involve
a journal of mathematics

Properties of RNA secondary structure matching models

Nicole Anderson, Michael Breunig, Kyle Goryl,
Hannah Lewis, Manda Riehl and McKenzie Scanlan

msp

2019 vol. 12, no. 2

INVOLVE 12:2 (2019)
dx.doi.org/10.2140/involve.2019.12.257

Properties of RNA secondary structure matching models

Nicole Anderson, Michael Breunig, Kyle Goryl,
Hannah Lewis, Manda Riehl and McKenzie Scanlan

(Communicated by Anant Godbole)

RNA secondary structures have been modeled using a variety of matching families. We first explore the intersections of different matching families which are models for likely RNA secondary structures. We then introduce their respective enumeration sequences and prove our proposed equations for enumeration. Next, we prove a formula for the number of matchings with a given crossing number for a variety of matching families. Then we develop a new statistic called the pseudoknot number and find the maximum pseudoknot number on a given set of matchings. We end by providing a comparison between the crossing number, nesting number, and pseudoknot number for three matching families on nine edges.

1. Introduction

1A. ***Background Information.*** Ribonucleic acid (RNA) is an important molecule in all living organisms that directly codes for amino acids and acts as a messenger between DNA and ribosomes to make proteins. RNA is a single-stranded molecule made up of the nucleotides adenine, guanine, cytosine, and uracil. This string of nucleotides folds over on itself and forms secondary bonds between some of its nucleotides [Martinez and Riehl 2017]. Biologists and mathematicians have spent a great deal of time studying these RNA secondary structures [Condon et al. 2004; Jefferson 2015]. In mathematics we represent the secondary structures using a type of graph called a matching. A matching is a graph-theoretical structure in which each vertex has at most degree 1. In our matchings each vertex represents a nucleotide and each edge represents a secondary bond. Since RNA is a string of ordered nucleotides connected by bonds, we list all the vertices of the matching in order along a horizontal line (the backbone) where the edges connect above the vertices. To simplify the graph, we omit any nucleotides that are not part of a

---

MSC2010: 05C30, 92E10.
Keywords: RNA secondary structure, perfect matchings, crossing number, nesting number, pseudoknot.

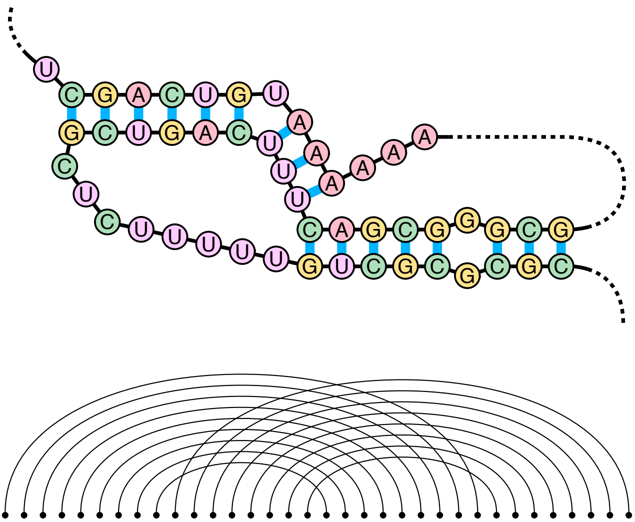

Figure 1. An RNA motif (top) [Ronan 2006] and its matching (bottom).

secondary bond, making each matching a complete matching on $2n$ vertices and n edges. In a complete matching, sometimes called a perfect matching, every vertex is incident to one edge. To recreate an RNA structure with isolated vertices from its matching we can add them into the complete matching in their corresponding spots. Figure 1 demonstrates an RNA motif represented as a matching.

For an edge, i, in a matching, we write $i = (i_1, i_2)$, where i_1, i_2 are the positions of the left- and right-endpoints of the edge, respectively. A pair of edges $i = (i_1, i_2)$ and $j = (j_1, j_2)$ in a matching, M, are said to be crossing if $i_1 < j_1 < i_2 < j_2$. A pair of edges are said to be nested if $i_1 < j_1 < j_2 < i_2$. Similarly, if $j_1 = i_1 + 1$ and $j_2 = i_2 + 1$, then i and j form a hairpin. Finally, a nested sequence of n edges is called a ladder [Martinez and Riehl 2017]. In Section 4 we will be particularly interested in examining pseudoknot numbers for certain matching families along with all perfect matchings.

A number of statistics can be applied to the matchings (e.g., crossing number and nesting number). The crossing number, $\mathrm{cr}(M)$, counts the number of times a pair of edges in the matching cross. The nesting number for one edge counts the number of edges nested under said edge. The nesting number statistic for a matching, $\mathrm{ne}(M)$, is the sum of the nesting numbers for each edge in the matching.

We introduce here a biologically motivated statistic called the pseudoknot number, $\mathrm{pknot}(M)$. A pseudoknot occurs in a strand of RNA when the strand folds on itself and forms secondary bonds between nucleotides, and then the same strand wraps around and forms secondary bonds again. See Figure 1 for an example. These formations are important because viruses have pseudoknots to infiltrate a host cell

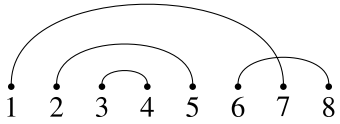

Figure 2. A matching on four edges that can be represented as the permutation $\sigma = 17253468$.

to cause such diseases as hepatitis C, HIV, and SARS [Brierley et al. 2007]. The pseudoknot number of a matching, $\mathrm{pknot}(M)$, counts the number of pseudoknots on the RNA motif by deflating any ladders in the matching and then finding the crossing number on the resulting matching.

Matchings can also be represented as permutations of length $2n$. The permutation is constructed by numbering the vertices (i.e., the nucleotides along the backbone) from 1 to $2n$. We then work from left to right recording the left vertex number followed by the right vertex number of each edge that we encounter. This creates a permutation in which the odd indices are monotonically increasing. For example, the matching in Figure 2, written as a permutation, is 17253468. We represented matchings as permutations in our code that generated data and helped us arrive at conjectures.

Definition 1.1. The *Catalan numbers* are a sequence that appears in mathematics in many different settings. This sequence is enumerated by the formula

$$C_n = \frac{2n}{(n+1)!\,n!},$$

where C_n represents the n-th Catalan number. This sequence begins

$$1, 2, 5, 14, 132, \dots.$$

Example 1.1. Given n pairs of parentheses, there are C_n many ways to arrange these parentheses such that every open parenthesis has a matching closed parenthesis after it. Table 1 shows the ways to arrange n pairs of parentheses where $1 \le n \le 4$.

There is clear bijection between these n pairs of parentheses and the noncrossing matchings on n edges. A left parenthesis corresponds to where an edge starts and a right parenthesis corresponds to where an edge ends. For example, Figure 3 corresponds to the arrangement of five pairs of parentheses given by ()(()())().

Therefore, it is known that the Catalan numbers also count the noncrossing matchings.

In Section 1B we briefly describe families of matchings that represent RNA secondary structures where unlikely biological structures are significantly restricted. Many of these families of matchings are inductively built starting from a single

n	arrangements	no. of arrang.
1	()	1
2	()(), (())	2
3	()()(), ()(()), (())(), (()()), ((()))	5
4	()()()(), (())()(), ()(())(), ()()(()), ()(()()), (())(()), (()())(), ((()))(), ()((())), (()()()), ((())()), (()(())), ((()())), (((())))	14

Table 1. The Catalan-many ways to arrange n pairs of parentheses, where $1 \leq n \leq 4$ [Davis 2016].

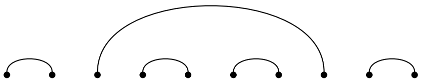

Figure 3. The matching corresponding to the arrangement of five pairs of parentheses given by ()(()())().

edge or a hairpin. Algebraic generating functions for the five main families of matchings have been found [Jefferson 2015]. In Sections 2 and 3 we will investigate intersections of these families.

1B. ***Hairpin-only Families.*** Mathematical biologists have defined five hairpin-only families of matchings that model RNA secondary structures. These families (often named after the people who first studied them) are: LHF (largest hairpin family), D&P (Dirks and Pierce), R&G (Rieder and Giegerich), C&C (Cao and Chen), and L&P (Lyngsø and Pedersen) (shown in Figure 4 from least restrictive to most restrictive) [Jefferson 2015].

Suppose we have two matchings M and N with $2m$ and $2n$ vertices respectively. For a vertex i in M, the *vertex insertion in M by N after i* is the matching with

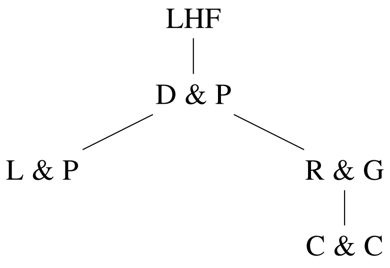

Figure 4. The inclusions among hairpin-only families.

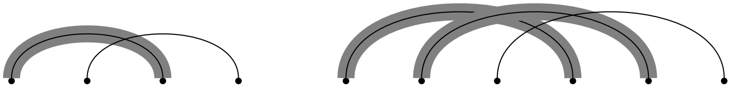

Figure 5. A hairpin with the left edge inflated by a hairpin.

$2m+2n$ vertices in which a copy of N is inserted between i and $i+1$ in M [Jefferson 2015]. Now, given any matching Q with q edges and a matching R with r edges, if (i, j) is an edge of Q then the *edge inflation of Q by R* at (i, j) is the matching on the vertices $2q+2r-2$ in which the vertex i is replaced by the left-endpoints of R, the vertex j is replaced by the right-endpoints of R, and the edges between these two sets of vertices are defined as they are in R [Jefferson 2015]. An example of edge inflation by a matching is shown in Figure 5.

LHF matchings. The *largest hairpin family* is formed by matchings that begin with either a single edge or a hairpin. Then arbitrary vertex insertions and edge inflations by matchings may be performed.

An example of a matching in LHF is shown on the right-hand side of Figure 5. To construct this matching we start with a hairpin and inflate one of its edges by another hairpin.

The D&P (Dirks and Pierce) family. The *D&P family* can be constructed inductively by beginning with either a single edge or a hairpin. Then the following procedures can be performed: (1) inflate an edge by a ladder and (2) insert one D&P matching into another. The D&P family differs from LHF because instead of being closed under edge inflation, only edge inflations by ladders are allowed. An example of a D&P matching is shown in Figure 6. To construct this matching we start with a hairpin and then inflate its left edge by a ladder. Then we insert a hairpin between the two edges of the ladder.

The R&G (Rieder and Giegerich) family. Starting from either a single edge or a hairpin, matchings of the *R&G family* can be built inductively via the following operations: (1) insert R&G matchings into the places shown in Figure 7, followed by (2) inflating the original edge(s) by ladders. An example of an R&G matching is shown in Figure 8. To construct this matching we start with a hairpin and insert another hairpin into the middle.

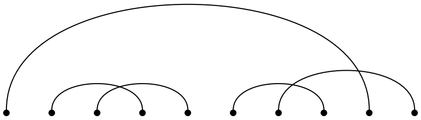

Figure 6. A D&P matching on five edges.

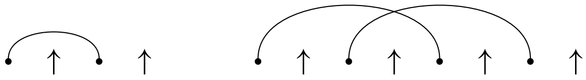

Figure 7. The allowable places for insertion in the R&G family.

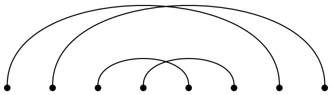

Figure 8. An R&G matching on four edges.

The C&C (*Cao and Chen*) *family.* Starting from either a single edge or a hairpin, matchings of the *C&C family* can be built inductively via the following operations: (1) insert C&C matchings in the allowable places (as shown in Figure 9) and then (2) inflate the original edge(s) by ladders. An example of a C&C matching is shown in Figure 10. To construct this matching we start with a single edge and insert a hairpin under the edge.

The L&P (*Lyngsø and Pedersen*) *family.* Starting from either a single edge or a hairpin, the matchings of the *L&P family* can be built inductively via the following operations: (1) inflate an edge of an L&P matching by a ladder and (2) insert a noncrossing matching into an L&P matching. An example of an L&P matching is shown in Figure 11. To construct this matching we start with a hairpin and insert a noncrossing matching of size 2 into the middle of the hairpin.

We studied different intersections of the hairpin-only families. The Venn diagram in Figure 12 depicts the different intersections in which we have labeled each distinct region using script letters. For example, $\mathcal{A}$ is the group of matchings that are both L&P and C&C. $\mathcal{B}$ is the group of matchings that are L&P and R&G but not C&C. In Sections 2 and 3 we will describe $\mathcal{A}$ and $\mathcal{B}$ and the formulae that enumerate these matchings. In Section 4 we will examine the crossing number, nesting number, and

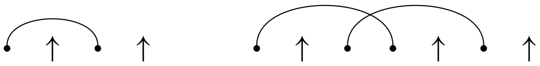

Figure 9. The allowable places for insertion in the C&C family.

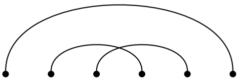

Figure 10. A C&C matching on three edges.

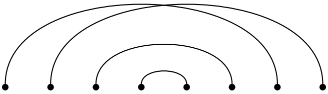

Figure 11. An L&P matching on four edges.

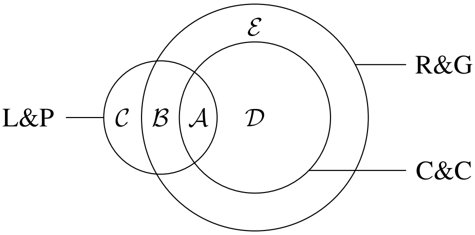

Figure 12. The Venn diagram that shows the different regions we have analyzed.

pseudoknot number statistics on L&P matchings, C&C matchings, and all perfect matchings. We will then compare the distributions of these statistics for each of the given families on nine edges.

2. Matchings that are L&P and C&C (i.e., in $\mathcal{A}$)

Matchings in $\mathcal{A}$ must be both L&P and C&C matchings. Figure 13 shows all 11 matchings in $\mathcal{A}$ with three edges.

For a small n, where n is the number of edges in each matching, the sequence of matchings in $\mathcal{A}$ starting at $n = 1$ is given by, see [Goryl 2017b],

$$1, 3, 11, 39, 134, 456, 1557, 5364, \dots . \tag{1}$$

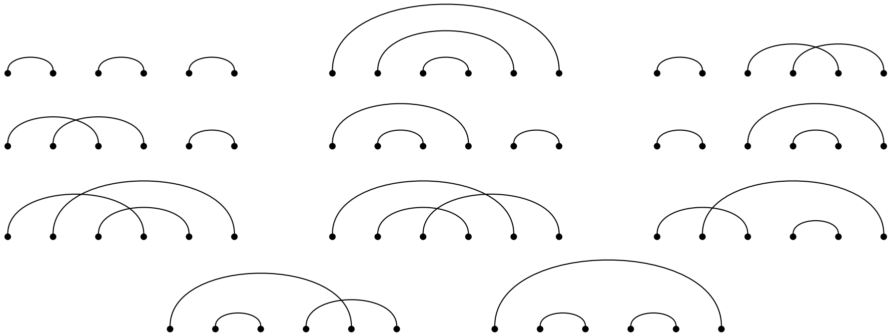

Figure 13. The 11 matchings in $\mathcal{A}$ with three edges.

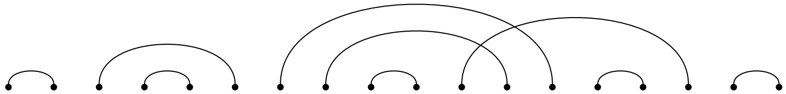

Figure 14. Matching in $\mathcal{A}$ with nine edges.

The matchings that are both L&P and C&C can be created inductively through the following operations in the given order:

- We begin with either a single edge or a hairpin.
- We can insert noncrossing matchings, to ensure the matching is L&P, into the insertion spots as shown in Figure 10, to ensure the matching is C&C.
- Finally, we can inflate the original edge(s) by ladders. Inflating the original edges last ensures that the matchings are C&C as we cannot insert matchings between ladders.

Each matching in $\mathcal{A}$ can be split into two separate parts: a (possibly empty) noncrossing matching and a (possibly empty) hairpin first matching. Let $\mathcal{NCM}_n$ be the noncrossing matchings on n edges and let $\mathrm{NCM}_n = |\mathcal{NCM}_n|$. Let $\mathcal{HF}_n$ be the matchings in $\mathcal{A}_n$ whose leftmost edge is part of a hairpin and let $\mathrm{HF}_n = |\mathcal{HF}_n|$.

For example, in Figure 14, the leftmost edge is part of a noncrossing matching consisting of three edges, so these three edges are in $\mathcal{NCM}_3$. The next part of the matching has its leftmost edge as part of a hairpin. So these six edges are in $\mathcal{HF}_6$.

Lemma 2.1. *The number of ways to insert (possibly empty) noncrossing matches into three possible positions with m total edges is $C_{m+2} - C_{m+1}$.*

Proof. Since the edges we insert are noncrossing and the noncrossing matchings are counted by the Catalan numbers, the summation of these products of three Catalan numbers with subscripts i, j, k would enumerate the number of ways we could insert a total of m edges, made of noncrossing matchings. It can be proved using generating functions [Sloane 2010b] for this sequence that

$$\sum_{\substack{i,j,k\geq 0\\ i+j+k=m}} C_i C_j C_k = C_{m+2} - C_{m+1}. \qquad \square$$

Lemma 2.2. *Matchings in $\mathcal{A}$ where the leftmost edge is part of the hairpin can be enumerated by the formula*

$$HF_n = \sum_{j=1}^{n} C_j - n.$$

Proof. Consider matchings in $\mathcal{A}$ where the leftmost edge is part of a hairpin. Since these matchings are C&C, there are only three insertion spots, and $n-2$ edges left to insert. By Lemma 2.1, there are $C_n - C_{n-1}$ ways to do so. If we have either side inflated by one ladder, it can either inflate the left or right side of the hairpin, and then we have $n-3$ edges left to insert, so that's $2(C_{n-1} - C_{n-2})$ more matchings. If we continue on with this process, the number of ways to insert noncrossing matchings into a hairpin, now taking ladders into consideration, is given by

$$\begin{aligned}&(C_n - C_{n-1}) + 2(C_{n-1} - C_{n-2}) + 3(C_{n-2} - C_{n-3}) + \cdots + (n-1)(C_2 - C_1)\\ &\quad = C_n - C_{n-1} + 2C_{n-1} - 2C_{n-2} + 3C_{n-2} - 3C_{n-3} + \cdots + (n-1)C_2 - (n-1)C_1\\ &\quad = C_n + C_{n-1} + C_{n-2} + \cdots + C_3 + C_2 - (n-1) \quad \text{since } C_1 = 1.\end{aligned}$$

In order to sum over all the Catalan numbers, we can add in C_1 and then subtract $C_1 = 1$ at the end, giving us

$$\sum_{j=1}^{n} C_j - n. \qquad \square$$

Lemma 2.3. *All matchings in $\mathcal{A}$ can be enumerated by*

$$\sum_{i=0}^{n} C_i \,\mathrm{HF}_{n-i},$$

with initial conditions

$$\mathrm{HF}_0 = 1, \quad \mathrm{HF}_1 = 0, \quad \mathrm{HF}_2 = 1, \quad C_0 = 1, \quad C_1 = 1,$$

where C_i are the Catalan numbers, and HF_n *represents the number of matchings that are both L&P, C&C, and have the leftmost edge as part of a hairpin.*

Proof. Matchings in $\mathcal{A}$ are completely described by all combinations of NCM_i and HF_{n-i}. Since we know that the Catalan numbers enumerate the noncrossing matchings, by [Sloane 2010a] we know all matchings in $\mathcal{A}$ are enumerated by

$$\sum_{i=0}^{n} C_i \,\mathrm{HF}_{n-i}\,. \qquad \square$$

Theorem 2.1. *The matchings that are both L&P and C&C can be enumerated by the formula*

$$\sum_{i=0}^{n-2} C_i \left(\sum_{j=1}^{n-i} C_j - (n-i) \right) + C_n,$$

where C_k is the k-th Catalan number.

Proof. From Lemma 2.3 we know the first sum will be

$$\sum_{i=0}^{n} C_i \,\mathrm{HF}_{n-i}\,.$$

Combining this with Lemma 2.2, we see that

$$|\mathcal{A}| = \sum_{i=0}^{n-2} C_i \left(\sum_{j=1}^{n-i} C_j - (n-i) \right) + C_n. \qquad \square$$

Example 2.1. Suppose we want to determine the number of matchings that are L&P and C&C with three edges.

Let $n = 3$. So,

$$\begin{aligned} \sum_{i=0}^{n-2} C_i \left(\sum_{j=1}^{n-i} C_j - (n-i) \right) + C_n &= \sum_{i=0}^{1} C_i \left(\sum_{j=1}^{3-i} C_j - (3-i) \right) + C_3 \\ &= C_0 \left(\sum_{j=1}^{3} C_j - 3 \right) + C_1 \left(\sum_{j=1}^{2} C_j - 2 \right) + C_3 \\ &= C_0(C_1 + C_2 + C_3 - 3) + C_1(C_1 + C_2 - 2) + C_3 \\ &= 1(1+2+5-3) + 1(1+2-2) + 5 = 11, \end{aligned}$$

which agrees with formula (1). The 11 matchings with three edges that are L&P and C&C are shown in Figure 13.

3. Matchings that are L&P and R&G but not C&C (i.e., in $\mathcal{B}$)

Matchings in $\mathcal{B}$ must be both L&P and R&G but not C&C. Figure 15 shows all eight matchings in $\mathcal{B}$ with four edges.

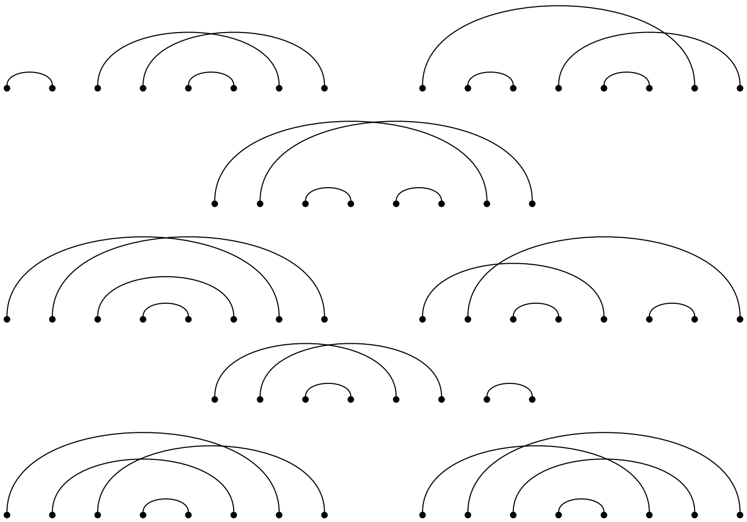

Figure 15. The eight matchings in $\mathcal{B}$ with 4 edges.

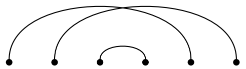

Figure 16. The smallest matching in $\mathcal{B}$.

For a small n, where n is the number of edges in each matching, the sequence of matchings in $\mathcal{B}$ starting at $n = 1$ is given by, see [Goryl 2017a],

$$0, 0, 1, 8, 42, 186, 759, 2970, 11369, \dots. \tag{2}$$

The group of matchings that are L&P and R&G but not C&C can be created inductively through the following operations in the given order:

- We must begin with a hairpin that has a single edge inserted into the middle of the hairpin as shown in Figure 16. The edge inserted into the middle of the hairpin will ensure that all matchings are not C&C.
- To ensure that the matchings are L&P, we next insert noncrossing matchings.
- Finally, we can inflate the edges of the hairpin by ladders. Inflating the edges of the hairpin last ensures that the matchings are R&G, in which you cannot insert matchings between ladders.

Lemma 3.1. *Let a_n be the number of matchings in $\mathcal{B}$ with n edges where the last step in its inductive creation is inflation by a ladder*, *denoted by* inflate(n). *Then* $a_n = 2a_{n-1} - a_{n-2}$.

Proof. In order for a matching to be R&G, inflations of the original edges must occur after all insertions. Since matchings in $\mathcal{B}$ start with a hairpin, the only edges we can inflate are the two edges that make up the hairpin. To obtain these matchings with n edges, we can take all matchings in $\mathcal{B}$ with $n-1$ edges and inflate each matching's left edge of the hairpin by a ladder. That gives us a_{n-1} matchings. Then, for the matchings in $\mathcal{B}$ with $n-1$ edges, we can inflate each matching's right edge of the hairpin by a ladder, which gives us another a_{n-1} matchings. In total, this gives us $2a_{n-1}$ matchings, but we have over-counted. Inflating the left side of the hairpin and then the right is the same as inflating the right side of the hairpin and then the left, so we want to remove those matchings that had one edge added to each half of the hairpin. If we take all the matchings with $n-2$ edges in $\mathcal{B}$ and inflate the left edge of the hairpin by one ladder and then right edge of the hairpin by one ladder, we get a_{n-2} matchings with n edges. So, if we subtract the over-counted matchings, we get $2a_{n-1} - a_{n-2}$ matchings in inflate(n). □

Lemma 3.2. *Matchings with n edges in $\mathcal{B}$ where the last step is an insertion*, *denoted by* ins(n), *are enumerated by*

$$\frac{6}{n+3}\binom{2n-1}{n-3}. \tag{3}$$

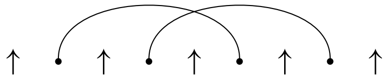

Figure 17. Five allowable places for insertion in $\mathcal{B}$.

Proof. Starting with a hairpin, we can insert noncrossing matchings into the five different insertion spots shown in Figure 17 given there is at least one element of NCM_1 in the middle of the hairpin. We know that the inserted noncrossing matchings cannot be inserted between ladders because the matchings must be both R&G and L&P. Note, for R&G matchings there are only four insertion spots because we cannot insert to the left of the hairpin. Noncrossing matchings inserted on the left of the hairpin can be thought of as starting with a noncrossing matching and inserting a matching containing a hairpin on the right. Therefore, we will be inserting noncrossing matchings into five insertion spots. Letting $i, j, k \geq 1$, we insert $i-1$ edges on the left side of the hairpin, j edges in the middle of the hairpin (since we must have at least one edge inserted here), and $k-1$ edges on the right side of the hairpin. Since there are already two edges included in the hairpin, $i-1+j+k-1=n-2$, which implies $i+j+k=n$.

First, consider noncrossing matchings that are on the left side of the hairpin. This includes both matchings that are on the far left outside of the hairpin and matchings that are under the left edge of the hairpin. There are $C_{i-1}C_0+C_{i-2}C_1+\cdots+C_1C_{i-2}+C_0C_{n-1}$ ways to insert $i-1$ edges into two insertion spots on the left side of the hairpin. It is known that sum is equal to C_i [Davis 2016]. Therefore, there are C_i ways to insert $i-1$ edges into the left side of the hairpin. Likewise, there are C_k ways to insert $k-1$ edges into the right side of the hairpin, which includes on the far right outside of the hairpin and matchings under the right edge of the hairpin. There are C_j ways to insert j edges into the middle of a hairpin because we can only insert noncrossing matchings. This implies we can enumerate matchings in $\operatorname{ins}(n)$ by

$$\sum_{\substack{i,j,k\geq 1\\ i+j+k=n}} C_iC_jC_k$$

It is known this is given by the closed form (3) [Sloane 1991]. □

Theorem 3.1. *The matchings that are L&P and R&G but not C&C can be enumerated by the formula*

$$a_n = 2a_{n-1}-a_{n-2}+\frac{6}{n+3}\binom{2n-1}{n-3}. \tag{4}$$

Proof. Since we inductively build matchings in $\mathcal{B}$ with exactly two operations, Lemmas 3.1 and 3.2 completely describe how to build all matchings in $\mathcal{B}$. Therefore, by Lemmas 3.1 and 3.2, we can add our two formulas together to obtain the recurrence relation (4), which enumerates all matchings in $\mathcal{B}$. □

This recurrence relation has no closed formula, although it can be written using hypergeometric series. Since there is a closed form for L&P matchings and we have now found equations for $\mathcal{A}$ and $\mathcal{B}$, we can use these to enumerate matchings in $\mathcal{C}$ using subtraction.

Example 3.1. Suppose we want to find the matchings in $\mathcal{B}$ with $n = 4$ edges. Then we can compute

$$a_4 = 2a_3 - a_2 + \tfrac{6}{7}\tbinom{7}{1} = 2(1) - 0 + 7\tfrac{6}{7} = 8,$$

which agrees with formula (2). The eight matchings with four edges in $\mathcal{B}$ can be found in Figure 15.

4. Statistics

We implemented the following statistics and constructions of each hairpin family in Python, and those programs are available upon request.

4A. ***L&P Matchings.*** L&P matchings insert only noncrossing matchings; therefore the only way to have two crossing edges is to begin with a hairpin. The only way to increase the crossing number is to inflate the edges of the hairpin. Because all of the crossings happen in the hairpin, we can count the number of crossings by taking the product of the number of edges that inflate the left side of the hairpin and the number of edges that inflate the right side of the hairpin.

Theorem 4.1. *The number of L&P matchings on n edges with a given crossing number $h > 0$ is*

$$\sum_{\substack{(\gamma_1,\gamma_2,\dots,\gamma_k)\\ \gamma_i \geq 0\\ \gamma_i \leq \beta_i \forall i}} \; \sum_{\substack{(\alpha_1,\alpha_2,\dots,\alpha_{(2E+1)})\\ \alpha_i \geq 0\\ \sum \alpha_i = n-E}} \; \prod_{i=1}^{2E+1} C_{\alpha_i},$$

where $h = p_1^{\beta_1} p_2^{\beta_2} \cdots p_k^{\beta_k}$ is the prime factorization of h, $p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k}$ is a factor of h, and E is $p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k} + p_1^{\beta_1-\gamma_1} p_2^{\beta_2-\gamma_2} \cdots p_k^{\beta_k-\gamma_k}$.

Proof. Let h be the desired crossing number and n be the number of edges in the L&P matching. Because $h > 0$ and the matchings are L&P we know that there is exactly 1 hairpin in the matching. Then let $h = p_1^{\beta_1} p_2^{\beta_2} \cdots p_k^{\beta_k}$ be its unique prime factorization. Each factor of h is then given by $p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k}$ with $0 \leq \gamma_i \leq \beta_i$. The product of the edges on the right and the left sides of the hairpin must be h so there must be $p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k}$ edges on the left side of the hairpin and $p_1^{\beta_1-\gamma_1} p_2^{\beta_2-\gamma_2} \cdots p_k^{\beta_k-\gamma_k}$ edges on the right side of the hairpin. Then there are $n - E$ edges left to insert into the matching in some of $2E + 1$ places. Then consider the number of noncrossing edges inserted into each spot of the matching as a vector $\vec{\alpha}$. If each α_i will be the number of edges inserted into the i-th spot of

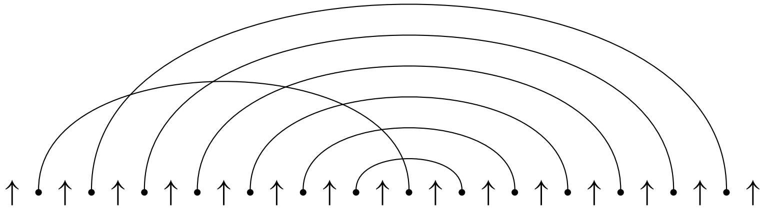

Figure 18. Places to insert in an L&P matching.

the matching, the insertion vector $\vec{\alpha}$ is given by $(\alpha_1, \alpha_2, \ldots, \alpha_{2E+1})$, where $\alpha_i \geq 0$ and $\sum_i \alpha_i = n - E$. The insertions must be noncrossing because the matchings are L&P; therefore there are Catalan-many ways to insert each noncrossing matching. Then there are $C_{\alpha_1} C_{\alpha_2} \cdots C_{\alpha_{2E+1}}$ ways to insert each $\vec{\alpha}$ into the matching. We can sum over all $\vec{\alpha}$ to find all the matchings with the given number of edges on the left and right sides of the hairpin. Then we sum over all the possible factors of h to get all possible L&P matchings on n edges with crossing number h. □

For example, to find the number of L&P matchings on $n = 9$ edges with a crossing number of $h = 6$ we will start by writing the prime factorization of h: $h = 2^1 * 3^1$. Using the prime factorization we can find all of the factors of h: 1, 2, 3, 6. We begin with our first factor, 1, being the number of edges on the left side of the hairpin. We know the product of the number of edges on the left and right sides of the hairpin must be 6, so there must be six edges on the right side of the hairpin. Then there are a total of seven edges in the hairpin, and there are two edges left to insert into the matching in the 15 places shown by the arrows in Figure 18. There are two edges left to insert into the inflated hairpin and they could be inserted into the same place or into two different places. If they are inserted into two different places the corresponding $\vec{\alpha}$ would contain two 1's, and 13 0's. There are $\binom{15}{2}$ distinct $\vec{\alpha}$ of this type. There are C_1 ways to arrange the first inserted edge and there are C_1 ways to arrange the second inserted edge. If both edges are inserted into the same place the corresponding $\vec{\alpha}$ would contain one 2 and 14 0's. There are 15 distinct $\vec{\alpha}$ of this type. There are C_2 ways to arrange the inserted noncrossing matching of size 2. The number of L&P matchings on nine edges with one edge on the left side of the hairpin and six edges on the right side of the hairpin is therefore $\binom{15}{2} C_1 C_1 + 15 C_2 = 135$. We repeat this process with each remaining factors of h, 2, 3, and 6, where each factor is the number of edges on the left side of the hairpin. After considering all factors we find that there are 4758 L&P matchings on nine edges with a crossing number of six.

Example 4.1. Suppose we want to find the number of L&P matchings with $n = 4$ edges and $h = 1$.

We start with the prime factorization of h: $h = 1$. Thus, the only factor of h is 1. This means there is one edge on each of the left and right sides of the hairpin. Hence,

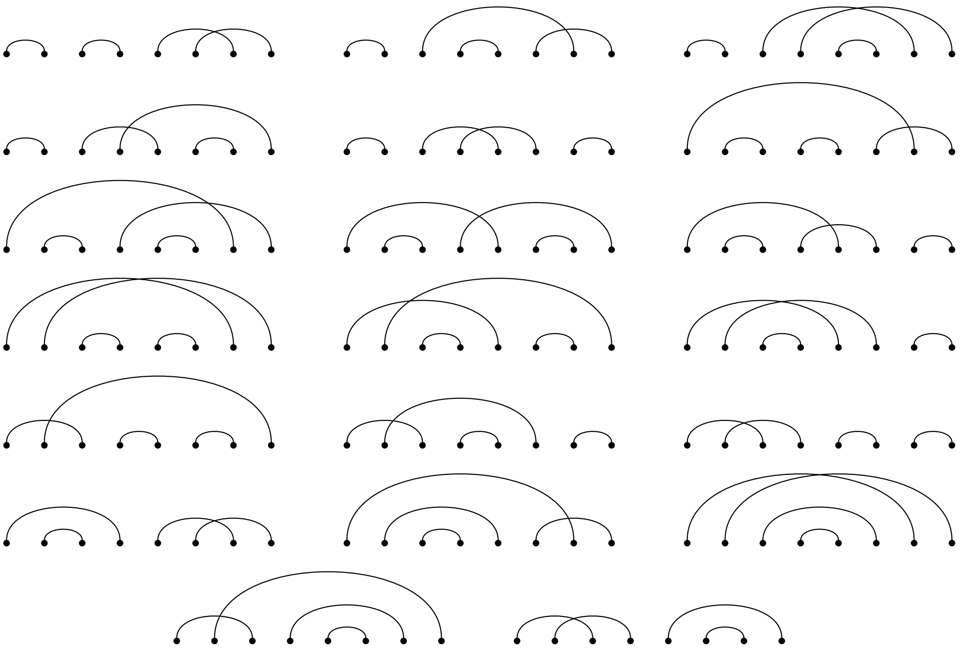

Figure 19. The 20 matchings that are L&P with $n = 4$ edges and $h = 1$.

there are two edges left to insert into five spots. Using the same process as before, we get $\binom{5}{2}C_1C_1 + 5C_2 = 20$ matchings. These 20 matchings are found in Figure 19.

4A1. *Maximum values.* To find the maximum crossing number of an L&P matching on n edges, we require that all edges be part of the initial hairpin and that there be as close to an equal number edges inflated on each side of the hairpin as possible. This results in a maximum crossing number of $\frac{1}{8}(2n^2 - 1 + (-1)^n)$, which is the closed form of $\sum \lfloor \frac{1}{2}n \rfloor + 1$, which is obtained by successively adding one edge to the previously maximally inflated hairpin. We note that the formula in Theorem 4.1 is zero if $h > \sum \lfloor \frac{1}{2}n + 1 \rfloor$.

The maximum nesting number of an L&P matching is the same as that of a noncrossing matching (i.e., a ladder on n edges), which is $\frac{1}{2}n(n + 1)$.

4B. ***C&C.*** The maximum crossing and nesting number of a C&C matching of n edges is the same as that of an L&P matching. This is because the C&C matchings contain the noncrossing matchings (i.e., the maximally nested matchings) and the matchings with the maximum crossing number are L&P and C&C.

4C. $\mathcal{A}$, *matchings that are L&P and C&C.*

Lemma 4.1. *If a C&C matching that contains exactly one hairpin is inserted to the right of a noncrossing matching, the obtained matching is C&C.*

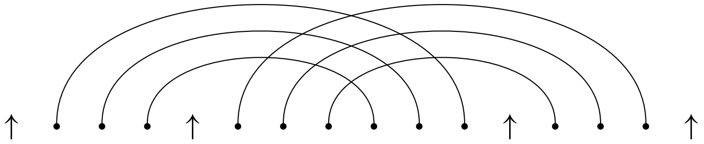

Figure 20. Allowable insertion spots for $\mathcal{A}$.

Proof. A noncrossing matching is C&C, and according to the inductive definition of C&C matchings we can insert another C&C matching to the right of it. □

Lemma 4.2. *There are four places to insert noncrossing matchings in a hairpin to obtain a matching in $\mathcal{A}$. Matchings cannot be inserted between ladders of the hairpin, and they cannot be inserted into the middle of the hairpin.*

Proof. By Lemma 4.1 if we insert a noncrossing matching to the far left of a C&C matching that contains exactly one possibly inflated hairpin, the obtained matching is C&C. By the rules of creating C&C matchings, if we insert a noncrossing matching into the spots shown by the three remaining arrows in Figure 20 it will still be C&C. We cannot insert noncrossing matchings in between ladders or in the middle of the hairpin because then it would no longer be C&C. Because there is only one possibly inflated hairpin and we are inserting only noncrossing matchings we are guaranteed that these matchings are also L&P, and therefore they are in $\mathcal{A}$. □

Theorem 4.2. *The number of matchings in $\mathcal{A}$ on n edges with crossing number $h > 0$ is given by*

$$\sum_{\substack{(\gamma_1,\gamma_2,\dots,\gamma_k)\\ \gamma_i\geq 0\\ \gamma_i\leq\beta_i\forall i}} \sum_{\substack{(\alpha_1,\alpha_2,\alpha_3,\alpha_4)\\ \alpha_i\geq 0\\ \sum\alpha_i=n-E}} \prod_{i=1}^{4} C_{\alpha_i},$$

where $h = p_1^{\beta_1} p_2^{\beta_2} \cdots p_k^{\beta_k}$ is the prime factorization of h, $p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k}$ is a factor of h, and E is $p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k} + p_1^{\beta_1-\gamma_1} p_2^{\beta_2-\gamma_2} \cdots p_k^{\beta_k-\gamma_k}$.

Proof. Let h be the given crossing number and n be the number of edges in the matching in $\mathcal{A}$. Because $h > 0$ and the matchings are L&P we know that there is exactly one possibly inflated hairpin in the matching. Then let $h = p_1^{\beta_1} p_2^{\beta_2} \cdots p_k^{\beta_k}$ be its unique prime factorization. Each factor of h is then given by $p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k}$. Since these matchings are L&P, the product of the edges on the right and the left sides of the hairpin must be h. Thus, there must be $p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k}$ edges on the left side of the hairpin and $p_1^{\beta_1-\gamma_1} p_2^{\beta_2-\gamma_2} \cdots p_k^{\beta_k-\gamma_k}$ edges on the right side of the hairpin. Then the total number of edges E in the hairpin is given by $p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k} + p_1^{\beta_1-\gamma_1} p_2^{\beta_2-\gamma_2} \cdots p_k^{\beta_k-\gamma_k}$. Then there are $n - E$ edges left to insert into the matching. By Lemma 4.2 there are four places to insert. The insertions must be noncrossing because the matchings are L&P; therefore there are Catalan-many ways to insert

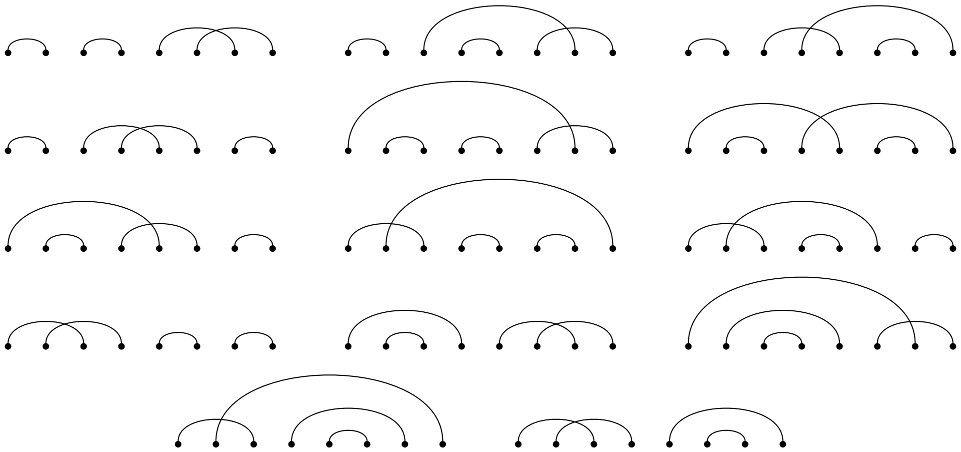

Figure 21. The 14 matchings in $\mathcal{A}$ with $n = 4$ edges and $h = 1$.

each noncrossing matching. If we think about the number of noncrossing edges inserted into each spot of the matching as a vector $\vec{\alpha}$ then each α_i will be the number of edges inserted into the i-th spot of the matching. The insertion vector $\vec{\alpha}$ is given by $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$, where $\alpha_i \geq 0$ and $\sum_i \alpha_i = n - E$. Then there are $C_{\alpha_1} C_{\alpha_2} C_{\alpha_3} C_{\alpha_4}$ ways to insert each $\vec{\alpha}$ into the matching. We can then sum over all $\vec{\alpha}$ to find all the matchings with the given number of edges on the left and right sides of the hairpin. Then we sum over all the possible factors of h to get all possible matchings in $\mathcal{A}$ on n edges with crossing number h. □

Example 4.2. Suppose we want to find the number of matchings in $\mathcal{A}$ with $n = 4$ edges and $h = 1$.

We start with the prime factorization of h: $h = 1$. Thus, the only factor of h is 1. This means there is one edge on each of the left and right sides of the hairpin. Hence, there are two edges left to insert into four spots. Using the same process as before, we get $\binom{4}{2} C_1 C_1 + 4C_2 = 14$ matchings. These 14 matchings are found in Figure 21.

4D. *$\mathcal{B}$, matchings that are R&G and L&P but not C&C.*

Lemma 4.3. *There are six places to insert noncrossing matchings in an inflated hairpin that has one edge inserted into the middle in order for a matching in $\mathcal{B}$. These places are shown in Figure 22.*

Proof. By Lemma 4.1 if we insert a noncrossing matching to the far left of a C&C matching that contains exactly one hairpin, the obtained matching is C&C. Therefore if we insert a noncrossing matching to the far left of an R&G matching that contains exactly one hairpin, then the obtained matching is R&G. We can't insert matchings between the ladders that inflate the hairpin because then the matching would not

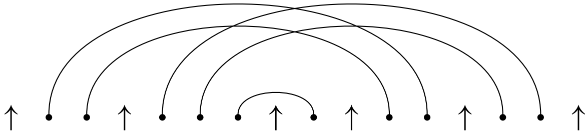

Figure 22. Allowable insertion spots for $\mathcal{B}$.

be R&G. The obtained matching is also L&P because it has exactly one hairpin, and only noncrossing matchings are inserted. The obtained matching is not C&C because of the guaranteed single edge inserted into the middle of the hairpin. We don't insert to the left of the single edge in the middle of the hairpin in order to prevent duplicates from being made. □

Theorem 4.3. *The number of matchings in $\mathcal{B}$ on n edges with crossing number $h > 0$ is given by the formula*

$$\sum_{\substack{(\gamma_1,\gamma_2,\ldots,\gamma_k)\\ \gamma_i\geq 0\\ \gamma_i\leq\beta_i\forall i}} \sum_{\substack{(\alpha_1,\alpha_2,\ldots,\alpha_6)\\ \alpha_i\geq 0\\ \sum\alpha_i=n-E-1}} \prod_{i=1}^{6} C_{\alpha_i},$$

where $h = p_1^{\beta_1} p_2^{\beta_2} \cdots p_k^{\beta_k}$ is the prime factorization of h, $p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k}$ is a factor of h, and E is $p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k} + p_1^{\beta_1-\gamma_1} p_2^{\beta_2-\gamma_2} \cdots p_k^{\beta_k-\gamma_k}$.

Proof. Let h be the given crossing number and n be the number of edges in the matching in $\mathcal{B}$. Because $h > 0$ and the matchings are L&P we know that there is exactly one possibly inflated hairpin in the matching. Then let $h = p_1^{\beta_1} p_2^{\beta_2} \cdots p_k^{\beta_k}$ be its unique prime factorization. Each factor of h is then given by $p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k}$. Since these matchings are L&P, the product of the edges on the right and the left sides of the hairpin must be h. Thus, there must be $p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k}$ edges on the left side of the hairpin and $p_1^{\beta_1-\gamma_1} p_2^{\beta_2-\gamma_2} \cdots p_k^{\beta_k-\gamma_k}$ edges on the right side of the hairpin. Then the total number of edges E in the hairpin is given by $p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k} + p_1^{\beta_1-\gamma_1} p_2^{\beta_2-\gamma_2} \cdots p_k^{\beta_k-\gamma_k}$. Then there are $n - E - 1$ edges left to insert into the matching. We subtract 1 since a single edge is guaranteed to be inserted into the middle of the hairpin. By Lemma 4.3 there are six places to insert. The insertions must be noncrossing because the matchings are L&P. Therefore there are Catalan-many ways to insert each noncrossing matching. If we think about the number of noncrossing edges inserted into each spot of the matching as a vector $\vec{\alpha}$, then each α_i will be the number of edges inserted into the i-th spot of the matching. The insertion vector $\vec{\alpha}$ is given by $(\alpha_1, \alpha_2, \ldots, \alpha_6)$, where $\alpha_i \geq 0$ and $\sum_i \alpha_i = n - E - 1$. Then there are $C_{\alpha_1} C_{\alpha_2} \cdots C_{\alpha_6}$ ways to insert each $\vec{\alpha}$ into the matching. We can then sum over all $\vec{\alpha}$ to find all the matchings with the given number of edges on the left and right sides of the hairpin. Then we sum over all the possible factors of h to get all possible matchings in $\mathcal{B}$ on n edges with crossing number h. □

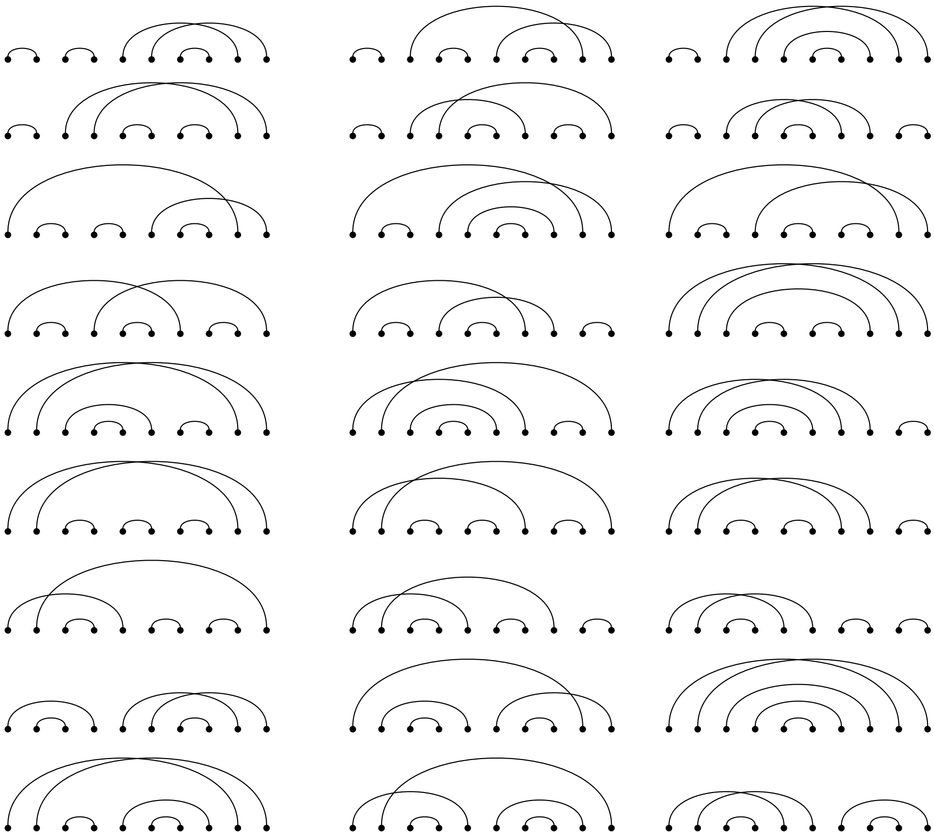

Figure 23. The 27 matchings in $\mathcal{B}$ with $n = 5$ edges and $h = 1$.

Example 4.3. Suppose we want to find the number of matchings in $\mathcal{B}$ with $n = 5$ edges and $h = 1$.

We start with the prime factorization of h: $h = 1$. Thus, the only factor of h is 1. This means there is one edge on each of the left and right sides of the hairpin. Hence, there are two edges left to insert into four spots. Using the same process as before, we get $\binom{6}{2}C_1C_1 + 6C_2 = 27$ matchings. These 27 matchings are found in Figure 23.

4E. ***Maximum Pseudoknot.*** The pseudoknot number of a matching is found by deflating the ladders of every edge in the matching and finding the crossing number of the new matching. To find the maximum pseudoknot number of a matching on n edges, we want a matching that has the highest crossing number once its edges have been deflated.

Let M_n be the set of all perfect matchings on n edges. Let CC_n be the set of all C&C matchings on n edges. For a set of matchings A, let $\mathrm{pknot}_{\max}(A) = \max\{\mathrm{pknot}(M) \mid M \in A\}$.

Figure 24. Matching that gives the maximum pseudoknot number on four edges for the C&C family.

Definition 4.1. We define a deflation as follows. If we have two edges $i = (i_1, i_2)$ and $j = (j_1, j_2)$ such that $i_1 = j_1 - 1$ and $i_2 = j_2 + 1$, then we delete edge i.

Theorem 4.4. *The maximum pseudoknot number of the set of C&C matchings on n edges is given by*

$$\mathrm{pknot}_{\max}(\mathrm{CC}_n) = \begin{cases} \frac{1}{2}n, & n \in \mathbb{N} \text{ and is even}, \\ \frac{1}{2}(n-1), & n \in \mathbb{N} \text{ and is odd}. \end{cases}$$

Proof. In C&C matchings, the maximum pseudoknot number occurs when we maximize the number of noninflated hairpins since we cannot insert between ladders in C&C matchings.

Case 1: n is even. Since n is even, every two edges must be involved in distinct hairpins in the matching in order to maximize the number of noninflated hairpins. Then every matching must have $\frac{1}{2}n$ hairpins. Because our matching M consists of strictly hairpins, there are no deflations. Hence, $\mathrm{pknot}(M) = \mathrm{cr}(M)$. Since all the hairpins are not inflated, to find $\mathrm{cr}(M)$ we can count the number of hairpins. Thus $\mathrm{cr}(M) = \frac{1}{2}n = \mathrm{pknot}_{\max}(\mathrm{CC}_n)$.

Case 2: n is odd. When n is odd, every two edges except for one must be involved in distinct hairpins in order to maximize the number of noninflated hairpins. Then the matching M must have $\frac{1}{2}(n-1)$ hairpins. Whether the final edge not involved in a hairpin is inserted or inflates another edge has no effect on the pseudoknot number. After deflations on all edges are made, there are still guaranteed to be $\frac{1}{2}(n-1)$ hairpins, and $\mathrm{pknot}(M)$ is the crossing number of the deflated matching, which is the number of hairpins. Thus $\mathrm{pknot}_{\max}(\mathrm{CC}_n) = \frac{1}{2}(n-1)$. □

Example 4.4. Suppose we want to calculate the maximum pseudoknot numbers for the C&C families with four and five edges.

First, let $n = 4$. According to Theorem 4.4, $\mathrm{pknot}_{\max}(M) = 2$. Figure 24 shows the matching with this pseudoknot number.

Now, let $n = 5$. By Theorem 4.4, the maximum pseudoknot will be just like $n = 4$. With the extra edge, we can either insert it into the allowable space or inflate the original matching. If we conduct an insertion, we will not increase the crossing number, unlike inflating. However, this one edge will become deflated. Figure 25 shows matchings which produce a maximum pseudoknot number of 2 on $n = 5$ edges.

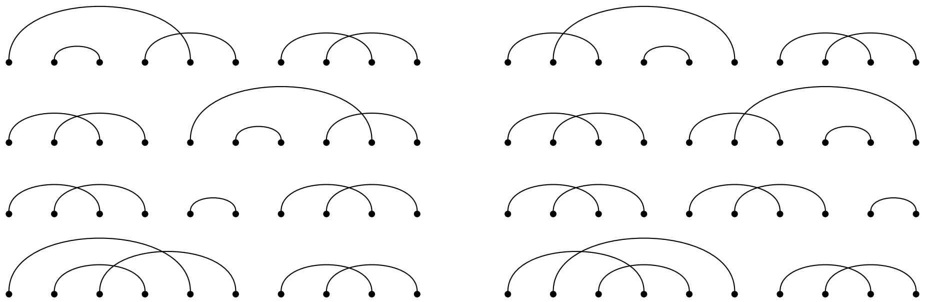

Figure 25. Matchings that give the maximum pseudoknot number on five edges for the C&C family.

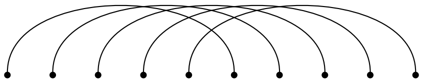

Figure 26. Matching that gives the maximum pseudoknot number on five edges for all perfect matchings.

Theorem 4.5. *The maximum pseudoknot number of the set of all perfect matchings on n edges is given by*

$$\operatorname{pknot}_{\max}(M_n) = \tfrac{1}{2}n(n-1)$$

Proof. The only restrictions we have on all perfect matchings are that an edge cannot cross itself or go below the horizon. In order to maximize $\operatorname{pknot}(M_n)$ in a perfect matching, none of the edges should be involved in ladders, and each edge should cross every edge in the matching once. The way to maximize $\operatorname{pknot}(M_n)$ for any given n is to use the matching on $n-1$ edges with the maximum pseudoknot number, and cross all of those edges with the new edge we add. Thus, we can represent this as the recurrence relation $\operatorname{pknot}_{\max}(M_n) = \operatorname{pknot}_{\max}(M_{n-1}) + (n-1)$ with the initial condition $\operatorname{pknot}_{\max}(M_2) = 1$ and where $(n-1)$ represents how many edges our new edge will cross. Solving this recurrence relation gives $\frac{1}{2}n(n-1)$. □

Example 4.5. Suppose we want to find $\operatorname{pknot}_{\max}(M_n)$ for all perfect matchings on $n = 5$ edges.

According to Theorem 4.5, $\operatorname{pknot}_{\max}(M_5) = \frac{5(4)}{2} = 10$. Figure 26 shows the matching with the maximum pseudoknot number.

4F. *Comparison.* Pseudoknots are not found as commonly in RNA secondary structures as they are in perfect matchings. One desired result of the inductive definitions of L&P and C&C is that the models more accurately reflect the likelihood of pseudoknots observed in nature. The following graphs in Figures 27, 29, and 28 display the distributions of the crossing number, pseudoknot number, and nesting

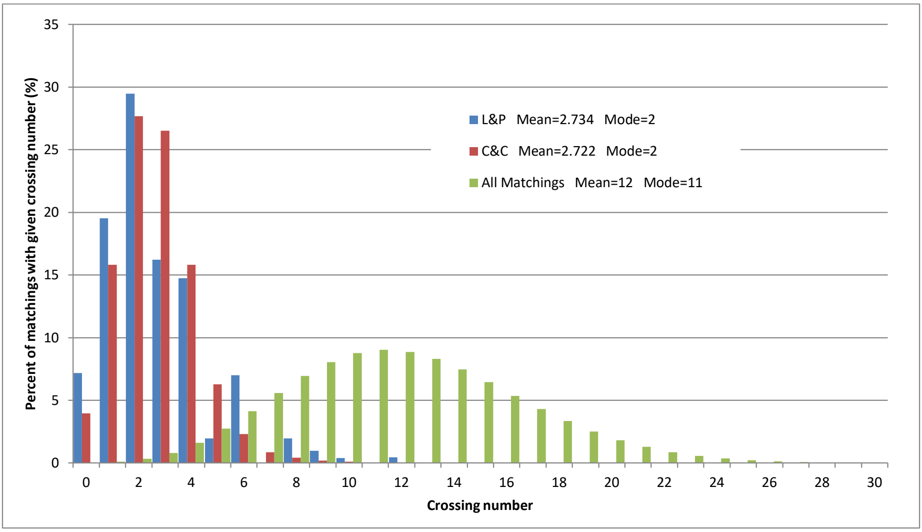

Figure 27. Percent of matchings on nine edges with given crossing number.

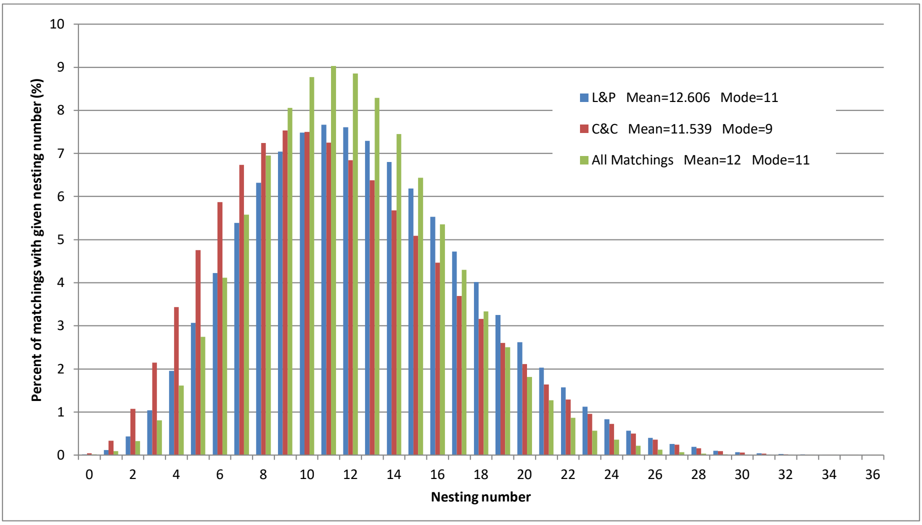

Figure 28. Percent of matchings on nine edges with given nesting number.

number for each given family on nine edges as a percentage of the total matchings in that family. The motivation for this section is to examine how limiting the number of pseudoknots affects the crossing, nesting, and pseudoknot number distribution compared to all perfect matchings.

In all perfect matchings the crossing and nesting number statistics are symmetrically distributed and their joint distribution has been an area of recent study.

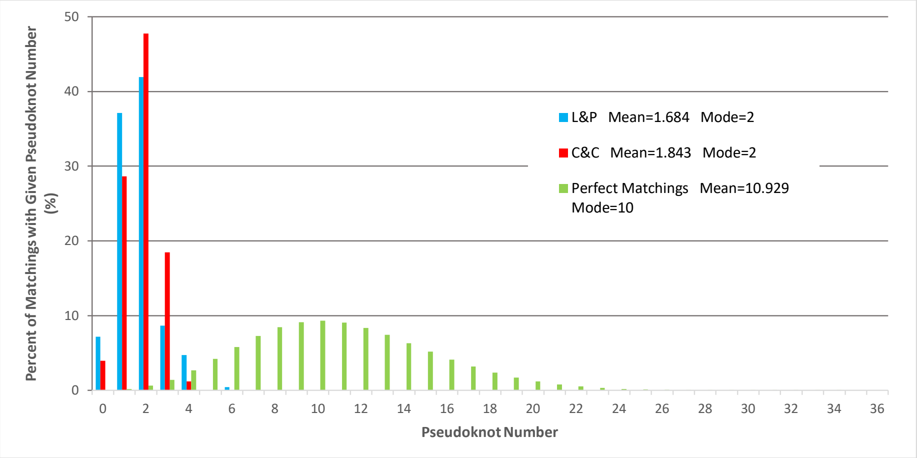

Figure 29. Percent of matchings on nine edges with given pseudoknot number.

Note that in all perfect matchings on n edges, as the crossing number increases, the nesting number decreases, and vice versa. These two statistics are inversely related and symmetrically distributed when considering all perfect matchings [Chen et al. 2007]. However, when we limit the number of pseudoknots by looking at L&P and C&C matchings, the distributions for crossing number and pseudoknot number become skewed right. Because the crossing and nesting number statistics are inversely related when considering all perfect matchings we would predict that the nesting number distribution would become skewed left when considering these L&P and C&C matchings. Surprisingly, the distributions of nesting numbers for L&P and C&C remain largely similar to the distribution of nesting numbers for all matchings. However, recall that these matching families limit the unlikely biological structures, in this case a high number of pseudoknots. Pseudoknots are represented in matchings by a hairpin, ignoring the additional ladders. Therefore these matchings limit the number of hairpins, so we should expect there to be more matchings with a smaller crossing number. These matching families don't limit the ladders that can occur on edges or newly inserted edges in a hairpin, so the nesting number stays relatively symmetrically distributed.

Acknowledgments

Thank you to the National Science Foundation DMS-1451801 Grant for funding the program that enabled this research. Also thanks to the Office of Research and Sponsored Programs at the University of Wisconsin-Eau Claire for providing additional support for student research.

References

[Brierley et al. 2007] I. Brierley, S. Pennell, and R. J. C. Gilbert, "Viral RNA pseudoknots: versatile motifs in gene expression and replication", *Nature Rev. Microbiol.* **5** (2007), 598–610.

[Chen et al. 2007] W. Y. C. Chen, E. Y. P. Deng, R. R. X. Du, R. P. Stanley, and C. H. Yan, "Crossings and nestings of matchings and partitions", *Trans. Amer. Math. Soc.* **359**:4 (2007), 1555–1575. MR Zbl

[Condon et al. 2004] A. Condon, B. Davy, B. Rastegari, S. Zhao, and F. Tarrant, "Classifying RNA pseudoknotted structures", *Theoret. Comput. Sci.* **320**:1 (2004), 35–50. MR Zbl

[Davis 2016] T. Davis, "Catalan numbers", preprint, 2016, available at http://www.geometer.org/mathcircles/catalan.pdf.

[Goryl 2017a] K. Goryl, "A289031: Number of perfect matchings on $n+3$ edges which represent RNA secondary folding structures characterized by the Reeder and Giegerich and the Lyngso and Pedersen families, but not the family characterized by Cao and Chen", entry A289031 in *The on-line encyclopedia of integer sequences* (http://oeis.org), 2017.

[Goryl 2017b] K. Goryl, "A289834: Number of perfect matchings on n edges which represent RNA secondary folding structures characterized by the Lyngso and Pedersen (L&P) family and the Cao and Chen (C&C) family", entry A289834 in *The on-line encyclopedia of integer sequences* (http://oeis.org), 2017.

[Jefferson 2015] A. F. Jefferson, *The substitution decomposition of matchings and RNA secondary structures*, Ph.D. thesis, University of Florida, 2015, available at https://tinyurl.com/jeffephd.

[Martinez and Riehl 2017] M. A. Martinez and M. Riehl, "A bijection between the set of nesting-similarity classes and L&P matchings", *Discrete Math. Theor. Comput. Sci.* **19**:2 (2017), art. id. 1. MR

[Ronan 2006] P. Ronan, "A pseudoknot structure in human telomerase RNA", diagram, 2006, available at https://tinyurl.com/pseuknpic.

[Sloane 1991] N. J. A. Sloane, "A003517: Number of permutations of $[n+1]$ with exactly 1 increasing subsequence of length 3", entry A003517 in *The on-line encyclopedia of integer sequences* (http://oeis.org), 1991.

[Sloane 2010a] N. J. A. Sloane, "A000108: Catalan numbers", entry A000108 in *The on-line encyclopedia of integer sequences* (http://oeis.org), 2010.

[Sloane 2010b] N. J. A. Sloane, "A000245: $a(n) = 3(2n)!/((n+2)!\,(n-1)!)$", entry A000245 in *The on-line encyclopedia of integer sequences* (http://oeis.org), 2010.

Received: 2017-07-22 Revised: 2018-04-03 Accepted: 2018-04-04

andersni9728@uwec.edu *University of Wisconsin, Eau Claire, WI, United States*

breunims@uwec.edu *University of Wisconsin, Eau Claire, WI, United States*

gorylk24@gmail.com *University of Wisconsin, Milwaukee, WI, United States*

lewishk0885@uwec.edu *University of Wisconsin, Eau Claire, WI, United States*

riehl@rose-hulman.edu *Department of Mathematics, Rose-Hulman Institute of Technology, Terre Haute, IN, United States*

scanlamd0729@uwec.edu *University of Wisconsin, Eau Claire, WI, United States*

involve

msp.org/involve

INVOLVE YOUR STUDENTS IN RESEARCH

Involve showcases and encourages high-quality mathematical research involving students from all academic levels. The editorial board consists of mathematical scientists committed to nurturing student participation in research. Bridging the gap between the extremes of purely undergraduate research journals and mainstream research journals, *Involve* provides a venue to mathematicians wishing to encourage the creative involvement of students.

MANAGING EDITOR

Kenneth S. Berenhaut Wake Forest University, USA

BOARD OF EDITORS

Colin Adams	Williams College, USA
Arthur T. Benjamin	Harvey Mudd College, USA
Martin Bohner	Missouri U of Science and Technology, USA
Nigel Boston	University of Wisconsin, USA
Amarjit S. Budhiraja	U of N Carolina, Chapel Hill, USA
Pietro Cerone	La Trobe University, Australia
Scott Chapman	Sam Houston State University, USA
Joshua N. Cooper	University of South Carolina, USA
Jem N. Corcoran	University of Colorado, USA
Toka Diagana	Howard University, USA
Michael Dorff	Brigham Young University, USA
Sever S. Dragomir	Victoria University, Australia
Behrouz Emamizadeh	The Petroleum Institute, UAE
Joel Foisy	SUNY Potsdam, USA
Errin W. Fulp	Wake Forest University, USA
Joseph Gallian	University of Minnesota Duluth, USA
Stephan R. Garcia	Pomona College, USA
Anant Godbole	East Tennessee State University, USA
Ron Gould	Emory University, USA
Sat Gupta	U of North Carolina, Greensboro, USA
Jim Haglund	University of Pennsylvania, USA
Johnny Henderson	Baylor University, USA
Natalia Hritonenko	Prairie View A&M University, USA
Glenn H. Hurlbert	Arizona State University,USA
Charles R. Johnson	College of William and Mary, USA
K. B. Kulasekera	Clemson University, USA
Gerry Ladas	University of Rhode Island, USA
David Larson	Texas A&M University, USA
Suzanne Lenhart	University of Tennessee, USA
Chi-Kwong Li	College of William and Mary, USA
Robert B. Lund	Clemson University, USA
Gaven J. Martin	Massey University, New Zealand
Mary Meyer	Colorado State University, USA
Emil Minchev	Ruse, Bulgaria
Frank Morgan	Williams College, USA
Mohammad Sal Moslehian	Ferdowsi University of Mashhad, Iran
Zuhair Nashed	University of Central Florida, USA
Ken Ono	Emory University, USA
Timothy E. O'Brien	Loyola University Chicago, USA
Joseph O'Rourke	Smith College, USA
Yuval Peres	Microsoft Research, USA
Y.-F. S. Pétermann	Université de Genève, Switzerland
Jonathon Peterson	Purdue University, USA
Robert J. Plemmons	Wake Forest University, USA
Carl B. Pomerance	Dartmouth College, USA
Vadim Ponomarenko	San Diego State University, USA
Bjorn Poonen	UC Berkeley, USA
Józeph H. Przytycki	George Washington University, USA
Richard Rebarber	University of Nebraska, USA
Robert W. Robinson	University of Georgia, USA
Javier Rojo	Oregon State University, USA
Filip Saidak	U of North Carolina, Greensboro, USA
James A. Sellers	Penn State University, USA
Hari Mohan Srivastava	University of Victoria, Canada
Andrew J. Sterge	Honorary Editor
Ann Trenk	Wellesley College, USA
Ravi Vakil	Stanford University, USA
Antonia Vecchio	Consiglio Nazionale delle Ricerche, Italy
Ram U. Verma	University of Toledo, USA
John C. Wierman	Johns Hopkins University, USA
Michael E. Zieve	University of Michigan, USA

PRODUCTION

Silvio Levy, Scientific Editor

Cover: Alex Scorpan

See inside back cover or msp.org/involve for submission instructions. The subscription price for 2019 is US $195/year for the electronic version, and $260/year (+$35, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Involve (ISSN 1944-4184 electronic, 1944-4176 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

Involve peer review and production are managed by EditFLOW® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

© 2019 Mathematical Sciences Publishers

involve

2019 vol. 12 no. 2