

## Lattice Paths, Riordan Matrices and RNA Numbers

Asamoah Nkwanta

Department of Mathematics

Morgan State University, Baltimore, MD 21251

email: asamoah.nkwanta@morgan.edu

ABSTRACT: Two infinite lower-triangular arrays, each of whose entries count unit-step lattice paths, are presented and denoted by  $R^*$  and  $R^{**}$ . The paths of the leftmost columns of  $R^*$  and  $R^{**}$  are counted by the RNA numbers. A bijection is then constructed between the paths. Thus, two types of lattice paths are enumerated by the RNA numbers. In addition, lattice path and RNA connections to other combinatorial objects are also given.

AMS SUBJECT CLASSIFICATION: Primary 05A15, Secondary 92D20

KEY WORDS: lattice path, Riordan matrix, RNA secondary structure

### 1. INTRODUCTION

We consider two infinite lower-triangular arrays where the leftmost columns count the following sequence of positive integers

$$\{s_n\}_{n \geq 0} = \{1, 1, 1, 2, 4, 8, 17, 37, 82, \dots\}. \quad (1)$$

The arrays are denoted by  $R^*$  and  $R^{**}$  and the first few entries are given below. Note that the leading 1 of the sequence is not included in the leftmost column of  $R^{**}$ .

$$R^* = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ 2 & 3 & 3 & 1 & 0 & 0 \\ 4 & 6 & 6 & 4 & 1 & 0 \\ 8 & 13 & 13 & 10 & 5 & 1 \end{pmatrix}, \quad R^{**} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 2 & 2 & 1 & 0 & 0 & 0 \\ 4 & 4 & 3 & 1 & 0 & 0 \\ 8 & 9 & 7 & 4 & 1 & 0 \\ 17 & 20 & 17 & 11 & 5 & 1 \end{pmatrix} \quad (2)$$

The numbers given by (1) also count possible ribonucleic acid (RNA) secondary structures of length  $n$  ( $n$  equals the number of nucleotides) from molecular biology and are called *RNA numbers*. See Sloane [14],[15] for more information on these numbers. The single-stranded RNA molecule consists of a sequence of base pairs derived from one of four nucleotides called: adenine (A), cytosine (C), guanine (G), and uracil (U) where A bonds or pairs with U and G bonds or pairs with C. RNA sequences are words defined over the four-letter genetic alphabet  $\{A, U, G, C\}$ . A linear sequence of such bases is called the *primary structure*. When an RNA molecule folds back on itself and forms new hydrogen bonds that form helical regions, the sequence is then referred to as the *secondary structure*. The following RNA sequence

*CAGCAUCACAUCCGCGGGUAAACGCU*

is an example of a primary secondary structure of length  $n = 27$ . This sequence when folded in two dimensions is referred to as a cloverleaf, in the biological literature, and is the secondary structure assumed by transfer RNA molecules, see Schmitt and Waterman [10]. The best pointer to the functionality of an RNA molecule is its overall three dimensional structure called the *tertiary structure*. We note that the three dimensional RNA folding problem is a difficult problem in computational biology. However, important information can be obtained from knowledge of its' primary and secondary structures - the linear sequence and Watson-Crick pairing of constituent bases. RNA molecules have important roles in regulating protein-coding genes and catalysis [2].

The enumeration of RNA secondary structures was first studied, from a graph theoretic point of view, by Waterman [20].

**Definition 1.1.** A secondary structure is a graph on the set of  $n$  labeled points  $\{1, 2, \dots, n\}$  such that the adjacency matrix  $A = (a_{ij})$  has the following three properties: (i)  $a_{i,i+1} = 1$  for  $1 \leq i \leq n-1$ , (ii) For all fixed  $i$ ,  $1 \leq i \leq n$ , there is at most one  $a_{ij} = 1$  where  $j \neq i \pm 1$ , and (iii) If  $a_{ij} = a_{kl} = 1$ , where  $i < k < j$ , then  $i \leq l \leq j$ .

If  $a_{ij} = 1$ ,  $i$  and  $j$  are said to be *bonded*. Thus, if  $s(n)$  denotes the total number of secondary structures defined on  $n$  labeled points, then the associated recurrence relation is

$$s(0) = 1, s(1) = 1, s(2) = 1 \text{ and for } n > 2$$

$$s(n+1) = s(n) + \sum_{j=1}^{n-1} s(j-1)s(n-j).$$

Donaghey [1] notes that the RNA numbers can also be computed by the following sum

$$\sum_{k \geq 1} \frac{1}{n-k} \binom{n-k}{k} \binom{n-k}{k-1}.$$

For  $n, k > 0$

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

counts the possible number of RNA structures of length  $n$  with exactly  $k$  basepairs [10],[21]. See Jin *et. al.* [4] for an interesting extension and new proof of this formula. The generating function for the RNA numbers, which can be derived from the recurrence relation, is

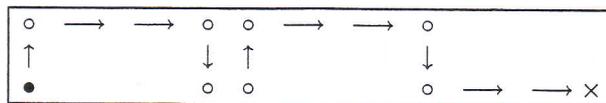
$$s(z) = \sum_{n \geq 0} s_n z^n = \frac{1 - z + z^2 - \sqrt{1 - 2z - z^2 - 2z^3 + z^4}}{2z^2} \quad (3)$$

Proofs of the recurrence relation and generating function can be found in the references [3], [8], [20], and [21].

We now define the type of lattice paths discussed in this paper.

**Definition 1.2.** A NSE lattice path is a sequence of contiguous and reversible unit steps that traverse the two dimensional integral lattice  $\mathbb{Z}^2$ . The step directions are  $(0, 1) = N$  (North or up),  $(0, -1) = S$  (South or down), and  $(1, 0) = E$  (East or right). All paths begin at the origin and move unit steps according to the following restrictions. The paths are considered to be in the first quadrant of the  $(x, y)$  plane and never pass below the  $x$ -axis. The length of each path is the number of unit steps, and the height corresponds to the  $y$  value of the endpoint  $(x, y)$  of the path.

We denote  $\text{NSE}^*$  as the set of unit step lattice paths that satisfy Definition 1.2 and do not have consecutive pairs of N and S steps. It is known that the  $n \times k$  array  $R^*$  counts unit step  $\text{NSE}^*$  lattice paths where  $n$  is the length of the path and  $k$  the height [6],[7]. For instance, illustrated below, NEESNEESEE is a  $\text{NSE}^*$  path of length 10 ending at height  $k = 0$ .



NSE\* Path

**Theorem 1.1.** ([6],[7]) There is a bijection between the set of unit-step  $\text{NSE}^*$  lattice paths of length  $n$  ending at height  $k = 0$  and the set of RNA secondary structures of length  $n$ .

Thus, the paths of the leftmost column of  $R^*$  are counted by the RNA numbers and bijective with RNA secondary structures. In this paper we will show that the  $n \times k$  array  $R^{**}$  also counts a certain subset of unit step NSE lattice paths with certain restrictions. Moreover, we will construct a bijection between the lattice paths counted by the leftmost columns of  $R^*$  and  $R^{**}$ . As a result of this two types of lattice paths enumerated by the RNA numbers are given. In addition, lattice path and RNA connections to other combinatorial objects are given.

## 2. LATTICE PATH INTERPRETATION OF $R^{**}$

A lattice path interpretation of  $R^{**}$  is given in this section, and we prove  $R^{**}$  is a Riordan matrix. A Riordan matrix is now defined and depends upon certain formal power series.

**Definition 2.1.** An infinite matrix  $L = (l_{n,k})_{n,k \geq 0}$  with complex entries  $\mathbb{C}$  is called a Riordan matrix if the  $k$ th column satisfies

$$\sum_{n \geq 0} l_{n,k} z^n := g(z) (f(z))^k$$

where  $g(z) = 1 + g_1 z + g_2 z^2 + \dots$  and  $f(z) = f_1 z + f_2 z^2 + f_3 z^3 + \dots$  belong to the ring of formal power series  $\mathbb{C}[[z]]$  and  $f_1 \neq 0$ .

The coefficients of the formal power series corresponding to  $g(z)$  and  $g(z) (f(z))^k$  represent the column entries of a Riordan matrix. The concept of representing columns of infinite matrices by coefficients of formal power series is not new and

goes back to Schur's paper on Faber polynomials [11]. A formal power series in auxiliary variable  $z$  of the form

$$b(z) = b_0 + b_1z + b_2z^2 + \dots = \sum_{n \geq 0} b_n z^n$$

is called an *ordinary generating function* of the sequence  $\{b_n\}$ . For instance, as earlier stated Equation (3) is the ordinary generating function for the sequence of RNA numbers. We note that a Riordan matrix can be defined by a pair of generating functions as  $L = (g(z), f(z))$ . Pascal's triangle, written in lower triangular form, denoted by

$$P = (1/(1-z), z/(1-z))$$

is typically given as an example of a Riordan matrix. Note that  $R^*$  is also a Riordan matrix [6],[7]. We mention here that the set of all Riordan matrices forms a noncommutative group call the *Riordan group*. See Shapiro et. al. [12] and Sprugnoli [16] for more on this interesting group.

One key point when working with Riordan matrices is to find the matrix formation rule. A *formation rule*, which we denote by  $[\mathbf{Z}; \mathbf{A}]$ , is a recurrence relation which defines the way entries of a Riordan matrix are computed. The notation  $[\mathbf{Z}; \mathbf{A}]$  means " $\mathbf{Z}$ " coincides with the formation of the zeroth column and " $\mathbf{A}$ " coincides with the formation of the other columns. In addition to this notation, formation rules are also denoted by dot diagrams. See Merlini, et. al. [5] for dot diagrams and related properties of Riordan matrices.

Following Rogers [9] and Merlini, et. al. [5], with minor adjustments, we give two useful characterizations of a given Riordan matrix. The formation rules which determine Riordan matrices are called  $\mathbf{A}$ - and  $\mathbf{Z}$ -sequences. The  $\mathbf{Z}$ -sequence ( $\mathbf{Z} = \{z_0, z_1, \dots\}$ ) characterizes the zeroth column. This means every element  $l_{n+1,0}$  can be expressed as a linear combination of all the elements in the preceding row, i.e.,

$$l_{n+1,0} = z_0 l_{n,0} + z_1 l_{n,1} + z_2 l_{n,2} + \dots$$

The  $\mathbf{A}$ -sequence ( $\mathbf{A} = \{a_0, a_1, \dots\}$ ,  $a_0 \neq 0$ ) characterizes the other columns. In this case every element  $l_{n+1,k+1}$  can be expressed as a linear combination with coefficients in  $\mathbf{A}$  of the elements in the preceding row, starting from the preceding column on, i.e.,

$$l_{n+1,k+1} = a_0 l_{n,k} + a_1 l_{n,k+1} + a_2 l_{n,k+2} + \dots$$

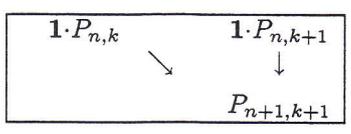
Thus, if  $A(y)$  and  $Z(y)$  are the generating functions (in auxiliary variable  $y$ ) of the  $\mathbf{A}$ - and  $\mathbf{Z}$ -sequences, respectively, then for a given Riordan matrix the generating functions  $g(z)$  and  $f(z)$  are the solutions of the functional equations

$$f(z) = zA(f(z)) \text{ and } g(z) = g_0 / (1 - zZ(f(z))). \quad (4)$$

Conversely,  $A(y)$  and  $Z(y)$  can be determined by letting  $y = f(z)$  and eliminating  $z$  from

$$A(y) = y/z \text{ and } Z(y) = (g(z) - g_0) / zg(z). \quad (5)$$

**Example 2.1.** The formation rule of the Pascal matrix  $P$  is  $[1, 0; 1, 1]$  where  $Z(y) = 1$  and  $A(y) = 1 + y$ . In general,  $P_{n+1, k+1}$  is computed as illustrated below.



$A$  - sequence

The formation rule of  $R^{**}$  is now given. As examples of the way the entries of  $R^{**}$  are formed, we refer back to (2) and observe that the second column entry 9 is computed by  $4 + 4 + 3 - 2$  and the leftmost column entry 8 is computed by  $4 + 4$ . The following diagrams:



illustrate the formation rule of the above examples. These patterns continue to form all of  $R^{**}$ . In general, the  $(n, k)$ th entry of  $R^{**}$  is formed and computed recursively by Proposition 2.1.

Before proving the proposition we denote  $NSE^{**}$  as the set of unit step lattice paths that satisfy Definition 1.2 and do not have consecutive pairs of S and N steps. Thus there are no paths with SN steps. For instance, the path NEESNEESEE given above is not an example of a  $NSE^{**}$  path. By letting  $s(n, k)$  denote the number of unit-step  $NSE^{**}$  lattice paths of length  $n$  and height  $k$ , the recurrence relation below is proved combinatorially in terms of  $NSE^{**}$  paths .

**Proposition 2.1.** Given the initial condition  $s(0, 0) = 1$ , then for  $n \geq 0$  and  $k \geq 1$ ,  $s(n + 1, k)$  satisfies the following relations where (a) is defined for the leftmost column of  $R^{**}$  and (b) is defined for the other columns of  $R^{**}$ :

$$\begin{aligned} \text{(a)} \quad s(n + 1, 0) &= s(n, 0) + s(n, 1) \\ \text{(b)} \quad s(n + 1, k) &= \begin{cases} s(n, k - 1) + s(n, k) + s(n, k + 1) - s(n - 1, k) \\ 0, \text{ if } k > n + 1 \end{cases} \end{aligned}$$

**Proof.** Suppose a unit-step  $NSE^{**}$  lattice path of length  $n$  and height  $k$  is given. Then, to form a new path of length  $(n + 1)$  and height  $k$  consider the following cases. Case (i): if the given path has length  $n$  and height  $k - 1$ , then on the last step there is 1 choice for height  $k - 1$  (the N step). In this case, all paths whose last step is N are counted by  $s(n, k - 1)$ . Case (ii): if the given path has length  $n$  and height  $k$ , then on the last step there is 1 choice for height  $k$  (the E step). In this case, all paths whose last step is E are counted by  $s(n, k)$ . Case (iii): if the given path has length  $n$  and height  $k + 1$ , then on the last step there is also 1 choice for

height  $k + 1$  (the S step). In this case, all paths whose last step is S are counted by  $s(n, k + 1)$ . Case (iv): if the given path has length  $(n - 1)$  and height  $k$ , then the last possible sequence of steps for height  $k$  is SN (south, north steps). The paths with SN steps are bad paths since paths with consecutive S and N steps are not allowed. Therefore, these paths are removed from the count by  $-s(n - 1, k)$ . In this particular case, all paths whose last consecutive sequence of steps is SN are counted by  $s(n - 1, k)$ . The over count occurs as a result of those paths counted by  $s(n, k - 1)$  whose next to last step is S. Combining all of the cases give all possible ways of forming a new  $(n + 1)$ st path of height  $k$ . Applying the addition principle, recurrence relation (a) is proved. Part (b) is easy to prove by similar reasoning.  $\square$

**Remark 2.1.** *The  $\mathbf{A}$ -sequence of  $R^{**}$  is not unique since it is easy to show  $R^{**}$  has another  $\mathbf{A}$ -sequence that is the same as the  $\mathbf{A}$ -sequence of  $R^*$ . See [6] and [7] for more details.*

We now prove  $R^{**}$  is a Riordan matrix. The generating functions that form the columns of  $R^{**}$  are now derived. Recall from Definition 2.1 if  $R^{**}$  is Riordan, then each column is of the form  $g(z)(f(z))^k$  for  $k > 0$ . By the formation rule of  $R^{**}$  (i.e., Proposition 2.1), the  $k$ th column generating function is defined as

$$gf^k = z(gf^{k-1} + gf^k + gf^{k+1}) - z^2gf^k.$$

Solving for  $f$  gives  $f = z + (z - z^2)f + zf^2$ . Now, solving  $f$  in terms of  $f(z)$  and simplifying,  $f(z)$  becomes  $f(z) = zs(z)$  where  $s(z)$  is defined by Equation (3). Similarly, the leftmost column generating functions is defined as  $g = 1 + z(g + gf)$ . Simplifying this equation and expressing  $g$  in terms of  $g(z)$  gives

$$g(z) = z^{-1}(s(z) - 1) = \frac{1 - z - z^2 - \sqrt{1 - 2z - z^2 - 2z^3 + z^4}}{2z^3}.$$

The coefficients of the generating function  $g(z)$  count unit-step NSE<sup>\*\*</sup> lattice paths of length  $n$  ending at height  $k = 0$ . These coefficients are indeed the RNA numbers given by (1) minus the leading 1. Thus,  $R^{**}$  is given a lattice path interpretation and defined as the Riordan matrix

$$R^{**} = (z^{-1}(s(z) - 1), zs(z)).$$

### 3. BIJECTION BETWEEN NSE\* AND NSE\*\* PATHS

A bijection is constructed between unit-step NSE\* and NSE\*\* lattice paths in this section. Since both sets count the RNA numbers, there exists a bijection between the sets.

**Theorem 3.1.** *There is a bijection between the set of unit-step NSE\* lattice paths of length  $n + 1$  ending at height  $k = 0$  and the set of unit-step NSE\*\* lattice paths of length  $n$  ending at height  $k = 0$ .*

**Proof.** First, we show how the unit steps of a NSE\*\* lattice path of length  $n$  and height  $k = 0$  are assigned. To establish the required correspondence, let  $l^*$  be a unit-step NSE\* lattice path of length  $n + 1$  and height  $k = 0$ . Recall from Theorem

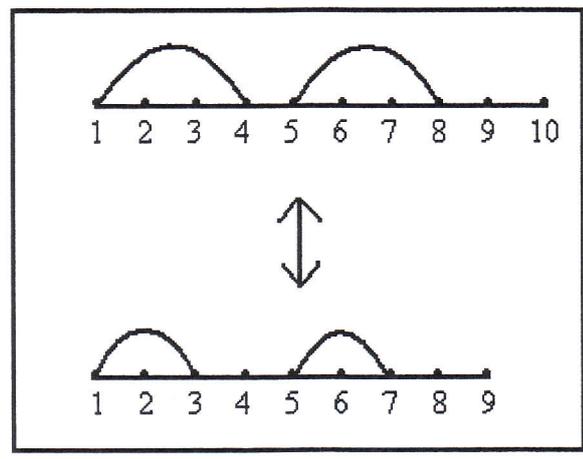


FIGURE 1

1.1 that the  $NSE^*$  paths are isomorphic to RNA secondary structures of length  $n$ , and the paths have a linear representation. Using this representation, write  $l^*$  in its linear form as a sequence of integers increasing in order from left to right along a horizontal axis where the corresponding N and S steps are joined by an arc. By the way  $NSE^*$  paths are defined no two adjacent points are connected by an arc and no two arcs intersect. Thus, in the construction  $l^*$  is considered to be in linear form.

To form a  $NSE^{**}$  path  $l^{**}$  of length  $n$ , the correspondence is set up according to the following rules. Given  $l^*$  in linear form, write  $n$  of the integer points of  $l^*$  as a sequence of integers increasing in order from left to right along a horizontal axis omitting the  $(n + 1)$ st point. Then starting on the left, join the arcs by leaving the left arc point in place and shifting each right arc point one unit to the left. By doing this, a linear representation is formed for  $l^{**}$  where the joining arcs represent from left to right NS steps and the remaining unjoined points represent E steps. In this construction there is always an E step or a sequence of E steps between consecutive S and N steps since the paths are of height 0 and never go below the  $x$ -axis. The special cases of the correspondence are now given. The origin of the  $NSE^*$  paths is assigned to the empty (zero) path of the  $NSE^{**}$  paths. And, the  $NSE^*$  path of length 1 which is the E step is assigned to the origin of the  $NSE^{**}$  paths. Therefore, a  $NSE^{**}$  path  $l^{**}$  of length  $n$  is formed. The correspondence is constructed and reversible. Thus, the correspondence is one-to-one and the theorem is proved.  $\square$

As an example of the correspondence consider  $l^*$  to be the unit-step  $NSE^*$  path NEESNEESEE previously given above. By the rules of the correspondence, the unit-step  $NSE^{**}$  path NESENESEE of length 9 in linear form is obtained. See Figure 1 where from left to right joined arcs are denoted by N and S steps and

non-arcs by E steps. See [6] and [7] for more details on the correspondence between RNA structures and  $NSE^*$  paths.

**Remark 3.1.** *We point out here that some  $NSE^{**}$  paths are consistent with RNA structures and some are not. For instance, the  $NSE^{**}$  path  $NESENESEE$  associated with Figure 1 is consistent with RNA. However,  $NSE^{**}$  paths with NS steps are not consistent with RNA structures. The RNA structures associated with these type of paths violate the Watson-Crick pairing. Although some  $NSE^{**}$  paths are also  $NSE^*$  paths, a subset of  $NSE^{**}$  paths are not related to RNA. The interesting point here is that the  $NSE^*$  paths are bijective with RNA while the  $NSE^{**}$  paths are not.*

#### 4. NONCROSSING PARTITIONS AND OTHER PATH BIJECTIONS

To find a combinatorial relation between the RNA numbers and noncrossing partitions consider the set  $[n] := \{1, 2, \dots, n\}$ . A partition  $\pi$  of  $[n]$  is said to be *noncrossing* if  $1 \leq a < b < c < d \leq n$  and if  $B_1$  and  $B_2$  are blocks of  $\pi$  such that  $a, c \in B_1$  and  $b, d \in B_2$ , then  $B_1 = B_2$ . That is, given that the conditions are satisfied,  $a, b, c$  and  $d$  are all in the same block. As an example of a noncrossing partition of  $[6] = \{1, 2, 3, 4, 5, 6\}$  consider

$$\pi = 15/24/3/6 \tag{6}$$

where the slashes separate the blocks. The linear representation of  $\pi$  can be illustrated where successive elements in the same block are joined by arcs.

Following Simion and Ullman [13], a word  $w$  of length  $n - 1$  over the alphabet  $\{b, e, l, r\}$  can be associated with a noncrossing partition  $\pi$ . See the reference for detailed definitions of each letter in the alphabet. By eliminating the letter  $r$  and the consecutive pair of letters  $b$  and  $e$  from any potential word, another word  $w^*$  can be defined over the alphabet  $\{b, e, l\}$ . The word  $w^*$  can also be associated with a noncrossing partition  $\pi$ . For example, the noncrossing partition given by (6) is associated with the word  $w^*(\pi) = bbleel$ . A one-to-one correspondence between noncrossing partitions associated with  $w^*$  and the  $NSE^*$  lattice paths is constructed. The correspondence is constructed according to the following rules:  $b \rightarrow N$ ,  $e \rightarrow S$ , and  $l \rightarrow E$  where the arrow means "corresponds to".

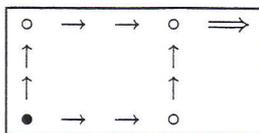
In addition to the correspondence between noncrossing partitions associated with  $w^*$  and the  $NSE^*$  paths, a one-to-one correspondence is also constructed between noncrossing partitions associated with another word  $w^{**}$  and the  $NSE^{**}$  paths. In this case,  $w^{**}$  is defined over the alphabet  $\{b, e, l, r\}$  where now the letter  $r$  is not eliminated. However, the consecutive pair of letters  $b$  and  $e$  remain eliminated from the formation of any potential word. This particular correspondence is set up according to the following rules:  $b \rightarrow N$ ,  $e \rightarrow S$ ,  $l \rightarrow E$  and  $r \rightarrow NS$ . For example, the noncrossing partition given by (6) is now associated with the word  $w^{**}(\pi) = brel$ . Thus, by way of both sets of lattice paths ( $NSE^*$  and  $NSE^{**}$ ), combinatorial relations are established between certain subsets of noncrossing partitions and unit-step lattice paths.

A topic of interest in lattice path combinatorics is the enumeration of pairs of nonintersecting lattice paths. Thus, another lattice path bijection is constructed

where the correspondence is set up between NSE\*\* paths and pairs of nonintersecting paths. Moreover, a one-to-one correspondence between pairs of nonintersecting unit-step lattice paths of length  $n$  and unit-step NSE\*\* lattice paths of length  $n$  and height 0 is constructed.

Consider pairs of nonintersecting paths in the integer lattice  $\mathbb{Z}^2$  both starting at the origin and proceeding in either a unit step E or N (east or north) direction such that the pair do not meet again after leaving the origin until the  $n$ th step. The noncrossing condition is relaxed in the sense that two paths may touch but may not cross. The path steps are restricted such that there are no pair of N steps, and no pair of N and E steps follow a pair of E and N steps. This means there are no double north steps and no parting pair of paths last step is the  $n$ th step.

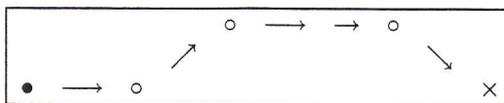
For a given NSE\*\* path, the correspondence is set up according to the following rules:  $N \rightarrow NE$ ,  $S \rightarrow EN$ , and  $E \rightarrow EE$  where the arrow means "corresponds to". For the NE pair, the first path proceeds in the N direction and the second in the E direction. Likewise for the EN pair, the first path proceeds in the E direction and the second in the N direction. As an example, the NSE\*\* path NNSSE corresponds with the path pair NE-NE-EN-EN-EE. The figure below illustrates the noncrossing path pair of the example where the symbol " $\bullet$ " denotes the origin of the path, the up arrow " $\uparrow$ " denotes N steps, the horizontal arrow " $\rightarrow$ " denotes E steps, and the horizontal arrow " $\Rightarrow$ " denotes double E steps.



Nonintersecting Path

The correspondence is reversible. Thus, the correspondence is one-to-one and the bijection between path pairs and NSE\*\* paths is established.

The RNA numbers also count peak free Motzkin paths. These are unit-step lattice paths with steps  $(1, 1) = U$  (Up),  $(1, -1) = D$  (Down), and  $(1, 0) = L$  (Level) that stay in the upper half plane starting at the origin and ending on the  $x$ -axis. These paths have no peaks in the sense that there are no paths with consecutive U and D steps. For instance, illustrated below, LULLD is an example of a peak free Motzkin path of length 5 ending at height  $k = 0$ .



Motzkin Path

A bijection can easily be constructed between these Motzkin paths and NSE\*\* paths of length  $n$ . The correspondence is constructed according to the following rules:

$$U \rightarrow N, D \rightarrow S, \text{ and } L \rightarrow E.$$

## 5. OTHER APPEARANCES OF THE RNA NUMBERS

The Narayana numbers are also of combinatorial interest and are defined as  $N(n, k) = \frac{1}{n} \binom{n}{k} \binom{n}{k-1}$  for  $n \geq 1$  and  $k \geq 1$ . These numbers can be put into infinite lower-triangular matrix form, denoted as  $N$ . The triangle  $N$  is not Riordan, and the first few entries are

$$N = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 3 & 1 & 0 & 0 \\ 1 & 6 & 6 & 1 & 0 \\ 1 & 10 & 20 & 10 & 1 \end{pmatrix}.$$

It is known that the row sums of  $N$  are the  $n$ th Catalan numbers

$$c_n = \frac{1}{1+n} \binom{2n}{n}$$

[19]. The numbers  $c_n$  occur in a wide variety of combinatorial problems and algebraic applications (see [17]). A combinatorial interpretation of  $N$  is that  $N(n, k)$  counts the number of Dyck paths of length  $2n$  with  $k$  peaks. A *Dyck* path is a path in the first quadrant, which begins at the origin, ends at  $(2n, 0)$ , and consists of north-east and south-east steps. Note, the sequence of diagonal slices  $1, 1, 1+1, 1+3, 1+6+1, \dots$  of the  $N$  triangle give the first few RNA numbers. This can be proved by using generating functions, where the bivariate GF associated with the Narayana numbers is noted by Stanley [17].

Several appearances of the RNA numbers related to planar trees are in [1]. Other combinatorial applications and interpretations of the RNA numbers are discussed in [18].

## Acknowledgments

The author thanks Professor Louis W. Shapiro and the combinatorics research group of Howard University for useful comments and suggestions on early drafts of this paper. Special thanks is also given to Sumitra Rudra for useful comments.

## REFERENCES

- [1] R. Donaghey, *Automorphisms on Catalan trees and bracketings*, J. Combinatorial Theory, Series B **29** (1980) 75-90.
- [2] H.H. Gan, S. Pasquali, and T. Schlick, *Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design*, Nucleic Acids Res. **31** (2003) 2926-2943.
- [3] J.A. Howell, T.F. Smith and M.S. Waterman, *Computation of generating functions for biological molecules*, SIAM J. Appl. Math. **39** (1980) 119-133.
- [4] E.Y. Jin, J. Qin, and C.M. Reidys, *Combinatorics of RNA structures with pseudoknots*, Bull. Math. Biol. **70** (2008) 45-67.
- [5] D. Merlini, D. G. Rogers, R. Sprugnoli, and M. C. Verri, *On some alternative characterizations of Riordan arrays*, Can. J. Math. **49** (1997) 301-320.
- [6] A. Nkwanta, *Lattice paths and RNA secondary structures*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science **34** (1997) 137-147.
- [7] A. Nkwanta, *Lattice paths, generating functions, and the Riordan group*, Ph.D. Thesis, Howard University, Washington, DC 1997.

- [8] F.S. Roberts and B. Tesman, *Applied Combinatorics*, 2nd Edition, Pearson Prentice Hall, New Jersey, 2005.
- [9] D.G. Rogers, *Pascal triangles, Catalan numbers and renewal arrays*, *Discrete Math.* **22** (1978) 301-310.
- [10] W.R. Schmitt and M.S. Waterman, *Linear trees and RNA secondary structures*, *Discrete Appl. Math.* **51** (1994) 317-323.
- [11] I. Schur, *On Faber polynomials*, *Amer. J. Math.* **67** (1945) 33-41.
- [12] L.W. Shapiro, S. Getu, W.J. Woan and L. Woodson, *The Riordan Group*, *Discrete Appl. Math.* **34** (1991) 229-239.
- [13] R. Simion and D. Ullman, *On the structure of the lattice of noncrossing partitions*, *Discrete Math.* **98** (1991) 193-206.
- [14] N.J.A. Sloane and S. Plouffe, *The Encyclopedia of Integer Sequences*, Academic Press, San Diego, 1995.
- [15] N.J.A. Sloane, *The On-line Encyclopedia of Integer Sequences*, <http://www.research.att.com/njas/sequences/index.html>, 2001.
- [16] R. Sprugnoli, *Riordan arrays and the Abel-Gould identity*, *Discrete Math.* **142** (1995) 213-233.
- [17] R.P. Stanley, *Enumerative Combinatorics*, Vol. 2, Cambridge Studies in Advanced Mathematics **62**, Cambridge University Press, Cambridge, 1999.
- [18] P.R. Stein and M.S. Waterman, *On some new sequences generalizing the Catalan and Motzkin numbers*, *Discrete Math.* **26** (1979) 261-272.
- [19] T.V. Narayana, *Lattice Path Combinatorics with Statistical Applications*, University of Toronto Press, 1979
- [20] M.S. Waterman, *Secondary structure of single-stranded nucleic acids*, *Studies in Foundations & Combinatorics, Advances in Mathematics Supplementary 1*, (1978) 167-212.
- [21] M.S. Waterman, *Introduction to Computational Biology: Maps, Sequences and Genomes*, Chapman & Hall/CRC Press, Boca Raton, 2000.