

# Towards the Algorithmic Molecular Self-Assembly of Fractals by Cotranscriptional Folding<sup>\*</sup>

Yusei Masuda, Shinnosuke Seki<sup>\*\*</sup>, and Yuki Ubukata

Department of Computer and Network Engineering, The University of  
Electro-Communications, 1-5-1, Chofugaoka, Chofu, Tokyo, 1828585, Japan  
`s.seki@uec.ac.jp`

**Abstract.** RNA cotranscriptional folding has been just experimentally proven capable of self-assembling a rectangular tile at nanoscale *in vivo* (RNA origami). We initiate the theoretical study on the algorithmic self-assembly of shapes by cotranscriptional folding using a novel computational model called the oritatami system. We propose an oritatami system that folds into an arbitrary finite portion of the Heighway dragon fractal, also-known as the paperfolding sequence  $P = \text{RRLRRLLR} \dots$ . The  $i$ -th element of  $P$  can be obtained by feeding  $i$  in binary to a 4-state DFA with output (DFAO). We implement this DFAO and a bit-sequence bifurcator as modules of oritatami system. Combining them with a known binary counter yields the proposed system.

## 1 Introduction

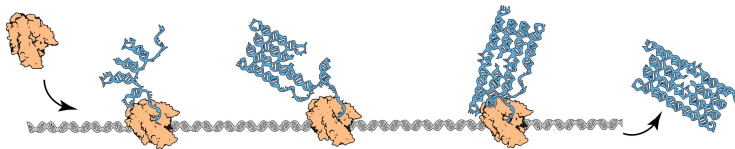
An RNA sequence, over nucleotides of four kinds A, C, G, U, is synthesized (*transcribed*) from its template DNA sequence over A, C, G, T nucleotide by nucleotide by an RNA polymerase (RNAP) enzyme according to the one-to-one mapping  $A \rightarrow U$ ,  $C \rightarrow G$ ,  $G \rightarrow C$ , and  $T \rightarrow A$  (for details, see, e.g., [2]). The yield, called *transcript*, starts folding immediately after it emerges from RNAP. This is the *cotranscriptional folding* (see Fig. 1). Geary, Rothmund, and Andersen have recently demonstrated the capability of cotranscriptional folding to self-assemble an RNA molecule of an intended shape at nano-scale [7]. They actually proposed an architecture of a DNA sequence whose transcript folds cotranscriptionally into an RNA tile of specific rectangular shape highly likely *in vitro*.

Algorithms and computation are fundamental to molecular self-assembly as illustrated in an enormous success of their use in DNA tile self-assembly (see, e.g., [4, 12, 15] and references therein). The Sierpinski triangle fractal was algorithmically self-assembled even *in vitro* from coalescence of DNA tiles that compute XOR [13]. Cotranscriptional folding exhibits highly sophisticated computational

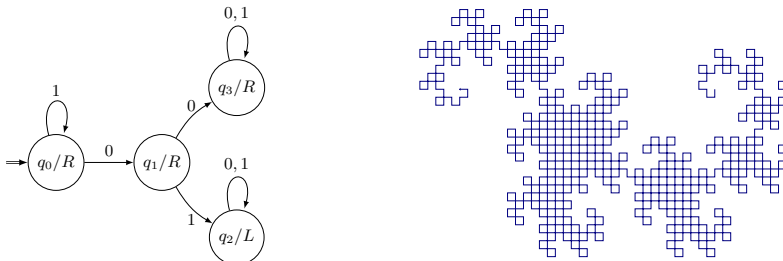
---

<sup>\*</sup> This work is in part supported by JST Program to Disseminate Tenure Tracking System, MEXT, Japan, No. 6F36 and by JSPS KAKENHI Grant-in-Aid for Young Scientists (A) No. 16H05854 to S. S.

<sup>\*\*</sup> Corresponding author



**Fig. 1.** RNA cotranscriptional folding. An RNA polymerase attaches to a template DNA sequence (gray spiral), scans it through, and synthesizes its RNA copy. The RNA sequence begins to fold upon itself immediately as it emerges from polymerase.



**Fig. 2.** (Left) DFAO to output the direction (L/R) of  $i$ -th turn of the Heighway dragon given  $i \geq 0$  in binary from the LSB. (Right) The first  $2^{10}-1$  turns of the dragon.

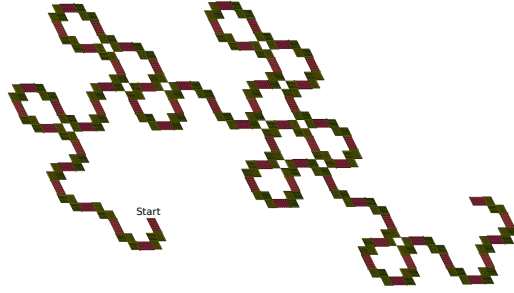
and algorithmic behaviors as well. Indeed, fluoride riboswitches in *Bacillus cereus* bacteria cotranscriptionally fold into a terminator stem or do not, in order to regulate gene expression [14]. This is just one example but should be enough to signify both the context-sensitivity of cotranscriptional folding and shapes thus self-assembled. Geary et al. have proved the capability of context-sensitivity to count in binary using a novel mathematical model of cotranscriptional folding called *oritatami system* (abbreviated as OS) [6].

We shall initiate theoretical study on algorithmic self-assembly of shapes by cotranscriptional folding using oritadaki system. Sierpinski triangle would allow our study to borrow rich insights from the DNA tile self-assembly. However, in order to cut directly to the heart of algorithmic self-assembly by cotranscriptional folding, shapes of choice should be traversable somehow algorithmically. One such way is to feed a turtle program (see [1]) with an *automatic sequence* as commands (drawing a line segment, rotation, etc.), whose  $i$ -th bit can be obtained by giving  $i$  in binary from the least significant bit (LSB) to one DFA with output (DFAO) [3]. Shapes thus describable include the Heighway dragon [3] and von Koch curve [10]. A DFAO for the Heighway dragon is illustrated in Fig. 2. It outputs the following sequence, given  $i = 0, 1, 2, \dots$  in binary:

$$P = \text{RRLRRLLRRRLRLLRRRLRRLLRLL} \cdots$$

(The notation  $P$  is after its appellative *paperfolding sequence* [3]). For instance, given  $i = 2$  in binary from the LSB as 01, the DFAO transitions as  $q_0 \rightarrow q_1 \rightarrow q_2$  and hence  $P[2] = L$ . A turtle should interpret an L (resp. R) as “move forward by

unit distance and turn left (resp. right) 90 degrees.” Any portion of the dragon can be represented by a factor of  $P$ ; for instance, Fig. 2 (Right) depicts the portion  $P[0..1022]$ , i.e., the first  $2^{10} - 1$  turns of the dragon.



**Fig. 3.** The portion  $P[0..62]$  of the Heighway dragon folded by the proposed oritatami system.

proposed in [6] so that it increments a given count  $i$  exactly by 1 while folding into a (red) line segment. At the end of the segment comes a DFAO module, which computes the turn direction  $P[i]$  and propagates it along with the count  $i$  to the next module for turn. A (green) L-shaped block is the turning module. It is a concatenation of three bit-sequence bifurcators, each of which folds into a rhombus, bifurcates  $i$  leftward as well as rightward, and guides further folding according to the turning direction.

The generic design proves the next theorem (for terminologies, see Sect. 2).

**Theorem 1.** *For any finite portion  $P[i..j]$  of the Heighway dragon, there exist a scaling factor  $c \in \mathbb{N}^+$  and a deterministic cyclic oritatami system of delay 3 and arity 3 that weakly folds into the  $c$ -rhombus scaling of  $P[i..j]$ .*

A JavaScript program to run this OS is available at <https://wolves13.github.io>.

## 2 Preliminaries

Let  $\Sigma$  be a set of types of abstract molecules, or *beads*, and  $\Sigma^*$  be the set of finite sequences of beads. A bead of type  $a \in \Sigma$  is called an  $a$ -bead. Let  $w = b_1 b_2 \dots b_n \in \Sigma^*$  be a string of length  $n$  for some integer  $n$  and bead types  $b_1, \dots, b_n \in \Sigma$ . The *length* of  $w$  is denoted by  $|w|$ , that is,  $|w| = n$ . For two indices  $i, j$  with  $1 \leq i \leq j \leq n$ , we let  $w[i..j]$  refer to the subsequence  $b_i b_{i+1} \dots b_{j-1} b_j$ ; if  $i = j$ , then we simplify  $w[i..i]$  as  $w[i]$ . For  $k \geq 1$ ,  $w[1..k]$  is called a *prefix* of  $w$ .

Oritatami systems fold their transcript, a sequence of beads, over the triangular grid graph  $\mathbb{T} = (V, E)$  as suggested in Fig. ?? cotranscriptionally based on hydrogen-bond-based interactions (*h-interactions* for short) which the system allow for between adjacent beads of particular types. A directed path  $P =$

In this paper, we propose a generic design of oritatami system for the algorithmic cotranscriptional folding of an arbitrary finite portion of the Heighway dragon. Fig. 3 shows the portion  $P[0..62]$  thus folded (the dragon is slanted but this is because the OS operates on the triangular grid). The transcript is a repetition of three modules: a catenation of binary counters, DFAO module, and turning module. The counter is a technical modification of the one proposed

$p_1 p_2 \cdots p_n$  in  $\mathbb{T}$  is a sequence of *pairwise-distinct* points  $p_1, p_2, \dots, p_n \in V$  such that  $\{p_i, p_{i+1}\} \in E$  for all  $1 \leq i < n$ . Its  $i$ -th point is referred to as  $P[i]$ .

A *conformation*  $C$  is a triple  $(P, w, H)$  of a directed path  $P$  in  $\mathbb{T}$ ,  $w \in \Sigma^*$  of the same length as  $P$ , and a set of h-interactions  $H \subseteq \{\{i, j\} \mid 1 \leq i, i+2 \leq j, \{P[i], P[j]\} \in E\}$ . This is to be interpreted as the sequence  $w$  being folded in such a manner that its  $i$ -th bead  $w[i]$  is placed on the  $i$ -th point  $P[i]$  along the path and there is an h-interaction between the  $i$ -th and  $j$ -th beads if and only if  $\{i, j\} \in H$ . The condition  $i+2 \leq j$  represents the topological restriction that two consecutive beads along the path cannot form an h-interaction between them. A *rule set*  $\mathcal{H} \subseteq \Sigma \times \Sigma$  is a symmetric relation over the set of pairs of bead types, that is, for all bead types  $a, b \in \Sigma$ ,  $(a, b) \in \mathcal{H}$  implies  $(b, a) \in \mathcal{H}$ . An h-interaction  $\{i, j\} \in H$  is *valid with respect to*  $\mathcal{H}$ , or simply  *$\mathcal{H}$ -valid*, if  $(w[i], w[j]) \in \mathcal{H}$ . This conformation  $C$  is  *$\mathcal{H}$ -valid* if all of its h-interactions are  $\mathcal{H}$ -valid. For an integer  $\alpha \geq 1$ ,  $C$  is of *arity*  $\alpha$  if it contains a bead that forms  $\alpha$  h-interactions and no bead of  $C$  forms more. By  $\mathcal{C}_{\leq \alpha}$ , we denote the set of all conformations of arity at most  $\alpha$ .

Oritatami systems grow conformations by elongating them under their own rule set. Given a rule set  $\mathcal{H}$  and an  $\mathcal{H}$ -valid finite conformation  $C_1 = (P, w, H)$ , we say that another conformation  $C_2$  is an *elongation of  $C_1$  by a bead  $b \in \Sigma$* , written as  $C_1 \xrightarrow{\mathcal{H}}_b C_2$ , if  $C_2 = (Pp, wb, H \cup H')$  for some point  $p$  not along the path  $P$  and set of h-interactions  $H' \subseteq \{\{i, |w|+1\} \mid 1 \leq i < |w|, \{P[i], p\} \in E, (w[i], b) \in \mathcal{H}\}$ , which can be empty. Note that  $C_2$  is also  $\mathcal{H}$ -valid. This operation is recursively extended to the elongation by a finite sequence of beads as: for any conformation  $C$ ,  $C \xrightarrow{\mathcal{H}}_{\lambda}^* C$ ; and for a finite sequence of beads  $w \in \Sigma^*$  and a bead  $b \in \Sigma$ , a conformation  $C_1$  is elongated to a conformation  $C_2$  by  $wb$ , written as  $C_1 \xrightarrow{\mathcal{H}}_{wb}^* C_2$ , if there is a conformation  $C'$  that satisfies  $C_1 \xrightarrow{\mathcal{H}}_w^* C'$  and  $C' \xrightarrow{\mathcal{H}}_b C_2$ .

A finite *oritatami system* (OS) is a 5-tuple  $\Xi = (\mathcal{H}, \alpha, \delta, \sigma, w)$ , where  $\mathcal{H}$  is a rule set,  $\alpha$  is an arity,  $\delta \geq 1$  is a parameter called the *delay*,  $\sigma$  is an initial  $\mathcal{H}$ -valid conformation of arity  $\alpha$  called the *seed*, upon which its finite *transcript*  $w \in \Sigma^*$  is to be folded by stabilizing beads of  $w$  one at a time so as to minimize energy collaboratively with the succeeding  $\delta - 1$  nascent beads. The energy of a conformation  $C = (P, w, H)$ , denoted by  $\Delta G(C)$ , is defined to be  $-|H|$ ; the more h-interactions a conformation has, the more stable it gets. The set  $\mathcal{F}(\Xi)$  of conformations *foldable* by this system is recursively defined as: the seed  $\sigma$  is in  $\mathcal{F}(\Xi)$ ; and provided that an elongation  $C_i$  of  $\sigma$  by the prefix  $w[1..i]$  be foldable (i.e.,  $C_0 = \sigma$ ), its further elongation  $C_{i+1}$  by the next bead  $w[i+1]$  is foldable if

$$C_{i+1} \in \arg \min_{C \in \mathcal{C}_{\leq \alpha} \text{ s.t. } C_i \xrightarrow{\mathcal{H}}_{w[i+1]} C} \min \left\{ \Delta G(C') \mid C \xrightarrow{\mathcal{H}}_{w[i+2..i+k]}^* C', k \leq \delta, C' \in \mathcal{C}_{\leq \alpha} \right\}. \quad (1)$$

We say that the bead  $w[i+1]$  and the h-interactions it forms are *stabilized* according to  $C_{i+1}$ . Note that an arity- $\alpha$  OS cannot fold any conformation of arity larger than  $\alpha$ . A conformation foldable by  $\Xi$  is *terminal* if none of its elongations is foldable by  $\Xi$ . The OS  $\Xi$  is *deterministic* if for all  $i \geq 0$ , there

exists at most one  $C_{i+1}$  that satisfies (1). Thus, a deterministic OS folds into a unique terminal conformation. An OS is cyclic if its transcript is of the form  $u^i u_p$  for some  $i \geq 2$  and a prefix  $u_p$  of  $u$ . The cyclic OSs are considered to be one of the practical classes of OS because a periodic RNA transcript is likely to be transcribed out of a circular DNA sequence [5].

Denote the points in  $\mathbb{R}^2$  that correspond to vertices of  $\mathbb{T}$  by  $\mathbb{Z}_\Delta^2$ . A *shape*  $S$  is a set of points on the triangular grid. For an integer  $c \geq 1$ , let  $Rhomb_c = \{(x, y) \in \mathbb{Z}_\Delta^2 \mid x, y \leq c\}$ . Let  $S' = \{(cx, cy) \mid (x, y) \in S\}$ . The *c-rhombus scaling* of  $S$  is the union over all  $\mathbf{p} \in S'$  of sets of points  $Rhomb_c + \mathbf{p} = \{\mathbf{v} \in \mathbb{R}^2 \mid \mathbf{v} = \mathbf{r} + \mathbf{p} \text{ for some } \mathbf{r} \in Rhomb_c\}$ . Let  $\diamond_c(S)$  be the *c-rhombus scaling* of  $S$ . We say that an OS  $\Xi$  *weakly folds* (or “self-assembles”)  $\diamond_c(S)$  if every terminal assembly of  $\Xi$  puts at least one bead in  $Rhomb_c + \mathbf{p}$  for all  $\mathbf{p} \in S$  and no bead in  $Rhomb_c + \mathbf{q}$  for all  $\mathbf{q} \notin S$ .

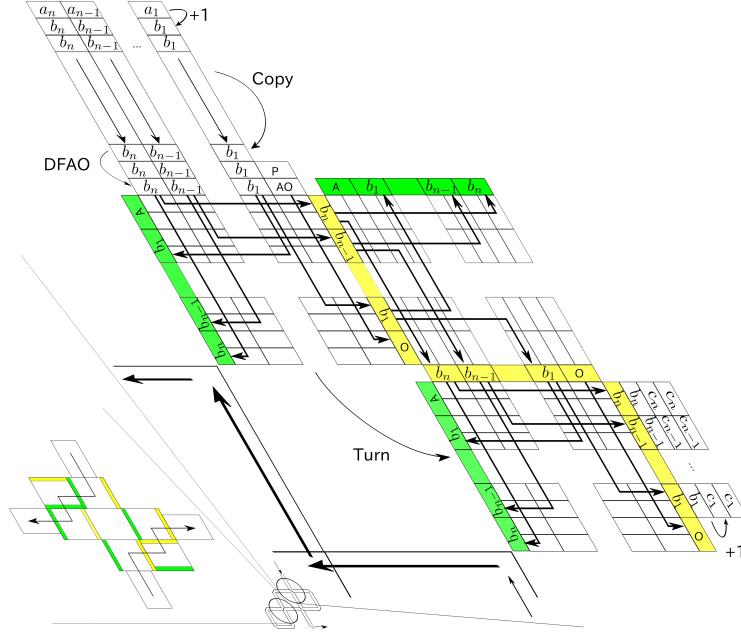
### 3 Folding the $n$ -bit Heighway dragon

We propose a generic design of deterministic OS that allows us to fold an arbitrary finite portion  $P[j_1..j_2]$  of the slanted Heighway dragon. Independently of  $j_1, j_2$ , the design sets both delay and arity to 3 and employs 567 bead types with a fixed rule set  $\mathcal{H}$  (some of the bead types might be saved but not easily due to the NP-hardness of minimizing the number of bead types [8]). It also challenges to make the resulting OS cyclic. Otherwise, one could simply design left and right-turn modules and concatenate their copies according to the (non-periodic) sequence  $P$ . Such a “hardcoding” however goes against the spirit of algorithmic self-assembly. Furthermore, it is highly unlikely that an OS for the infinite Heighway dragon, if any, could adopt this approach in order to be describable by a finite mean.

One period is semantically divided into six parts (*modules*) as  $CD_vTCD_hT$ , where

- $C$  is called a *counter module*, which increments  $i$  by 1 and propagates it;
- $D_v$  and  $D_h$  are called a *DFAO module*, which computes  $P[i]$  and interprets it properly as A(cute) or O(btuse);
- $T$  is called a *turning module*, which makes a turn according to A/O.

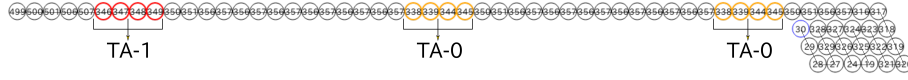
The first  $C$  and  $D_v$  modules fold into a vertical line segment, while the second  $C$  and  $D_h$  fold into the next line segment, which is guaranteed to be horizontal. The DFAO modules  $D_v, D_h$  differ only in their way to interpret their intermediate outcome  $P[i]$ . The slanted dragon involves two types of left turn as well as two types of right turn: acute and obtuse. Observe that after (slanted) vertical line segments, the dragon turns left obtusely and right acutely, whereas after horizontal ones, it turns left acutely and right obtusely. Moreover, vertical and horizontal segments alternate on the dragon. Therefore, it suffices for  $D_v$  and  $D_h$  to compute  $P[i] \in \{L, R\}$  in the same way and interpret it as A(cute)/O(btuse) in an opposite fashion as  $D_v$  converts  $L$  into  $O$  and  $R$  into  $A$ , while  $D_h$  converts  $L$  into  $A$  and  $R$  into  $O$ .



**Fig. 4.** Folding of one segment plus turn of the Highway dragon, flow of information through it, and the two ways of collision avoidance between two turns.

Before explaining the modules, one issue intrinsic to the folding by oritatami systems should be pointed. It rises when the dragon makes a turn where it has already turned before. By definition, OSs cannot put a bead anywhere occupied by another bead. Being scaled sufficiently large, a rhombus corresponding to a point affords two turning modules, which otherwise collide, as long as they fold into an L-shape (Figs. 3, 4). A turning module has its three bifurcator submodules direct further folding obtusely one after another guided by the signal O from the previous DFAO module, as colored in yellow in Figs. 4, ??, ??, or acutely one after another by A, as colored in green, and folds into two L-shape conformations. Note that two turns which share a point are both acute or both obtuse.

Having outlined the generic design, now we explain how the design implements an OS for a specific target portion  $P[j_1..j_2]$ , or more precisely, how the modules  $C, D_v, D_h, T$  and their submodules are implemented, interlocked with each other, and collaborate. Let  $n = \min\{m \mid j_2 < 2^m\}$ . The length of  $C, T$ , or more precisely, of their transcript, are of length  $O(n^2)$  while  $D_v, D_h$  are of length  $O(n)$ . Thus, the period of the OS is of length proportional to  $n^2$ . Each of these modules consists of submodules of constant size. Using a simulator developed for [9], we verified that all of the submodules fold as intended in all possible environments, which are abstracted in Figs. 4, ??, and ??.



**Fig. 5.** The seed conformation for the 3-bit Heighway dragon encoding the initial count 100.

*Seed conformation* encodes the initial count  $j_1$  in its binary representation as

$$499 \rightarrow 500 \rightarrow 501 \rightarrow 506 \rightarrow 507 \rightarrow \bigodot_{k=n}^1 (\mathbf{w}_{t,b_k} \rightarrow 350 \rightarrow 351 \rightarrow (356 \rightarrow 357 \rightarrow)^6) \quad (2)$$

where  $\mathbf{w}_{t,0} = 338 \rightarrow 339 \rightarrow 344 \rightarrow 345$  and  $\mathbf{w}_{t,1} = 346 \rightarrow 347 \rightarrow 348 \rightarrow 349$ .

*Counter module* is borrowed from [6] with technical modification to let it operate in the dynamics (1), which is more prevailing [8, 9, 11] though less tractable. We hence leave all its details to the Appendix (Sect. B), and just describe its input and output. It takes the current count  $i$  formatted as (2), which is fed by the seed or by the previous turner module, increments the count by 1 unless it is preceded by the seed, and outputs the resulting count in its binary representation  $a_n a_{n-1} \cdots a_1$  in the following format:

$$??? \quad (3)$$

where  $\mathbf{w}_{c,0} = 51 \rightarrow 52 \rightarrow 57 \rightarrow 58 \rightarrow 63 \rightarrow 64$  and  $\mathbf{w}_{c,1} = 59 \rightarrow 60 \rightarrow 61 \rightarrow 62 \rightarrow 63 \rightarrow 64$ .

*DFAO modules*  $D_v, D_h$  receive the current count  $i$  in the format (3) from the previous counter module, compute  $P[i]$  and interpret it as A or O properly, and output the interpretation along with the count  $i$  in the following format:

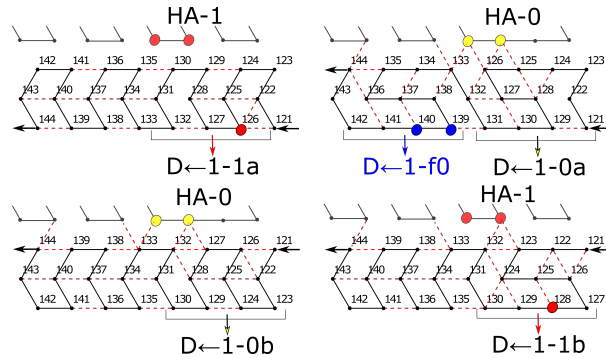
$$??? \quad (4)$$

What the DFAO in Fig. 2 really does for computing  $P[i]$  is to search for the first 0 from the LSB and check whether it is followed by 0 ( $P[i] = R$ ) or by 1 ( $P[i] = L$ ). See Fig. ?? .  $D_v$  employs the six submodules: DFAO-zig1, -zag1, -zig2, -zag2, PFS, and  $\text{AO}_v$  (resp.  $\text{AO}_h$ ), which are interleaved by spacers, as well as corners, which guide the transcript into two zigzags and one more zig. The first zigzag is for the search, the second zigzag is for the check, and at the beginning of the third zig, as shown in Fig. ??,  $\text{AO}_v$  interprets the output L of PFS as O and R as A, while  $\text{AO}_h$  interprets them the other way around otherwise.

*Turner module*

## References

1. Harold Abelson and Andrea diSessa. *Turtle*



**Fig. 6.** The four bricks of Dzig1: (top) Dzig1-1 and Dzig1-f0; (bottom) Dzig1-20 and Dzig1-21.

- Geometry The Computer as a Medium for Exploring Mathematics*. MIT Press Series in Artificial Intelligence. The MIT Press, 1981.
2. Bruce Alberts, Alexander Johnson, Julian Lewis, David Morgan, Martin Raff, Keith Roberts, and Peter Walter. *Molecular Biology of the Cell*. Garland Science, 6th edition, 2014.
  3. Jean-Paul Allouche and Jeffrey Shallit. *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge University Press, 2003.
  4. David Doty. Theory of algorithmic self-assembly. *Communications of the ACM*, 55(12):78–88, 2012.
  5. C. Geary and E. S. Andersen. Design principles for single-stranded RNA origami structures. In *Proc. DNA20*, volume LNCS 8727, pages 1–19. Springer, 2014.
  6. C. Geary, P.-É. Meunier, N. Schabanel, and S. Seki. Programming biomolecules that fold greedily during transcription. In *Proc. MFCS2016*, LIPIcs 58, pages 43:1–43:14, 2016.
  7. Cody Geary, Paul W. K. Rothemund, and Ebbe S. Andersen. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science*,



- 345(6198):799–804,  
2014.
8. Y-S. Han and H. Kim. Ruleset optimization on isomorphic oritatami systems. In *Proc. DNA23*, LNCS 10467, pages 33–45. Springer, 2017.
9. Yo-Sub Han, Hwee Kim, Makoto Ota, and Shinnosuke Seki. Non-deterministic seedless oritatami systems and hardness of testing their equivalence. In *Proc. 22nd International Conference on DNA Computing and Molecular Programming (DNA22)*, volume 9818 of *LNCS*, pages 19–34. Springer, 2016.
10. Jun Ma and Judy Holdener. When Thue-Morse meets Koch. *Fractals*, 13:191–206, 2005.
11. Makoto Ota and Shinnosuke Seki. Ruleset design problems for oritatami systems. *Theoretical Computer Science*, 671:26–35, 2017.
12. Matthew J. Patitz. Self-assembly of fractals. In *Encyclopedia of Algorithms*, pages 1918–1922. Springer, 2016.
13. Paul W. K. Rothemund, Nick Papadakis, and Erik Winfree. Algorithmic self-assembly of DNA Sierpinski triangle. *PLoS Biology*, 2(12):e424, 2004.
14. Kyle E. Watters, Eric J. Strobel, Angela M. Yu,

- John T. Lis, and Julius B. Lucks. Co-transcriptional folding of a riboswitch at nucleotide resolution. *Nature Structural & Molecular Biology*, 23(12):1124–1133, 2016.
15. Erik Winfree. *Algorithmic Self-Assembly of DNA*. PhD thesis, California Institute of Technology, June 1998.

## A Module automaton

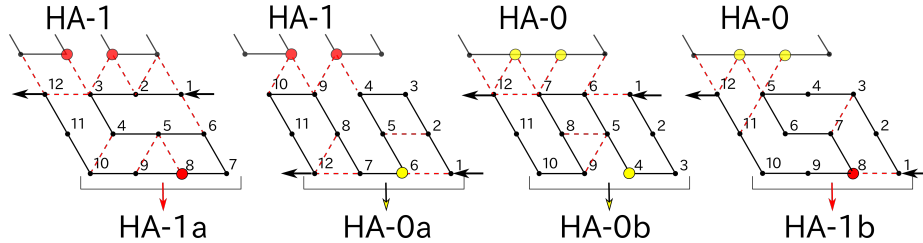


**Fig. 7.** Module automaton

## B Counter module

One period of the transcript of the  $n$ -bit counter module is the concatenation of the following submodules:

- $n$  submodules  $C_{HA}$ ; the half-adder;
- $n - 1$  submodules  $C_{SPzig}$  interleaved between the half-adders;
- Submodule  $C_{LT}$ ; the left-turner;
- $n$  submodules  $C_F$ , the formatter;
- $n - 1$  submodules  $C_{SPzag}$  interleaved between the formatters;
- Submodule  $C_{RT}$ ; the right-turner.



**Fig. 8.** All conformations of the half-adder. The first and third are diverted to implement the body-lpx2 component of the turning module.

### B.1 Transcript