{2k}$ as the number of paths from $M^{(l)}(n)$ which have exactly $p$ plateaus. Summing over all $p$‘s from 1 to $k$, we get the number of paths in $M^{(l)}(n)$ with $k$ steps of the form $(1,1)$, whose all plateaus are at least $l$ steps long. Recalling our correspondence between such Motzkin paths and secondary structures of rank $l$ (Proposition 2), we get the claim of the theorem. 

**Corollary 8** $S_k^{(l)}(n) = N(n-k, k+1) = \frac{1}{k+1} \binom{n-k}{k} \binom{n-k-1}{k} = \frac{1}{n-k} \binom{n-k}{k} \binom{n-k}{k+1}$ for all $n, k \geq 0$. 

29
The first few rows of the triangle of \( S_k^{(1)}(n) \) numbers are given in Table 1. below. Closer look at this table reveals certain symmetry pattern. We see that \( S_k^{(1)}(n) = S_{n-2k-1}^{(1)}(2n - 3k - 1) \). This symmetry is a consequence of the symmetry property of Narayana numbers. We give here a combinatorial proof of this property.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( k )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>( \Sigma = S_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>3</td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>10</td>
<td>6</td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>15</td>
<td>20</td>
<td>1</td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>21</td>
<td>50</td>
<td>10</td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>28</td>
<td>105</td>
<td>50</td>
<td>1</td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>36</td>
<td>196</td>
<td>175</td>
<td>15</td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>45</td>
<td>336</td>
<td>490</td>
<td>105</td>
<td>1</td>
<td> </td>
<td> </td>
<td> </td>
</tr>
</tbody>
</table>

Table 1: The beginning rows of the triangle \( S_k^{(1)}(n) \)

**Proposition 9** There is a bijection between the sets \( S_k^{(1)}(n) \) and \( S_{n-2k-1}^{(1)}(2n - 3k - 1) \).

**Proof.** Take a secondary structure \( S \) from \( S_k^{(1)}(n) \). Map the structure \( S \) to a Dyck path \( P \) in \( \mathcal{D}_{n-2k}(n-k) \). Now invoke the bijection \( f: \mathcal{D}_{n-2k}(n-k) \to \mathcal{D}_{k+1}(n-k) \) that accounts for the symmetry property \( \mathcal{N}(n,k) = \mathcal{N}(n,n+1-k) \) of Narayana numbers and map \( P \) into \( P' \in \mathcal{D}_{k+1}(n-k) \). Finally, map the path \( P' \) back to a secondary structure. The obtained secondary structure lies in \( S_{n-2k-1}^{(1)}(2n - 3k - 1) \) and the correspondence is obviously bijective. ■

We say that the secondary structures from \( S_{n-2k-1}^{(1)}(2n - 3k - 1) \) are **dual** to the structures from \( S_k^{(1)}(n) \). The duality of sets \( S_1^{(1)}(5) \) and \( S_2^{(1)}(6) \) is illustrated on Fig. 6.
1.5 Some geometric applications

The various problems of polygon dissections have attracted much attention during last two centuries. Most of them require using of non-intersecting diagonals, i.e. diagonals which do not intersect in the interior of the considered polygon (e.g. [2, 3, 18]). Strengthening a little bit the condition of non-intersecting, and requiring that the the diagonals be totally disjoint, i.e. having no common point at all, we get an interesting variant of the polygon dissection problem, which, to the best of our knowledge, received no attention so far.

Problem
In how many ways one can dissect a convex \( n \)-gon using exactly \( k \) totally disjoint diagonals?

The problem also has an equivalent formulation as follows: How many ways are there of selecting \( k \) totally disjoint diagonals of an \( n \)-gon?

Let us denote the number of dissections of an \( n \)-gon by \( k \) totally disjoint diagonals by \( D_k(n) \), and the total number of all such dissections of an \( n \)-gon by \( D(n) \). Clearly,
\[ D(n) = \sum_{k=0}^{\left\lfloor \frac{n-2}{2} \right\rfloor} D_k(n). \] All the four dissections of an 8-gon by three totally disjoint diagonals are shown on Fig.7.

![Figure 7: Dissections of an 8-gon with totally disjoint diagonals](image)

We present here a simple solution of this problem, based on connections between the polygon dissections and secondary structures.

**Theorem 10** \[ D_k(n) = \frac{n}{k(k+1)} \binom{n-k-2}{k} \binom{n-1}{k-1}, \text{ for all } n \geq 3, 1 \leq k \leq \left\lfloor \frac{n-2}{2} \right\rfloor. \]

**Proof.** Let us consider a secondary structure (of rank 1) of \( k \)-th order in a molecule with \( n \) bases, \( k \geq 1 \). If the considered structure contains an \( h \)-bond connecting the bases 1 and \( n \), then the remaining \( k-1 \) \( h \)-bonds form a secondary structure in a molecule of size \( n-2 \). If the considered structure does not contain such an \( h \)-bond, then its \( h \)-bonds form a set of \( k \) totally disjoint diagonals in an \( n \)-gon obtained by adding the side \( 1-n \) to the molecule’s backbone. From this observation we get \( S_k^{(1)}(n) = S_{k-1}^{(1)}(n-2) + D_k(n), \) and then \( D_k(n) = S_k^{(1)}(n) - S_{k-1}^{(1)}(n-2). \) The claim now follows by substituting the explicit value for \( S_k^{(1)}(n) \) from Corollary 8. \( \blacksquare \)

Obviously, \( D_0(n) = 1 \), for all \( n \geq 3 \).

The numbers \( D_k(n) \) form a triangular number array. We tabulate here its first few rows:

The following results are immediate consequences of Theorem 10:

**Corollary 11**
\[
D(n) = S^{(1)}(n) - S^{(1)}(n-2), \quad n \geq 3 \\
\sum_{k \geq 0} D_k(n + k) = C_n - C_{n-1}, \quad n \geq 3.
\]
Table 2: The beginning rows of the triangle $D_k^{(1)}(n)$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$k$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>$\Sigma = D_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>1</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>1</td>
<td>9</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>1</td>
<td>14</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>29</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>1</td>
<td>20</td>
<td>40</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>65</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>1</td>
<td>27</td>
<td>90</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td>148</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>1</td>
<td>35</td>
<td>175</td>
<td>125</td>
<td>5</td>
<td></td>
<td></td>
<td>341</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>1</td>
<td>44</td>
<td>308</td>
<td>385</td>
<td>55</td>
<td></td>
<td></td>
<td>793</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>1</td>
<td>54</td>
<td>504</td>
<td>980</td>
<td>315</td>
<td>6</td>
<td></td>
<td>1860</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>1</td>
<td>65</td>
<td>780</td>
<td>2184</td>
<td>1274</td>
<td>91</td>
<td></td>
<td>4395</td>
</tr>
</tbody>
</table>

It can be checked by direct calculations that the numbers $D_k(n)$ are log-concave (and hence unimodal) in $k$. By expressing $D_k(n)$ in the form $D_k(n) = \frac{n(n-2k)(n-2k-1)}{(k+1)(n-k)^2(n-k-1)} \binom{n-k}{k}^2$, we can establish the following asymptotic behavior of the position of the maximal value of $D_k(n)$.

**Proposition 12** Let $k_n$ be the value of $k$ for which the maximal value of $D_k(n)$ is attained, for a fixed $n$. Then $\lim_{n \to \infty} \frac{k_n}{n} = \frac{5-\sqrt{5}}{10}$.

It is interesting to note that the same value, $\frac{5-\sqrt{5}}{10}$, appears as the asymptotic limit of several statistics on non-decreasing Dyck paths and corresponding trees [17].

Let us now impose an additional constraint on our dissections, requiring that the endpoints of any diagonal must be separated by at least $l \geq 1$ intervening vertices. In other words, we require that each ear of the dissection [12] be at least $l + 1$ edges long. Denote the number of such dissections of a convex $n$-gon by $k$ totally disjoint diagonals by $D_k^{(l)}(n)$. By a reasoning similar to that in Theorem 10, we obtain the following result.
Theorem 13 For all \( n \geq 3, 1 \leq k \leq \lfloor \frac{n-2}{2} \rfloor \) and \( l \geq 1 \) we have

\[
D_k^{(l)}(n) = S_k^{(l)}(n) - \sum_{m=1}^{l} m S_{k-1}^{(l)}(n-m-1).
\]

Recall that the dissections of a convex \( n \)-gon by diagonals that do not intersect in its interior are counted by the little Schröder numbers \( s_{n-2} \) (sequence A001003 in [16]). We can compare asymptotic behavior of the two types of dissections.

Proposition 14

\[
\frac{D_n}{s_{n-2}} \sim \sqrt{\frac{15 + 7\sqrt{5}}{6\sqrt{2} - 8}} \left(1 - \frac{2}{n}\right)^{-3/2} \left(\frac{3 + \sqrt{5}}{6 + 4\sqrt{2}}\right)^n.
\]

Proof. Follows by comparing the asymptotics \( D_n \sim S^{(1)}(n) \sim \sqrt{\frac{15 + 7\sqrt{5}}{8\pi}} \left(\frac{3 + \sqrt{5}}{2}\right)^n n^{-3/2} \) (Proposition 2.2 of [8]) with the asymptotics for the little Schröder numbers \( s_n \) given in [16].

It is also possible to think of polygon dissections in terms of polygonal clusters [6]. We say that a connected and bounded set \( W \subset \mathbb{R}^2 \) is a **polygonal cluster** if \( W \) is a finite union of convex polygons which are either pairwise disjoint or share a whole edge. Each polygon in a cluster is called a **cell**. The number of cells in a cluster is its **size**, and the number of edges in its boundary is its **circumference**. Borrowing a term from benzenoid chemistry, we say that a cluster is **catacondensed** if none of its vertices is shared by more than two of its cells. The minimal number of sides over all its cells is the **girth** of a given cluster. Example of a cluster is shown on Fig. 8 a), and a catacondensed cluster is represented on Fig. 8 b).

Proposition 15 The number of all catacondensed polygonal clusters of size \( s \) and circumference \( r \) is \( D_{s-1}(r) \). The number of all catacondensed polygonal clusters of size \( s \) with circumference \( r \) and girth at least 4 (i.e. without triangles) is \( D_{s-1}^{(2)}(n) \).

Let us present now another geometric consequence of our results on secondary structures. Consider a circle and a set of \( k \) randomly chosen chords on this circle. What is the probability that no two of them are intersecting?
Proposition 16 The probability that among \( k \) randomly chosen chords of a circle no two of them intersect is \( \frac{2^k}{(k+1)!} \).

Proof. Consider a circular representation of \( K_n - E(C_n) \), where \( E(C_n) \) stands for the set of edges of a cycle on \( n \) vertices. There are \( \frac{n(n-3)}{k} \) \( k \)-element matchings in this graph, and any of them corresponds to a choice of a \( k \)-element subset from the set of all diagonals in a convex (regular) \( n \)-gon. Only \( D_k(n) \) of such subsets are totally disjoint. Our claim follows by considering the quotient \( \frac{D_k(n)}{\binom{n(n-3)}{k}} \), using Theorem 10, and letting the parameter \( n \) tend to infinity.

Our final geometric application of secondary structures concerns the theory of polyominoes. (We refer the reader to [10] and [13] for basic definitions and results about polyominoes.)

A glacial landscape of size \( n \) is a Motzkin path on \( n \) steps without peaks and trenches. The name is motivated by the resemblance of such paths to the landscapes eroded by action of glaciers: the peaks are flattened, and the trenches are widened into valleys. The set of all glacial landscapes of size \( n \) is denoted by \( \mathcal{G}(n) \), and the set of all paths from \( \mathcal{G}(n) \) with exactly \( k \) Up steps we denote by \( \mathcal{G}_k(n) \). Obviously, \( \mathcal{G}(n) \subset \mathcal{M}^{(1)}(n) \).

Proposition 17

\[
|\mathcal{G}_k(n)| = \sum_{j=1}^{k} N(k,j) \binom{n-2j+1}{2k},
\]
for all \( n, k \geq 0 \).

**Proof.** We apply the same reasoning as in the proof of Theorem 7. Namely, there are \( N(k,j) \) Dyck paths on \( 2k \) steps with exactly \( j \) peaks (and hence with exactly \( j-1 \) trenches). By flattening the peaks and broadening the trenches with one Level step each, we get \( N(k,j) \) glacial landscapes of size \( 2k + 2j - 1 \), whose all plateaus and all valleys are exactly one step wide. Additional \( m \) Level steps can now be distributed in each of these landscapes in \( \binom{m+2k}{2k} \) different ways. The claim now follows by denoting the total number of steps by \( n \) and summing over all \( 1 \leq j \leq k \).

Hence, \( G_k(n) = |\mathcal{G}_k(n)| = \sum_{j=1}^{k} N(k,j) \binom{n-2j+1}{2k} \).

A closer look at the formula for \( G_k(n) \) reveals that \( G_k(n) = S^{(2)}(k)(n+1) \). In a similar way one can show that \( G^{(l)}(n) \), the number of all Motzkin paths on \( n \) steps whose all plateaus and valleys are at least \( l \) steps wide is equal to \( \sum_{k} \sum_{j} N(k,j) \binom{n-2lj+l}{2k} = S^{(2l)}(n+l) \). The numbers \( G^{(l)}(n) \) can also be interpreted as the numbers of all walks on nonnegative integers, beginning and ending at \( 0 \), with the steps \( 1 \), \( 0 \) and \( -1 \) with inertia \( l \). It means that the current direction of the walk cannot be quickly reverted.

Let us consider a glacial landscape \( M \) from \( \mathcal{G}_k(n) \). Clearly, the path \( M \) has \( k \) steps of the form \((1,1)\), and \( k \) steps of the form \((1,-1)\). Replacing these steps by the steps of the form \((0,1)\) and \((0,-1)\), respectively, we get a lattice path \( P' \) from \((0,0)\) to \((n-2k,0)\) with steps \((0,1),(0,-1)\) and \((1,0)\) which never falls below the \( x \)-axis. Now join the points \((0,0)\) and \((n-2k,0)\) by another lattice path, starting from \((0,0)\) by an \((0,-1)\) step, then adding \( n-2k \) steps of the form \((1,0)\) and finally a \((0,1)\) step to reach the point \((n-2k,0)\). Call this path \( P'' \). The figure \( P \) enclosed between the paths \( P' \) and \( P'' \) is a polyomino with \( p = n - 2k \) columns and circumference of \( 2r = 2n - 2k + 2 \). We call the polyominoes of such shape Manhattan polyominoes. An example of Manhattan polyominoes with the parameters \( p = 14, r = 20 \) is shown on Fig.9. The set of all Manhattan polyominoes with parameters \( p \) and \( r \) we denote by \( \mathcal{P}_p(r) \).

In this way to every glacial landscape from \( \mathcal{G}_k(n) \) we can assign a polyomino from \( \mathcal{P}_{n-2k}(n-k+1) \). It is obvious that this mapping is injective and that we can invert it. Hence, we have

**Proposition 18** There is a bijection between \( \mathcal{G}_k(n) \) and \( \mathcal{P}_{n-2k}(n-k+1) \).
Corollary 19 The number of Manhattan polyominoes with \( p \) columns and circumference \( 2r \) is given by

\[
|P_p(r)| = \sum_{j=1}^{r-p-1} N(r-p-1, j) \left( \frac{2(r-j) - p + 1}{2(r-p-1)} \right) = S_{r-p-1}^{(2)}(2r - p - 1),
\]

for \( r \geq p + 1, \ p \geq 0 \).

Reasoning in the same way as in Proposition 18, one can show that the number of all Motzkin paths of length \( n \) whose all plateaus are at least \( l \) steps long, and whose all valleys are at least \( m \) steps wide is

\[ S^{(l,m)}(n) = S^{(l+m)}(n + m). \]

1.6 Concluding remarks

In this paper we have collected various results about secondary structures and pointed out several connections between them and other combinatorial families. In particular, we have used enumerative results obtained in our previous work to give explicit formulas and asymptotic behavior for sequences enumerating various classes of polygon dissections and restricted polyominoes.

References


