Random *k*-noncrossing RNA structures

William Y. C. Chen, Hillary S. W. Han, and Christian M. Reidys[1]

Center for Combinatorics, LPMC-TJKLC, Nankai University, Tianjin 300071, People's Republic of China

Edited by Michael S. Waterman, University of Southern California, Los Angeles, CA, and approved September 24, 2009 (received for review June 30, 2009)

In this paper, we introduce a combinatorial framework that provides an interpretation of RNA pseudoknot structures as sampling paths of a Markov process. Our results facilitate a variety of applications ranging from the energy-based sampling of pseudoknot structures as well as the ab initio folding via hidden Markov models. Our main result is an algorithm that generates RNA pseudoknot structures with uniform probability. This algorithm serves as a steppingstone to sequence-specific as well as energy-based transition probabilities. The approach employs a correspondence between pseudoknot structures, parametrized in terms of the maximal number of mutually crossing arcs and certain tableau sequences. The latter can be viewed as lattice paths. The main idea of this paper is to view each such lattice path as a sampling path of a stochastic process and to make use of *D*-finiteness for the efficient computation of the corresponding transition probabilities.

Pseudoknots have long been known as important structural elements (see Fig. 1) (1). These cross-serial interactions between RNA nucleotides are functionally important in tRNAs, RNaseP (2), telomerase RNA (3), and ribosomal RNAs (4). Pseudoknots in plant virus RNAs mimic tRNA structures, and in vitro selection experiments have produced pseudoknotted RNA families that bind to the HIV-1 reverse transcriptase (5). Important general mechanisms, such as ribosomal frame shifting, are dependent upon pseudoknots (6).

Despite their biological importance, pseudoknots are typically excluded from large-scale computational studies. Although the problem has attracted considerable attention in the last decade and several software tools (7) have become available, the required resources have remained prohibitive for applications beyond individual molecules. Lyngso et al. (8) have shown that the prediction of general RNA pseudoknot structures is nondeterministic polynomial time NP-complete. In the literature, some variant of the dynamic programming (DP) paradigm is often employed (7). This DP method generates certain subclasses of pseudoknots. We will discuss below that the DP paradigm is ideally suited for an inductive, or context-free, structure class. However, because of the cross-serial bonds, we a priori know that the DP paradigm is only of limited applicability for RNA pseudoknot structures. In addition, DP-based approaches are not even particularly time efficient, a point in case being ref. 7's exhibition of a time complexity of $O(n^6)$. The algorithmic difficulties are confounded by the fact that the thermodynamics of pseudoknots are poorly understood.

The approach that we take here is based on the observation that pseudoknotted RNA structures are in a natural way related to well-understood combinatorial objects. The key algorithmic innovation is a Markov process that efficiently generates pseudoknotted structures with a uniform measure. Biophysical realism can be added by modifying the transition rates.

In order to put our approach into context, let us give a retrospective overview. Three decades ago Waterman et al. (9–11) analyzed RNA secondary structures. Secondary structures are coarse-grained RNA contact structures. They can be represented as diagrams, i.e., labeled graphs over the vertex set $[n] = \{1, \ldots, n\}$ with vertex degrees ≤ 1, represented by drawing their vertices on a horizontal line and their arcs (i,j) $(i < j)$ in the upper halfplane (see Figs. 1 and 2). Here, vertices and arcs correspond to the nucleotides **A**, **G**, **U**, and **C** and Watson–Crick (**A-U**, **G-C**) and (**U-G**) base pairs, respectively.

In a diagram, two arcs (i_1,j_1) and (i_2,j_2) are called "crossing" if $i_1 < i_2 < j_1 < j_2$ holds. Accordingly, a k-crossing is a sequence of arcs $(i_1,j_1),\ldots,(i_k,j_k)$ such that $i_1 < i_2 < \cdots < i_k < j_1 < j_2 < \cdots < j_k$, (Fig. 2).

We call diagrams containing at most $(k-1)$-crossings "k-noncrossing diagrams" (k-noncrossing partial matchings). RNA secondary structures have no crossings in their diagram representation, [see Fig. 2 (l.h.s.)], and are therefore 2-noncrossing diagrams.

The efficient minimum free energy (MFE) folding of secondary structures is a consequence of the following relation of the numbers of RNA secondary structures over n nucleotides, $S_2(n)$ (9),

$$S_2(n) = S_2(n-1) + \sum_{j=0}^{n-3} S_2(n-2-j)S_2(j), \qquad [1]$$

where $S_2(n) = 1$ for $0 \leq n \leq 2$. Accordingly, RNA secondary structures satisfy a constructive recursion. As mentioned above, this relation suggests the DP recursions used for the polynomial time folding of secondary structures (10) and has therefore profound algorithmic implications. The uniform generation of RNA secondary structures is well known (12) and can be derived in linear time by using the framework of Flajolet et al. (13).

k-noncrossing RNA structures (14, 15) are k-noncrossing diagrams without arcs of the form $(i,i+1)$ and represent a natural generalization. The notion of k-noncrossing stipulates that the complexity of a pseudoknot is related to the maximal number of mutually crossing bonds. Indeed, most natural RNA pseudoknots are 3-noncrossing (16). Because of the cross-serial interactions, the numbers of pseudoknot structures do not satisfy a recursion of the type of Eq. **1**, rendering the ab initio folding into MFE configurations (8) as well as the derivation of detailed statistical properties, a nontrivial task. Indeed, in order to derive statistical properties, the entire space of structures has to be exhaustively generated, which is only possible for small sequence lengths. Only a few statistical results, derived by using singularity analysis of the bivariate generating functions, are known (17).

There exists no general framework for the uniform generation of elements of a noninductive combinatorial class. However, in the context of graphs the subject of uniform generation via Markov processes has been studied. Work on the uniform generation of specific graphs in the context of parallel random access machine (PRAM) can be found in ref. 18, and Jerrum et al. (19, 20) studied approximation algorithms in the context of rapidly mixing Markov chains (21). We also refer to the paper of Wilf (22) as well as his book (23).

Our approach is as follows: We translate k-noncrossing diagrams into specific lattice walks and view the latter as sampling paths of a stochastic process (see Fig. 3).

The key observation is that the generating function of these walks is D-finite or equivalently, P-recursive (24). This means

Author contributions: C.M.R. and W.Y.C.C. designed research; and W.Y.C.C., H.S.W.H., and C.M.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: duck@santafe.edu.

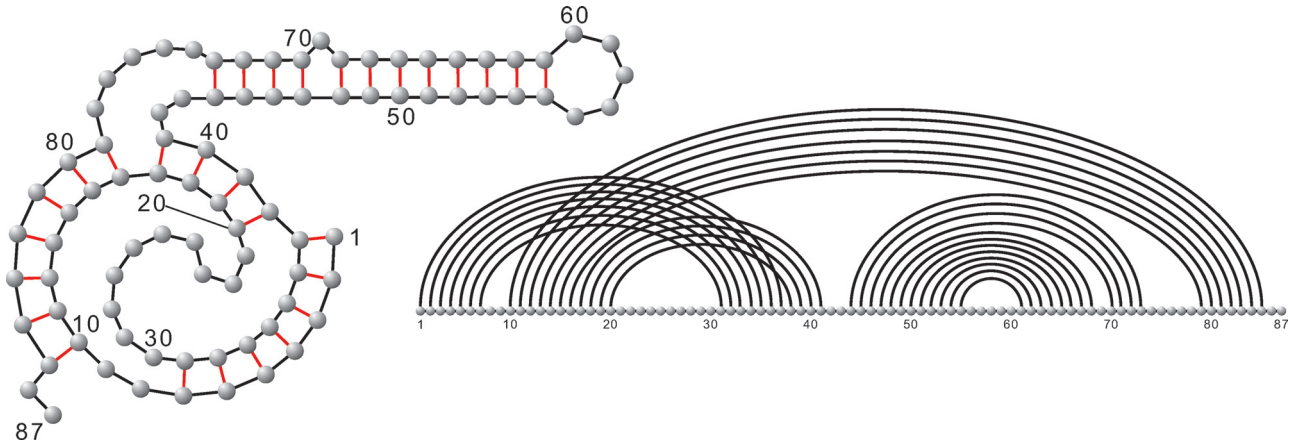

Fig. 1. The hepatitis delta virus pseudoknot structure represented as a planar graph and as a diagram. We display the structure as folded by the ab initio folding algorithm cross (*Left*) and the diagram representation (*Right*).

that there exists a finite recurrence relation with polynomial coefficients (see Corollary 1). Consequently, the numbers of these walks can be derived in linear time, and these allow us to compute the transition probabilities of the process displayed in Fig. 3. The implication is profound: The transition probabilities can be derived as a preprocessing step in polynomial time, after which a pseudoknot structure can be generated uniformly in linear time. Indeed, each structure is generated by the stochastic process having exactly n steps, each of which require constant time.

From Structures to Lattice Paths and Back

In this section, we translate RNA pseudoknot structures into lattice paths. For this purpose, we introduce shapes, $*$-tableaux, and the Robinson–Shensted–Knuth (RSK) algorithm (25).

A shape is a collection of squares arranged in left-justified rows with a weakly decreasing number of boxes in each row. A Young tableau is a filling of these squares by numbers. A Young tableau is weakly decreasing in each row and strictly decreasing in each column. A $*$-tableau of shape λ^n is a sequence of shapes $\varnothing = \lambda^0, \lambda^1, \ldots, \lambda^n$ such that for $1 \leq i \leq n$, λ^i is obtained from λ^{i-1} by either adding or removing one square or doing nothing (hesitating step) (see Fig. 4).

The RSK algorithm is a procedure which row-inserts elements into a Young tableau, T. Suppose we want to insert k into T. Let $T_{i,j}$ denote the element in the ith row and jth column. Let j be the largest integer such that $T_{1,j-1} \leq k$. (If $T_{1,1} > k$, then $j = 1$.) If $T_{1,j}$ does not exist, then simply add k at the end of the first row. Otherwise, if $T_{1,j}$ exists, then replace $T_{1,j}$ by k. Next, insert $T_{1,j}$ into the second row following the above procedure and continue until an element is inserted at the end of a row.

The RSK algorithm has also an inverse. Suppose we are given two shapes $\lambda^i \subsetneq \lambda^{i-1}$, which differ by exactly one square. Let T_{i-1} and T_i be Young tableau of shape λ^{i-1} and λ^i, respectively. Then there exists a unique j contained in T_{i-1} and a unique tableau T_i such that T_{i-1} is obtained from T_i by inserting j by using the RSK algorithm.

We are now ready to describe the correspondence between diagrams and $*$-tableaux discussed in ref. 26.

From k-Noncrossing Structures to $*$-Tableaux. Starting with an empty shape, consider the sequence $(n, n-1, \ldots, 1)$ and do the following:

- If j is the endpoint of an arc (i,j), then RSK-insert i.
- If j is the startpoint of an arc (j,s), then remove the square containing j.
- If j is an isolated point, then do nothing (see Fig. 5).

From $*$-Tableaux to k-Noncrossing Structures. Given a $*$-tableau of empty shape, $(\varnothing, \lambda^1, \ldots, \lambda^{n-1}, \varnothing)$, reading $\lambda^i \backslash \lambda^{i-1}$ from left to right, at step i we do the following:

- For a $+\square$-step, we insert i into the new square.
- For a $\varnothing$-step, we do nothing.
- For a $-\square$-step, we extract the unique entry, $j(i)$, of T^{i-1}.

The latter extractions generate the arc set $\{(i, j(i)) \mid i \text{ is a } \text{-}\square\text{-step}\}$ of a k-noncrossing diagram (see Fig. 6).

Therefore, each $*$-tableau of length n containing shapes with at most $(k-1)$-rows corresponds uniquely to a k-noncrossing partial matching on $[n]$ (26). We denote the numbers of $*$-tableaux and those without hesitating steps (oscillating tableaux) of shape λ^i and length $(n-i)$, by $\mathsf{O}_k^*(\lambda^i, n-i)$ and $\mathsf{O}_k^0(\lambda^i, n-i)$, respectively.

Reflection and D-Finiteness

The reflection principle (27–29) is a powerful technique in combinatorial enumeration. However, it is not directly applicable to RNA pseudoknot structures. Additional arguments (14) are needed for dealing with the nonreflectable, minimum arc length condition (see Lemma 1).

Given a $*$-tableau of shape λ, $(\lambda^i)_{i=0}^n$, we consider the number of squares in the sth row of shape λ^i, denoted by $x_s(i)$. It is evident that a $*$-tableau of shape λ with at most $(k-1)$ rows uniquely corresponds to a walk of length n that starts at $a = (k-1, k-2, \ldots, 1)$ and ends at $b = (k-1+x_1(n), \ldots, 1+x_{k-1}(n))$ and has steps $0, \pm e_i$, $1 \leq i \leq k-1$ such that $0 < x_{k-1} < \cdots < x_1$ at any step (see Fig. 7). That is, a $*$-tableau of shape λ with at most $(k-1)$ rows corresponds to a lattice path in $\mathbb{Z}^{k-1}$ that remains in the interior of the dominant Weyl chamber (28).

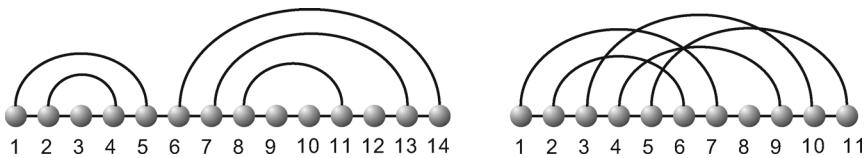

Fig. 2. k-noncrossing diagrams. A noncrossing (*Left*) and a 4-noncrossing diagram (*Right*) containing the three mutually crossing arcs (1, 7), (4, 9), and (5, 11).

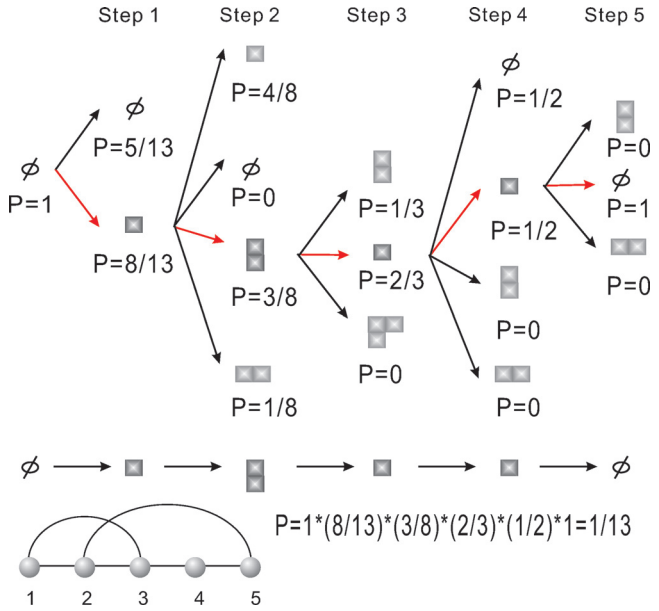

Fig. 3. Uniform generation. The stochastic process over shapes (*Top*), a sampling path (*Middle*) and its pseudoknot structure (*Bottom*). The transition probabilities are computed in Theorem 2 as a preprocessing step.

For $a,b \in \mathbb{Z}^{k-1}$, let $\Gamma(a,b)$ denote the set of walks without 0 steps of length n. Clearly, $\Gamma_n^0(a,b) = \mathsf{O}_k^0(\lambda,n)$, where λ represents the unique shape with at most $(k-1)$ rows that corresponds to the lattice point $b \in \mathbb{Z}^{k-1}$. Let $I_r(2x)$ denote the hyperbolic Bessel function of the first kind of order r.

By using the reflection principle, Grabiner (30) derived the following relation

$$\sum_{n\geq 0} \Gamma_n^0(a,b)\frac{x^n}{n!} = \det[I_{a_i-b_j}(2x) - I_{a_i+b_j}(2x)]_{i,j=1}^{k-1}. \qquad [2]$$

In ref. 14 it is shown, by using Eq. **2**, that for $k \geq 2$, the numbers of k noncrossing RNA pseudoknot structures with minimum arc length 2, $S_k(n)$, are P-recursive and given by

$$S_k(n) = \sum_{b\leq \lfloor \frac{n}{2} \rfloor} (-1)^b \binom{n-b}{b} \mathsf{O}_k^*(\varnothing^0, n-2b), \qquad [3]$$

where $\mathsf{O}_k^*(\lambda^i, n-i)$ satisfies

$$\mathsf{O}_k^*(\lambda^i,n-i) = \begin{cases} \sum_{l=0}^{\frac{n}{2}} \binom{n-i}{2l} \mathsf{O}_k^0(\lambda^i, n-i-2l), \\ \qquad\qquad \text{for } (n-i) \text{ even} \\ \sum_{l=0}^{\frac{n}{2}} \binom{n-i}{2l+1} \mathsf{O}_k^0(\lambda^i, n-i-2l-1). \\ \qquad\qquad \text{for } (n-i) \text{ odd.} \end{cases} \qquad [4]$$

Thus, the number of k-noncrossing RNA pseudoknot structures can be derived from the quantities $\mathsf{O}_k^0(\lambda^i,n)$ given by Eq. **2**.

Uniform Generation

Eq. **2**, combined with the fact that D-finite functions form an algebra (24) implies that the ordinary generating function $\sum_{n\geq 0} \Gamma_n^0(a,b)x^n$ is D-finite. Because D-finiteness is equivalent to the P-recursiveness (25) of its coefficients, we derive

Corollary 1: *For fixed shape λ with at most $(k-1)$ rows and $n \in \mathbb{N}$, there exists some $m \in \mathbb{N}$ and polynomials $p_0(n),\ldots,p_m(n)$ such that*

$$p_m(n)\mathsf{O}_k^0(\lambda,n+m) + \cdots + p_0(n)\mathsf{O}_k^0(\lambda,n) = 0. \qquad [5]$$

In particular, the numbers $\mathsf{O}_k^0(\lambda,n)$ can be computed in $O(n)$ time.

We remark that for fixed n and λ, the derivation of Eq. **5** is a preprocessing step. For special cases, we can employ Zeilberger's algorithm (31, 32). The recursions of Corollary 1 can be found empirically with the MAPLE package gfun.

Theorem 2. *A random k-noncrossing structure can be generated, after polynomial preprocessing time, with uniform probability in linear time. The algorithmic implementation (see Algorithm 1) has $O(n^{k+1})$ preprocessing time and $O(n^k)$ space complexity. Each k-noncrossing structure is generated with $O(n)$ space and time complexity.*

Let $\mathsf{W}_k^*(\lambda^i, n-i)$ denote the number of $*$-tableaux of shape λ^i with at most $(k-1)$ rows of length $(n-i)$ that do not contain any $(+\square_1, -\square_1)$-steps, and we then have Algorithm 1.

Algorithm 1.

1. PShape $\leftarrow$ ArrayP(n,k) (computation of $\mathsf{O}_k^*(\lambda^i, n-i)$, $i = 0,1,\ldots,n-1,\lambda^i$)
2. SShape $\leftarrow$ ArrayS(n,k) (computation of $\mathsf{W}_k^*(\lambda_j^i, n-i)$, $j = 0,1^+,1^-,\ldots,(k-1)^+,(k-1)^-; i = 0,1,\ldots,n-1$, stored in the $k \times n$ array *SShape*)
3. While $i < n$ do
4. Flag $\leftarrow 1$
5. $X[0] \leftarrow W_k^*(\lambda_0^{i+1}, n-(i+1))$
6. $X[1] \leftarrow W_k^*(\lambda_{1^+}^{i+1}, n-(i+1)) - W_k^*(\lambda_{1^-}^{i+2}, n-(i+2))$
7. If flag $= 0$ and $j = 2$ then
8. $X[2] \leftarrow 0$
9. Else
10. $X[2] \leftarrow W_k^*(\lambda_{1^-}^{i+1}, n-(i+1))$
11. End if
12. sum $\leftarrow X[0] + X[1] + X[2]$
13. For j from 2 to $k-1$ do
14. $X[2j-1] \leftarrow W_k^*(\lambda_{j^+}^{i+1}, n-(i+1))$
15. $X[2j] \leftarrow W_k^*(\lambda_{j^-}^{i+1}, n-(i+1))$
16. sum $\leftarrow$ sum $+ X[2j-1] + X[2j]$
17. End for
18. Shape $\leftarrow$ Random(sum) (Random generates the random shape λ_j^{i+1} with probability $X[j]$/sum)
19. $i \leftarrow i+1$
20. If Shape $= \lambda_{1^+}^i$ then
21. Flag $\leftarrow 0$

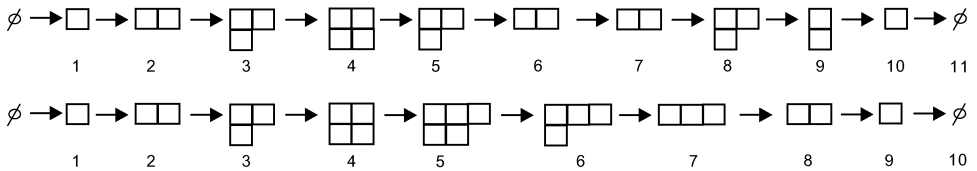

Fig. 4. $*$-tableaux with (*Upper*) and without (*Lower*) hesitating steps. The hesitation step in the *Upper* tableau is at (6, 7). In Fig. 6, we show how the *Upper* $*$-tableau induces a unique k-noncrossing structure.

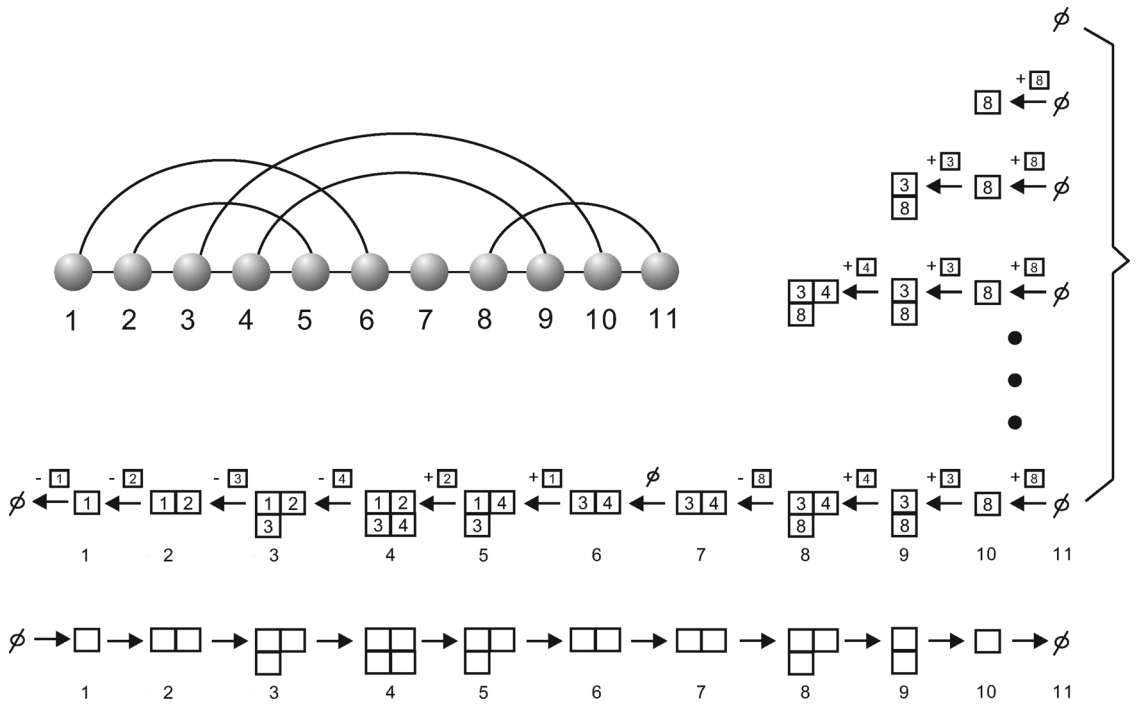

Fig. 5. From k-noncrosssing diagrams to $*$-tableaux using RSK insertion. Here we generate the *Upper* $*$-tableau of Fig. 4.

22. End if
23. Insert λ_j^{i+1} into Tableau
24. End while
25. Map(Tableau)

We remark that for $k = 3$, explicit formulas based on the work of refs. 29, 33, and 34 allow us to derive the transition probabilities.

Conclusion

This paper provides an interpretation of RNA pseudoknot structures as sampling paths of a Markov process. This point of view has the potential to offer radically new ways of dealing with the complex cross-serial interactions in RNA molecules.

It is not obvious that cross-serial interactions can be expressed in terms of a Markov process because the latter is by construction local, having no memory of the sampling path, except of the last step. Our construction shows why this is the case: $(k-1)$-crossings in RNA molecules can be expressed locally by a $(k-1)$ row of squares in the associated shape sequence. This locality holds for arbitrary k. The conclusion that cross-serial interactions are indeed local is good news for designing ab initio folding algorithms.

The framework presented here is a steppingstone towards the sampling with nonuniform transition probabilities. The

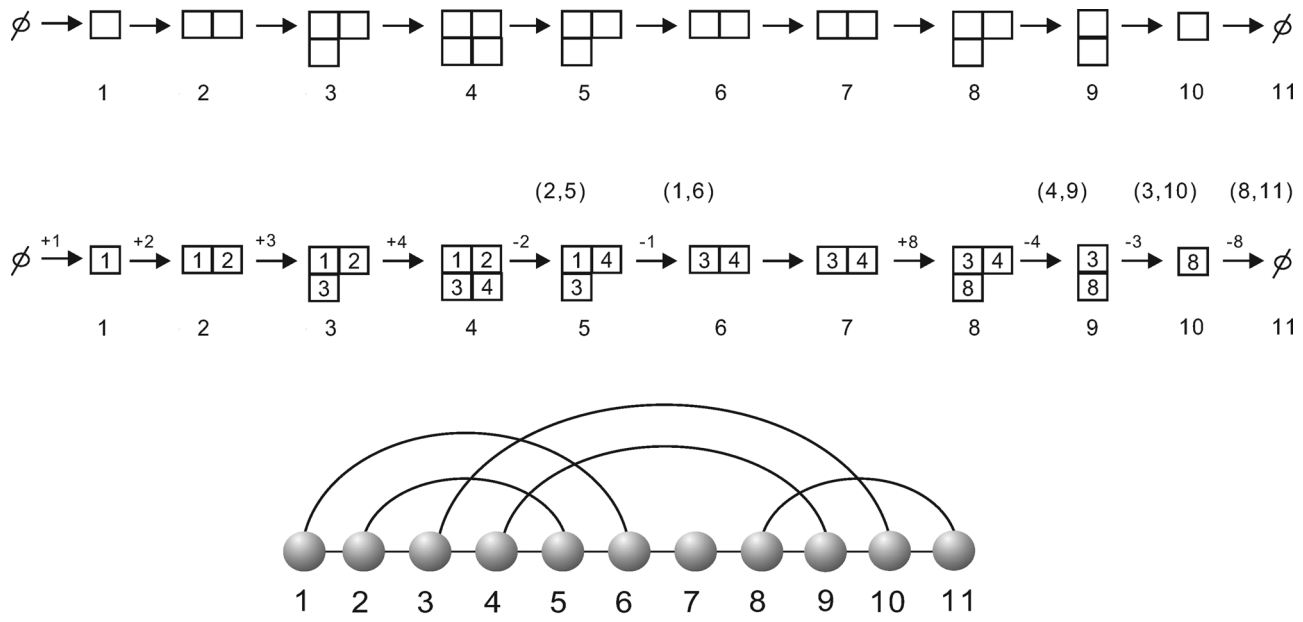

Fig. 6. From $*$-tableaux to partial matchings. If $\lambda^i \setminus \lambda^{i-1} = -\square$, then the unique number is extracted, RSK-insertion of this number into λ^i recovers the shape λ^{i-1}. This yields the arc set of a k-noncrossing partial matching.

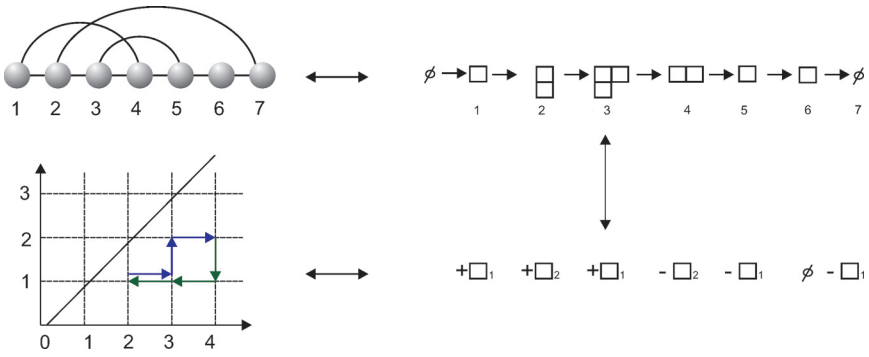

Fig. 7. From diagrams to lattice paths. A 3-noncrossing diagram is translated into a sequence of shapes (∗-tableaux) which in turn induces a walk that stays in the dominant Weyl chamber of $\mathbb{Z}^2$.

uniform sampling of pseudoknot structures, compatible with a given sequence, is displayed in Fig. 8. Here we insert the nucleotides into the squares and, in analogy to Theorem 2, consider compatible paths, thereby sampling uniformly. For instance, Theorem 2 immediately allows one to sample sequence-specific locally uniformly in linear time by setting all incompatible uniform transition probabilities to zero and then rescaling.

It is also possible to assign transitions that induce base pair's (i.e., extraction's) particular weights. This assignment leads to the energy-based sampling of pseudoknot structures, which can be made context dependent along the lines of ref. 35. Stacking bonds could be included as well in our framework. Higher-order Markov processes naturally model the dependencies of adjacent arcs.

Because our approach is path based, it offers the possibility to formalize the kinetics of the folding. In addition, we can generalize a class of stochastic context-free grammars for RNA secondary structures (36). Examining a set of known molecular foldings, it is now possible to derive the maximum-likelihood estimators of the model parameters (37) and to fold pseudoknot structures by using hidden Markov models (38). One important advantage of this approach is to avoid explicit knowledge of the energy parameters of pseudoknot loops.

Our algorithm is available in C and in MAPLE at http://www.combinatorics.cn/cbpc/unif.html.

Proof of the Main Result

A 1-arc corresponds to a subsequence of shapes $(\lambda^i, \lambda^{i+1}, \lambda^{i+2} = \lambda^i)$ obtained by first adding and then removing a square in the first row. This sequence corresponds to a pair of steps $(+\square_1, -\square_1)$ where $+\square_1$ and $-\square_1$ indicate that a square is added and subtracted in the first row, respectively. In terms of ∗-tableaux having at most $(k-1)$ rows, Eq. **3** can be rewritten as follows:

$$\mathsf{W}_k^*(\varnothing^0, n) = \sum_{b=0}^{\frac{n}{2}} (-1)^b \binom{n-b}{b} \mathsf{O}_k^*(\varnothing^0, n-2b).$$

In order to prove our main result, we have to generalize this relation from the empty shape, ∅ to arbitrary shapes, λ.

Lemma 1. *Let λ^i be an arbitrary shape with at most $(k-1)$ rows, then*

$$\mathsf{W}_k^*(\lambda^i, n-i) = \sum_{b=0}^{\frac{n-i}{2}} (-1)^b \binom{(n-i)-b}{b} \mathsf{O}_k^*(\lambda^i, n-i-2b). \quad [6]$$

Let $\mathcal{Q}_k^*(\lambda^i, n-i, j)$ denote the set of ∗-tableaux of shape λ^i of length $(n-i)$ having at most $(k-1)$ rows containing exactly j pairs $(+\square_1, -\square_1)$, and set $\mathsf{Q}_k^*(\lambda^i, n-i, j) = |\mathcal{Q}_k^*(\lambda^i, n-i, j)|$.

Proof: Let $(\lambda^s)_{s=0}^{(n-2b)-i}$ be a ∗-tableau of shape λ^i. We select from the set $\{0, \ldots, (n-2b)-i-1\}$ an increasing sequence of labels $(r_1, \ldots, r_b)$. For each r_s, we insert a pair $(+\square_1, -\square_1)$ after the corresponding shape λ^{r_s}. This insertion generates a ∗-tableau of length $(n-i)$ of shape λ^i.

Considering the above insertion for all sequences $(r_1, \ldots, r_b)$, we arrive at a family $\mathcal{F}_b$ of ∗-tableaux of length $(n-i)$ containing at least b pairs, $(+\square_1, -\square_1)$. Because we can insert at any position $0 \le h \le ((n-i)-2b-1)$, $\mathcal{F}_b$ has cardinality $\binom{(n-i)-b}{b} \mathsf{O}_k^*(\lambda^i, n-i-2b)$. By construction, each ∗-tableau $(\lambda^s)_{s=0}^{n-i} \in \mathcal{F}_b$ that exhibits exactly j pairs $(+\square_1, -\square_1)$ appears with multiplicity $\binom{j}{b}$, whence

$$\sum_{j \ge b} \binom{j}{b} \mathsf{Q}_k^*(\lambda^i, n-i, j) = \binom{(n-i)-b}{b} \mathsf{O}_k^*(\lambda^i, n-i-2b). \quad [7]$$

We consider $F_k(x) = \sum_{j \ge 0} \mathsf{Q}_k^*(\lambda^i, n-i, j) x^j$. By computing the Taylor expansion of $F_k(x)$ at $x = 1$, we obtain $\frac{1}{b!} F_k^b(1) = \sum_{j \ge b} \binom{j}{b} \mathsf{Q}_k^*(\lambda^i, n-i, j) 1^{j-b}$ and by computing the Taylor expansion of $F_k(x)$ at $x = 1$, we obtain

$$\begin{aligned} F_k(x) &= \sum_{b \ge 0} \frac{1}{b!} F_k^b(1) (x-1)^b \\ &= \sum_{b=0}^{\frac{n-i}{2}} \binom{(n-i)-b}{b} \mathsf{O}_k^*(\lambda^i, n-i-2b) (x-1)^b. \end{aligned}$$

Because $\mathsf{W}_k^*(\lambda^i, n-i) = \mathsf{Q}_k^*(\lambda^i, n-i, 0)$ is the constant term of $F_k(x)$, the lemma follows.

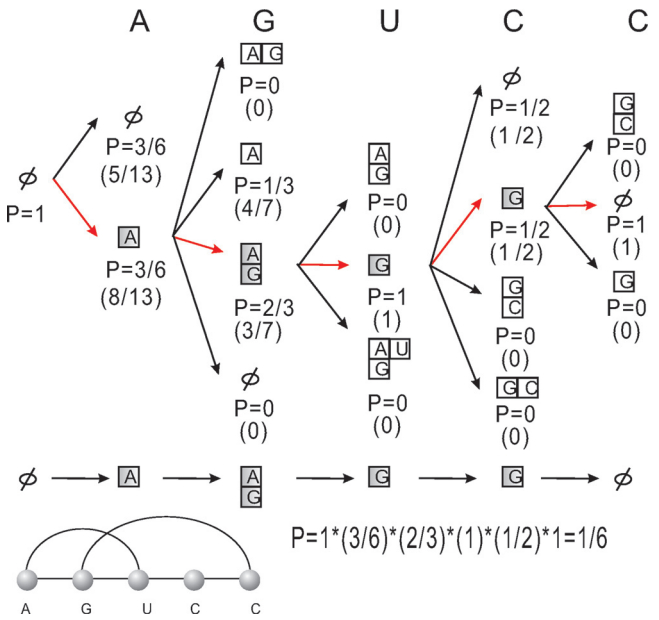

Fig. 8. Sequence-specific, uniform, and locally uniform sampling of RNA pseudoknot structures. Here we display the two sampling variants for the sequence AGUCC.

Proof of Theorem 2: The idea is to interpret $*$-tableaux without pairs of steps, $(+\square_1, -\square_1)$, (good $*$-tableaux) as paths of a stochastic process. To this end, we index the shapes λ^{i+1} according to their predecessors: Let $i = 0, 1, \ldots, n-1$ and $j \in \{0, 1^+, 1^-, \ldots, (k-1)^+, (k-1)^-\}$. Setting $\lambda_j^0 = \varnothing$, we write λ_j^{i+1}, if λ^{i+1} is obtained via

- doing nothing (λ_0^{i+1}),
- adding a square in the jth row $(\lambda_{j^+}^{i+1})$, or
- deleting a square in the jth row $(\lambda_{j^-}^{i+1})$.

With this notation, the number of good $*$-tableaux of shape $\lambda_{1^+}^{i+1}$ of length $(n-(i+1))$ is given as follows:

$$\mathsf{V}_k^*(\lambda_{1^+}^{i+1}, n-(i+1)) = \mathsf{W}_k^*(\lambda_{1^+}^{i+1}, n-(i+1)) - \mathsf{W}_k^*(\lambda_{1^-}^{i+2}, n-(i+2)).$$

In order to derive transition probabilities, we establish two equations: First, for any λ_j^i, where $j \neq 1^+$, we have $\mathsf{W}_k^*(\lambda_j^i, n-i) =$

$$\mathsf{V}_k^*(\lambda_{1^+}^{i+1}, n-(i+1)) + \mathsf{W}_k^*(\lambda_{1^-}^{i+1}, n-(i+1)) + \sum_{h=2}^{k-1} \left(\mathsf{W}_k^*(\lambda_{h^+}^{i+1}, n-(i+1)) + \mathsf{W}_k^*(\lambda_{h^-}^{i+1}, n-(i+1))\right) + \mathsf{W}_k^*(\lambda_0^{i+1}, n-(i+1))$$

and second, in the case of $j = 1^+$, we have $\mathsf{V}_k^*(\lambda_{1^+}^i, n-i) =$

$$\mathsf{V}_k^*(\lambda_{1^+}^{i+1}, n-(i+1)) + \mathsf{W}_k^*(\lambda_0^{i+1}, n-(i+1)) \sum_{h=2}^{k-1} \left(\mathsf{W}_k^*(\lambda_{h^+}^{i+1}, n-(i+1)) + \mathsf{W}_k^*(\lambda_{h^-}^{i+1}, n-(i+1))\right).$$

We are now in a position to specify the process $(X^i)_{i=0}^n$:

- $X^0 = X^n = \varnothing$ and X^i is a shape having at most $(k-1)$ rows.
- For $0 \le i \le n-1$, X^i and X^{i+1} differ by at most one square.
- There exists no subsequence $X^i, X^{i+1}, X^{i+2} = X^i$ obtained by first adding and second removing a square in the first row.
- For $j \neq 1^+$,

$$\mathbb{P}(X^{i+1} = \lambda_l^{i+1} \mid X^i = \lambda_j^i) = \begin{cases} \frac{\mathsf{W}_k^*(\lambda_l^{i+1}, n-(i+1))}{\mathsf{W}_k^*(\lambda_j^i, n-i)}, & \text{for } l \neq 1^+ \\ \frac{\mathsf{V}_k^*(\lambda_{1^+}^{i+1}, n-(i+1))}{\mathsf{W}_k^*(\lambda_j^i, n-i)}, & \text{for } l = 1^+. \end{cases} \quad [8]$$

- For $j = 1^+$,

$$\mathbb{P}(X^{i+1} = \lambda_l^{i+1} \mid X^i = \lambda_{1^+}^i) = \begin{cases} \frac{\mathsf{W}_k^*(\lambda_l^{i+1}, n-(i+1))}{\mathsf{V}_k^*(\lambda_{1^+}^i, n-i)}, & \\ & \text{for } l \neq 1^+, 1^- \\ \frac{\mathsf{V}_k^*(\lambda_{1^+}^{i+1}, n-(i+1))}{\mathsf{V}_k^*(\lambda_{1^+}^i, n-i)}, & \text{for } l = 1^+. \end{cases} \quad [9]$$

We observe that Eqs. **8** and **9** imply

$$\prod_{i=0}^{n-1} \mathbb{P}(X^{i+1} = \lambda^{i+1} \mid X^i = \lambda^i) = \frac{\mathsf{W}_k^*(\lambda^n = \varnothing, 0)}{\mathsf{W}_k^*(\lambda^0 = \varnothing, n)} = \frac{1}{\mathsf{W}_k^*(\varnothing, n)}. \quad [10]$$

Consequently, the process $(X^i)_{i=0}^n$ generates random k-noncrossing structures with uniform probability in $O(n)$ time and space. According to Corollary 1, we can for any λ^i, having at most $(k-1)$ rows, compute $\mathsf{O}_k^0(\lambda^i, n-i)$ in $O(n)$ time. Consequently, we can generate the arrays $(\mathsf{O}_k^*(\lambda^i, n-i))_{\lambda^i, n-i}$ and $(\mathsf{W}_k^*(\lambda^i, n-i))_{\lambda^i, n-i}$ in $O(n^2) + O(n^2)\, O(n^{k-1})$ time and $O(n^k)$ space.

A random k-noncrossing structure is then generated as a $*$-tableau with at most $(k-1)$ rows by using the array $(\mathsf{W}_k^*(\lambda^i, n-i))_{\lambda^i, n-i}$ with $O(n)$ time and space complexity.

ACKNOWLEDGMENTS. We thank M. Marz, M. Nebel, and P.F. Stadler for their help and suggestions. We also wish to thank the referees for valuable comments. This work was supported by the 973 Project, the PCSIR Project of the Ministry of Education, and the National Science Foundation of China.

1. Westhof E, Jaeger L, (1992) RNA pseudoknots. *Curr Opin Struct Biol* 2:327–333.
2. Loria A, Pan T (1996) Domain structure of the ribozyme from eubacterial ribonuclease P. *RNA* 2:551–563.
3. Staple DW, Butcher SE (2005) Pseudoknots: RNA structures with diverse functions. *PLoS Biol* 3:956–959.
4. Konings DAM, Gutell RR (1995) A comparison of thermodynamic foldings with comparatively derived structures of 16S and 16S-like rRNAs. *RNA* 1:559–574.
5. Tuerk C, MacDougal S, Gold L (1992) RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc Natl Acad Sci USA* 89:6988–6992.
6. Chamorro M, Parkin N, Varmus HE (1991) An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA J. *Proc Natl Acad Sci USA* 89:713–717.
7. Rivas E, Eddy SR (1999) A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J Mol Biol* 285:2053–2068.
8. Lyngso RB, Pedersen CNS (2000) RNA pseudoknot prediction in energy-based models. *J Comput Biol* 7:409–427.
9. Waterman MS (1979) Combinatorics of RNA hairpins and cloverleafs. *Stud Appl Math* 60:91–96.
10. Nussinov R, Pieczenik G, Griggs JR, Kleitman DJ (1978) Algorithms for loop matchings. *SIAM J Appl Math* 35:68–82.
11. Kleitman D (1970) Proportions of irreducible diagrams. *Stud Appl Math* 49:297–299.
12. Hofacker IL (2003) Vienna RNA secondary structure server. *Nucleic Acids Res* 31:3429–3431.
13. Duchon P, Flajolet P, Louchard G, Schaeffer G (2004) Boltzmann samplers for the random generation of combinatorial structures. *Combin Probab Comput* 13:577–625.
14. Jin EY, Qin J, Reidys CM (2008) Combinatorics of RNA structures with pseudoknots. *Bull Math Biol* 70:45–67.
15. Jin EY, Reidys CM (2008) Combinatorial design of pseudoknot RNA. *Adv Appl Math* 42:135–151.
16. Haslinger C, Stadler PF (1999) RNA structures with pseudo-knots: Graph-theoretical, combinatorial, and statistical properties. *Bull Math Bio* 61:437–467.
17. Jin EY, Reidys CM (2007) Central and local limit theorems for RNA structures. *J Theor Biol* 250:547–559.
18. Zito M, Pu I, Alnos M, Gibbons A (1996) RCN algorithms for the uniform generation of combinatorial structures. *Proceedings of the Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*. (Soc Ind Appl Math, Philadelphia) 429–437.
19. Jerrum MR, Valiant LG, Vazirant VV (1986) Random generation of combinatorial structures from a uniform distribution. *Theor Comput Sci* 43:169–188.
20. Jerrum M, Sinclair A (1990) Fast uniform generation of regular graphs. *Theor Comput Sci* 73:91–100.
21. Aldous D (1982) Random walks on finite groups and rapidly mixing Markov chains. Séminaire de Probabilités XVII. *Lecture Notes in Mathematics* (Springer, Berlin) Vol. 986, pp 243–297.
22. Wilf HS (1977) A unified setting for sequencing, ranking, and selection algorithms for combinatorial objects, *Adv Math* 24:281–291.
23. Nijenhuis A, Wilf HS (1978) *Combinatorial Algorithms* (Academic, New York).
24. Stanley RP (1980) Differentiably finite power series. *Eur J Combin* 1:175–188.
25. Stanley RP (1996) *Enumerative Combinatorics*. (Cambridge Univ Press, Cambridge, UK), Vol 1.
26. Chen WYC, Deng EYP, Du RRX, Stanley RP, Yan CH (2007) Crossings and nestings of matchings and partitions. *Trans Amer Math Soc* 359:1555–1575.
27. Renault M (2008) Lost (and found) in translation: André's actual method and its application to the generalized Ballot problem. *Am Math Monthly* 115: 358–363.
28. Gessel IM, Zeilberger D (1992) Random walk in a Weyl chamber. *Proc Am Math Soc* 115:27–31.
29. Lindström B (1973) On the vector representation of induced matroids. *Bull London Math Soc* 5:85–90.
30. Grabiner DJ, Magyar P (1993) Random walks in Weyl chambers and the decomposition of Tensor Powers. *J Algebr Combin* 2:239–260.
31. Zeilberger D (1990) A holonomic systems approach to special functions identities. *J Comput Appl Math* 32:321–368.
32. Salvy B, Zimmerman P (1994) GFUN: A Maple package for the manipulation of generating and holonomic functions in one variable. *ACM Trans Math Software* 20:163–177.
33. Gessel IM, Viennot XG (1989) *Determinants, paths, and plane partitions*, preprint.
34. Gouyou–Beauschamps D (1989) Standard Young tableaux of height 4 and 5. *Eur J Combin* 10:69–82.
35. Witwer C, Hofacker IL, Stadler PF (2004) Prediction of consensus RNA secondary structures including pseudoknots. *IEEE/ACM Trans Comput Biol Bioinform* 1:66–77.
36. Knudsen B, Hein J (1999) RNA secondary structure prediction using stochastic context-free grammars and evolutionary history. *Bioinformatics* 15:446–454.
37. Metzler D, Nebel ME (2008) Predicting RNA secondary structures with pseudoknots by MCMC sampling. *J Math Biol* 56:161–181.
38. Laube U, Nebel ME (2009) Maximum likelihood analysis of algorithms and data structures. *Theor Comp Sci*, doi: 10.1016/j.tcs.2009.09.025.