Self-attraction Removal from Oritatami Systems

Yo-Sub Han, Hwee Kim, Trent Rogers, Shinnosuke Seki

Self-attraction removal from oritatami systems

Yo-Sub Han[1⋆], Hwee Kim[1], Trent A. Rogers[2⋆⋆], and Shinnosuke Seki[3⋆⋆⋆†]

[1] Department of Computer Science, Yonsei University
50 Yonsei-Ro, Seodaemun-Gu, Seoul 03722, Republic of Korea
kimhwee,emmous@yonsei.ac.kr
[2] Department of Computer Sceince and Computer Engineering, University of Arkansas, Fayetteville, AR, 72701, USA tar003@uark.edu
[3] Department of Computer and Network Engineering, University of Electro-Communications, 1-5-1 Chofugaoka, Chofu, Tokyo, 1828585, Japan
s.seki@uec.ac.jp

Abstract. RNA cotranscriptional folding refers to the phenomenon in which an RNA transcript folds upon itself while being synthesized (transcribed). Oritatami is a computational model of this phenomenon, which lets its transcript, a sequence of beads (abstract molecules) fold cotranscriptionally via interactions between beads according to its ruleset. In this paper, we study the problem of removing self-attractions, which lets a bead interact with another bead of the same kind, from a given oritatami system without changing its behavior. We provide an algorithm for that with overhead linear in the delay parameter, which should be considerably smaller than the length of its transcript. We also show that this overhead is tight.

1 Introduction

Self-assembly is the process by which relatively simple components coalesce to form intricate and complex structures. Studying self-assembling systems can provide us with insights into everything from designing nanophotonic devices [11] to the origins of life [14]. A number of theoretical models of self-assembly have been proposed [2,12,15] and some models of self-assembly have been implemented in the laboratory to algorithmically build structures out of DNA [3,6,13]. One proposed model of self-assembly is called Oritatami (folding in Japanese) [8], which seeks to capture the fundamental dynamics of cotranscriptional folding. Transcription is the first step in gene expression in which an RNA polymerase

⋆ Supported by International Cooperation Program (2017K2A9A2A08000270) and Basic Science Research Program (2015R1D1A1A01060097) by NRF of Korea.
⋆⋆ This author's research was supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1450079, and National Science Foundation grants CAREER-1553166 and CCF-1422152.
⋆⋆⋆ In part supported by JST Program to Disseminate Tenure Tracking System No. 6F36 and JSPS KAKENHI Grant-in-Aid for Young Scientists (A) No. 16H05854
† Corresponding author

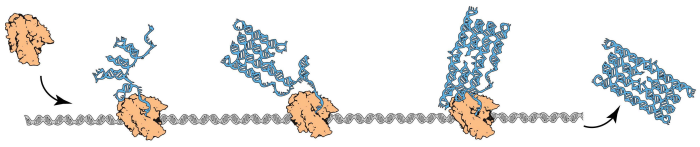

Fig. 1. RNA origami [9]: the cotranscriptional folding of an RNA tile in a laboratory. The gray spiral is a DNA sequence (template), the orange molecule attaching is the RNA polymerase, and the blue output of the polymerase is the RNA transcript.

attaches to a DNA sequence and sequentially produces RNA nucleotides (`A`, `C`, `G`, `U`) (see Fig. 1). Cotranscriptional folding refers to the folding of RNA during transcription. That is, as the RNA is transcribed, its nucleotides interact with each other via hydrogen bonds, resulting in the folding of the RNA.

Geary, Rothemund, and Andersen harnessed the power of cotranscriptional folding in order to self-assemble nanoscale tiles out of RNA (*RNA origami* [9]). Oritatami is a theoretical model to study the computational aspect of cotranscriptional folding. It models a single strand of RNA as a "strand" of abstract molecules, or *beads*. Each bead is of a certain type taken out of a finite alphabet Σ. The bead types along with a set of attraction rules specify which beads are attracted to one another. In addition, each oritatami system has a parameter called the *delay* δ. It models the speed at which cotranscriptional folding occurs. The folding of an oritatami system proceeds by "looking ahead" at the next δ beads on the strand and folds them so as to create the largest number of bonds. We can see in [1] an oritatami system of delay 3 fold a motif called *glider*.

The class of oritatami systems implementable in the laboratory by the cotranscriptional folding of RNA is limited by the properties of RNA. More specifically, the attraction rules of oritatami systems are limited by the types of allowable interactions between RNA nucleotides. Therefore, laboratory implementation may require to alter the system so that it fits certain criteria required for experimental implementation. For example, a physical implementation of oritatami systems might rely on the Watson-Crick complementarity (`G`-`C` and `A`-`U`). If we wanted to implement an oritatami system in this setting which had a rule specifying a bead type is attracted to itself, the self-attraction would need to be removed.

In addition to providing tools to make oritatami systems physically realizable, our results are the first set of results to show an oritatami system being "simulated" by another oritatami system with a different set of properties. Simulation has played an important role in determining the relative power of classes of systems in tile assembly and determining how classes of tile systems relate to each other [4,7]. In addition, the notion of simulation in tile assembly has given rise to a rich study of intrinsic universality [5,10,16] which has provided us with a deeper understanding of tile assembly. The results in this paper are a first step towards using "simulation" to develop a better understanding of the model.

In this paper, we examine the removal of rules specifying that a bead type is attracted to itself, which we call self-attraction rules, from Oritatami systems.

Given a system Ξ, the goal of self-attraction removal is to create another system Ξ' such that Ξ' behaves in the same way as Ξ, Ξ' produces the same set of conformations as Ξ, and Ξ' does not contain any self-attraction rules.

2 Preliminaries

Let Σ be a finite set of bead types. A bead of type a is called an a-bead. By Σ^* (resp. Σ^ω), we denote the set of finite (one-way infinite) sequences of bead types in Σ. A sequence $w \in \Sigma^*$ can be represented as $w = b_1 b_2 \cdots b_n$ for some $n \geq 0$ and bead types $b_1, b_2, \ldots, b_n \in \Sigma$, where n is the *length* of w and denoted by $|w|$. The sequence of length 0 is denoted by λ. For $1 \leq i, j \leq n$, the subsequence of w ranging from the i-th bead to j-th bead is denoted by $w[i..j]$, that is, $w[i..j] = b_i b_{i+1} \cdots b_j$; $w[i..j] = \lambda$ if $i > j$. This notation is simplified as $w[i]$ when $j = i$, referring to the i-th bead of w. For $k \geq 1$, $w[1..k]$ is called a *prefix*.

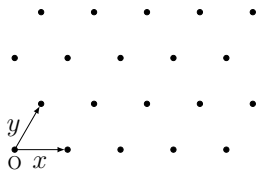

Fig. 2. Triangular lattice.

Oritatami systems fold their transcript, a sequence of beads, over the triangular lattice (Fig. 2) cotranscriptionally by letting nascent beads form as many hydrogen-bond-based interactions (*h-interactions*) as possible according to their own interaction rules. Let $\mathbb{T} = (V, E)$ be the triangular grid graph. A directed path $P = p_1 p_2 \cdots$ in $\mathbb{T}$ is a possibly-infinite sequence of pairwise-distinct points $p_1, p_2, \ldots \in V$ such that $\{p_i, p_{i+1}\} \in E$ for all $i \geq 1$. Its i-th point p_i is referred to as $P[i]$. Provided it is finite, by $|P|$, we denote the number of points in it.

A conformation instance, or *configuration*, is a triple (P, w, H) of a directed path P in $\mathbb{T}$, $w \in \Sigma^* \cup \Sigma^\omega$, and a set $H \subseteq \big\{(i, j) \;\big|\; 1 \leq i, i+2 \leq j, \{P[i], P[j]\} \in E\big\}$ of h-interactions. 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 the i-th bead interacts with the j-th bead if and only if $(i, j) \in H$. Configurations (P_1, w_1, H_1) and (P_2, w_2, H_2) are *congruent* provided $w_1 = w_2$, $H_1 = H_2$, and P_1 can be transformed into P_2 by a combination of a translation, a reflection, and rotations by 60 degrees. Given a configuration (P, w, H), the set of all configurations congruent to it, denoted by $[(P, w, H)]$, is called its *conformation*. We refer to w as its *primary structure*. A *ruleset* $\mathcal{H} \subseteq \Sigma \times \Sigma$ 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 called *valid with respect to* $\mathcal{H}$, or simply $\mathcal{H}$*-valid*, if $(w[i], w[j]) \in \mathcal{H}$. This conformation is $\mathcal{H}$*-valid* if all of its h-interactions are $\mathcal{H}$-valid. For $\alpha \geq 1$, this conformation is *of arity* α if the maximum number of h-interactions per bead is α, that is, if for any $k \geq 1$, $\big|\{i \mid (i, k) \in H\}\big| + \big|\{j \mid (k, j) \in H\}\big| \leq \alpha$ and the equation holds for some k. By $\mathcal{C}_{\leq \alpha}$, we denote the set of all conformations of arity at most α.

Oritatami systems grow conformations by elongating them under their own ruleset. For a finite conformation C_1, another finite conformation C_2 is an *elongation* of C_1 by a bead $b \in \Sigma$ under a ruleset $\mathcal{H}$, written as $C_1 \xrightarrow{\mathcal{H}}_b C_2$,

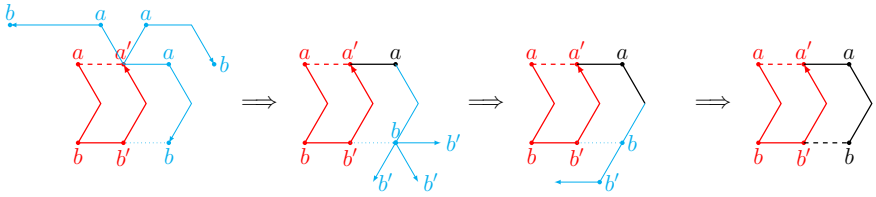

Fig. 3. Folding of a glider motif by a delay-3 deterministic oritatami system.

if there exists a configuration (P, w, H) of C_1 such that $(Pp, wb, H \cup H')$ is a configuration of C_2, where $p \in V$ is a point not along the path P and $H' \subseteq \big\{(i, |P|+1) \;\big|\; 1 \leq i \leq |P|-1, \{P[i], p\} \in E, (w[i], b) \in \mathcal{H}\big\}$. 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 $C' \xrightarrow{\mathcal{H}}_{b} C_2$.

An *oritatami system* over an alphabet Σ is a 6-tuple $\Xi = (\Sigma, w, \mathcal{H}, \delta, \alpha, \sigma)$, where $\mathcal{H}$ is a *ruleset*, $\delta \geq 1$ is a parameter called *delay*, and σ is an $\mathcal{H}$-valid initial *seed* conformation of arity at most α, upon which its *transcript* $w \in \Sigma^* \cup \Sigma^\omega$ 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 $U(C)$ of a conformation $C = [(P, w, H)]$ is defined to be $-|H|$; the more h-interactions a conformation has, the more stable it becomes. The set $\mathcal{F}(\Xi)$ of conformations *foldable* by this system is recursively defined as follows: the seed σ is in $\mathcal{F}(\Xi)$; and provided that an elongation C_i of σ by the prefix $w[1..i]$ is foldable (i.e., $C_0 = \sigma$), its further elongation C_{i+1} by the next bead $w[i+1]$ is foldable if

$$C_{i+1} \in \operatorname*{argmin}_{\substack{C \in \mathcal{C}_{\leq \alpha} \text{ s.t.} \\ C_i \xrightarrow{\mathcal{H}}_{w[i+1]} C}} \min \left\{ U(C') \;\middle|\; C \xrightarrow{\mathcal{H}}{}^{*}_{w[i+2..i+k]} C', k \leq \delta, C' \in \mathcal{C}_{\leq \alpha} \right\}. \qquad (1)$$

The bead $w[i+1]$ and h-interactions it forms are said to have been *stabilized* according to C_{i+1}. A conformation foldable by Ξ is *terminal* if none of its elongations is foldable by Ξ. The oritatami system Ξ is *deterministic* if for all $i \geq 0$, there exists at most one C_{i+1} that satisfies (1). Thus, a deterministic oritatami system folds into a unique terminal conformation.

Example 1 ([1]). See Fig. 3 for a delay-3 oritatami system Ξ to fold a motif called glider. Its transcript is a repetition of $a \bullet bb' \bullet a$ and its ruleset is $\{(a, a'), (b, b')\}$. Its seed is colored in red. The first 3 beads, $a \bullet b$, are transcribed and elongate the seed by the seed in all possible ways. The a-bead cannot form any h-interaction or the second bead is inert w.r.t. the ruleset. The third bead, b, can interact with the b'-bead in the seed but for that, the a-bead must be located to the east of the previous a'-bead; it is thus stabilized there. Then the next bead, b', is transcribed. After the three steps, the third bead, b, is stabilized. It is not until then that its h-interaction with the b'-bead is also stabilized.

2.1 Self-attraction-free oritatami system

A bead type $a \in \Sigma$ is *self-attractive* according to a ruleset $\mathcal{H}$ if $(a, a) \in \mathcal{H}$. An oritatami system is *free from self-attraction*, or *self-attraction-free*, if no bead type in its alphabet is self-attractive according to its ruleset.

We formulate the problem of removing self-attraction from a given oritatami system without changing the behavior in Problem 1. An isomorphism between conformations must be introduced. Conformations C_1 and C_2 are *isomorphic* if there exist an instance (P_1, w_1, H_1) of C_1 and an instance (P_2, w_2, H_2) of C_2 such that $P_1 = P_2$ and $H_1 = H_2$.

Problem 1 (Self-attraction removal). Let Ξ be an oritatami system. Design a self-attraction-free oritatami system Ξ' such that a conformation is foldable by Ξ if and only if the isomorphic conformation is foldable by Ξ'.

Replacing the beads in the seed and transcript with pairwise distinct bead types provides a trivial solution, but is not desirable.

The following approach to Problem 1 called *bead type modification* is promising: When a bead type a is found self-attractive, it is modified as $a_1, a_2, \ldots, a_t$. The ruleset is then modified so as not to allow for any self-attraction but to let a bead interact with any of its modifications; for example, if we make three copies a_1, a_2, a_3 of a bead a, none of $(a_1, a_1), (a_2, a_2), (a_3, a_3)$ is in the ruleset but all of (a_1, a_2), (a_2, a_3), and (a_3, a_1) are in it. Moreover, these modifications are to look to non-a beads as if they were identical, or the ruleset should be modified so. More precisely, for any $b \neq a$, the modified ruleset should include the rules $(a_1, b), (a_2, b), (a_3, b)$ if and only if (a, b) is in the original ruleset.

We propose a subproblem of Problem 1 based on this approach and establish the measure for the efficiency of the modification. It employs a *subscript-erasing homomorphism* $h : \Sigma' \to \Sigma$ defined as $h(x_i) = x$ for $x \in \Sigma$ and $i \geq 1$.

Problem 2 (Self-attraction removal by bead type modification). Solve Problem 1 on an oritatami system $\Xi = (\Sigma, w, \mathcal{H}, \delta, \alpha, \sigma)$ so that the resulting self-attraction-free system $\Xi' = (\Sigma', w', \mathcal{H}', \delta, \alpha, \sigma')$ also satisfies the following properties:

1. $\Sigma' = \{x_i \mid x \in \Sigma, 1 \leq i \leq c(x)\}$ for a positive integer $c(x)$ for x.
2. $\mathcal{H}' = \{(x_i, y_j) \mid (x, y) \in \mathcal{H}, x \neq y\} \cup \{(x_i, x_j) \mid (x, x) \in \mathcal{H}, i \neq j\}$.
3. h maps the primary structure of σ' to that of σ and $h(w') = w$.

We use the *copying ratio* $c = \max c(x)$ to measure the efficiency of Ξ'.

3 Bead type modification based on the event horizon

The bead stabilization is a local optimization. By definition, the stabilization of a bead in a delay-δ system is not affected by any bead outside the circle of radius $\delta + 1$ centered at the bead stabilized previously. The circle is so called the *event horizon*. Beads on its circumference can affect the process not geometrically but energetically by interacting with the bead transcribed most recently. The event horizon can encompass at most $3(\delta + 1)(\delta + 2)$ beads.

Algorithm 1 Removing self-attractions from a deterministic finite oritatami system

Require: A given oritatami system $\Xi = (\Sigma, w, \mathcal{H}, \delta, \alpha, \sigma)$ is deterministic.

$c \leftarrow 3(\delta+1)(\delta+2)+1$
$w' \leftarrow w$
$\Sigma' \leftarrow \{a_i \mid a \in \Sigma, 1 \leq i \leq c\}$
for all $(a,a) \in \mathcal{H}$ **do**
  $\mathcal{H}_{\text{temp}} \leftarrow \mathcal{H} \cup \big\{(a_i, a_j), (a_j, a_i) \;\big|\; 1 \leq i < j \leq c\big\}$
end for
for all $(a,b) \in \mathcal{H}$ such that $a \neq b$ **do**
  $\mathcal{H}_{\text{temp}} \leftarrow \mathcal{H}_{\text{temp}} \cup \big\{(a_i, b_j), (b_j, a_i) \;\big|\; 1 \leq i, j \leq c\big\}$
end for
for all $(a,c) \in \mathcal{H}$ **do**
  $\mathcal{H}_{\text{temp}} \leftarrow \mathcal{H}_{\text{temp}} \cup \big\{(a, c_i), (c_i, a) \;\big|\; 1 \leq i \leq c\big\}$
end for
$\sigma' \leftarrow 3\text{-COLOR}(\sigma)$
$\Xi' \leftarrow (\Sigma', w', \mathcal{H}_{\text{temp}}, \delta, \alpha, \sigma')$
for $i = 1$ to $|w|$ **do**
  Simulate Ξ' to compute the event horizon of the $\max(1, i-\delta+1)$-th step, at which the i-th bead is transcribed
  $a \leftarrow w'[i]$
  $m \leftarrow \min \left\{ k \geq 1 \;\middle|\; \begin{array}{l} \text{no } a_k\text{-bead is in the horizon and} \\ \text{for all } \max(1, i-\delta+1) \leq j < i,\ w'[i] \neq a_k \end{array} \right\}$
  $w'[i] \leftarrow a_m$
end for
$\mathcal{H}' \leftarrow \mathcal{H}_{\text{temp}} \setminus \big(\mathcal{H} \cup \big\{(a, c_i), (c_i, a) \;\big|\; 1 \leq i \leq c\big\}\big)$
return The self-attraction-free oritatami system $(\Sigma', w', \mathcal{H}', \delta, \alpha, \sigma')$

The event horizon varies from step to step; in the i-th step is centered at the $i-1$-th bead a_{i-1}, which was stabilized previously. The nascent fragment $a_i a_{i+1} \cdots a_{i+\delta-1}$ folds so as to stabilize a_i most stably inside the horizon. During this search, the bead $a_{i+\delta-1}$, which was just transcribed and hence is at the tip of the fragment, can observe every point in and on the horizon, if any, unless being hindered by other beads geometrically. The region observable by the bead $a_{i+\delta-1}$ is never widening but just narrowing as steps go by.

We use bead type modification along the transcript to remove self-attraction. Prior to the transcription of a bead of self-attractive type $a \in \Sigma$, an event horizon is queried for another a-bead, and if there is, then modified the bead to be transcribed into $a_1 \notin \Sigma$. Later, if another a-bead is about to be transcribed inside a horizon provided with both an a-bead and an a_1-bead, then modify its type as $a_2 \notin \Sigma$; without any a_1-bead around, the new type a_2 need not be introduced but an a_1-bead can be transcribed next. No more than $3(\delta+1)(\delta+2)+1$ modifications are needed per bead type due to the size of an event horizon.

Algorithm 1 is an implementation of the idea for deterministic oritatami systems that are finite in the sense that their transcript is finite. Transient systems

need be simulated in line 16 so that auxiliary rules which let a subscripted bead bind to an unsubscripted one are introduced in line 10 but eliminated in the end.

Lemma 1. *Let $C = [(P, u, H)]$ be a conformation that is not $\mathcal{H}$-valid. For any $v \in h^{-1}(u)$, the isomorphic conformation $C' = [(P, v, H)]$ is not $\mathcal{H}'$-valid.*

Proof. Being $\mathcal{H}$-invalid means that an a-bead interacts with b-bead in C though $(a, b) \notin \mathcal{H}$. Since $v \in h^{-1}(u)$, the corresponding interaction in C' is between an a_i-bead and b_j-bead for some i, j. By definition, $(a_i, b_j) \notin \mathcal{H}'$. □

Corollary 1. *If an elongation of the seed σ by $w[1..i]$ is not $\mathcal{H}$-valid, then its isomorphic conformation obtained by elongating σ' by $w'[1..i]$ is not $\mathcal{H}'$-valid.*

The inverse of the statement in Lemma 1 is not always true. Imagine in the conformation C, an a-bead is bound with another a-bead. Providing these beads with the same subscript results in an $\mathcal{H}'$-invalid conformation. Preventing them from being subscripted identically actually yields a valid conformation.

Lemma 2. *Let $C = [(P, u, H)]$ be an $\mathcal{H}$-valid conformation. For any $v \in h^{-1}(u)$, the conformation $C' = [(P, v, H)]$ is $\mathcal{H}'$-valid if for all $(i, j) \in H$, $h(v[i]) = h(v[j])$ implies $v[i] \neq v[j]$.*

Now we show that the resulting system Ξ' behaves as the given system Ξ does. For $0 \leq i \leq |w|$, let $C_i = [(P_i, w_\sigma w[1..i], H_i)]$ be the unique conformation foldable by Ξ among all the elongations of the seed σ by the transcript's prefix $w[1..i]$, where w_σ is the primary structure of σ. That is, $C_0 = \sigma$. The seed σ' of Ξ' is obtained from σ via the function 3-COLOR, which gives subscripts 0, 1, or 2 to beads of a given conformation so as for adjacent beads not to share a common subscript based on the 3-colorability of the triangular grid graph. It is hence $\mathcal{H}'$-valid due to Lemma 2 so that it is foldable by Ξ'. As an inductive hypothesis, assume that C'_i be the unique conformation foldable by Ξ' among all the elongations of the seed σ' by the subscripted transcript's prefix $w'[1..i]$ and C'_i be isomorphic to C_i. Corollary 1 justifies that the stabilization of the $(i+1)$-th bead in Ξ' counts out any elongation of C'_i by $w'[i+1..i+\delta]$ or by its prefix that is isomorphic to an $\mathcal{H}$-invalid elongation of C_i. On the contrary, an elongation of C'_i by $w'[i+1..i+\delta]$ or by its prefix isomorphic to an $\mathcal{H}$-valid elongation of C_i is $\mathcal{H}'$-valid due to Lemma 2 because line 18 of Algorithm 1 prevents any nascent bead in $w'[i+1..i+\delta]$ from being transcribed in the sight of another bead of identical type. Therefore, only the elongation of σ' by $w'[1..i+1]$ that is isomorphic to C_{i+1} is foldable by Ξ'. This concludes the inductive proof.

4 On deterministic finite oritatami systems

The quadratic copying ratio in Algorithm 1 can be reduced to linear. Algorithm 1 is overly cautious; it forbids an a_i-bead to be transcribed inside a horizon with an a_i-bead because an a_i-bead is not self-attractive while its original was. It suffices to guarantee that in at least one of the most stable elongations of

Algorithm 2 Linear-cost self-attraction removal from a deterministic finite oritatami system

Require: A given oritatami system $\Xi = (\Sigma, w, \mathcal{H}, \delta, \alpha, \sigma)$ is deterministic.
1: $c \leftarrow 4\delta + 2$
2: Run Algorithm 1 from line 2 up to line 14
3: Simulate Ξ and arbitrarily choose one representative $E_k = (P_k, w_k, H_k)$ among the most stable elongations of C_{k-1} by $w[k..k+\delta-1]$ for all $1 \leq k \leq |w|$.
4: **for** $i = 1$ **to** $|w|$ **do**
5: $\quad a \leftarrow w'[i]$
6: $\quad u_{\mathrm{done}} \leftarrow w_\sigma w'[1..i-1]$
7: $\quad$ **for** $j = \max(1, i - \delta + 1)$ **to** i **do**
8: $\quad\quad I \leftarrow \{1, 2, \ldots, c\} \setminus \{\ell \mid u_{\mathrm{done}}[k] = a_\ell$ for some $(k, |w_\sigma| + i) \in H_j\}$
9: $\quad$ **end for**
10: $\quad w'[i] \leftarrow a_{\min I}$
11: **end for**
12: $\mathcal{H}' \leftarrow \mathcal{H}_{\mathrm{temp}} \setminus \big(\mathcal{H} \cup \big\{(a, c_i), (c_i, a) \bigm| 1 \leq i \leq c\big\}\big)$
13: **return** The self-attraction-free oritatami system $(\Sigma', w', \mathcal{H}', \delta, \alpha, \sigma')$

each foldable conformation, not both of a beads interacting with each other are modified as a_i. The other elongations may get less stable but it does not affect the behavior of the resulting system because a given system is deterministic so that the next bead is to be stabilized uniquely point-wise and interaction-wise no matter which of the most stable elongations is referred to.

The modified algorithm is implemented as Algorithm 2. A given oritatami system $\Xi = (\Sigma, w, \mathcal{H}, \delta, \alpha, \sigma)$ is deterministic so that, for each $0 \leq i \leq |w|$, there exists at most one foldable elongation of the seed by the transcript's prefix $w[1..i]$; let us denote it by C_i. After setting c rather linearly, Algorithm 2 runs as Algorithm 1 up to line 14. It then chooses arbitrarily for each k one representative elongation E_k of C_{k-1}, according to which the next bead $w[k]$ is stabilized in the given system. The i-th bead is transcribed at the $\max(1, i - \delta + 1)$-th step and it is involved in the stabilization of the previous at most $\delta - 1$ beads until it is stabilized at the i-th step. By the i-th execution of the outer **for**-loop in line 4, the first $i - 1$ beads of w have been already subscripted somehow; the remaining beads have not been given subscripts yet. The inner **for**-loop examines how the i-th bead is bound to preceding (already-subscripted) beads in the j-th representative for all $\max(1, i - \delta + 1) \leq j \leq i$ and chooses a proper subscript m out of the set I. All of the representatives $E_{\max(1,i-\delta+1)}, \ldots, E_j$ may have to be considered because the i-th bead may not be bound to the same bead in all of them. Note that the i-th bead can interact with 5 beads in the first representative but with at most 4 beads in the others. Therefore, $4\delta + 2$ subscripts suffice.

Lemma 3. *The following statements hold:*

- *For $1 \leq i \leq |w|+1$, the elongation C'_{i-1} of σ' by $w'[1..i-1]$ that is isomorphic to C_{i-1} is foldable by Ξ';*
- *For $1 \leq i \leq |w|$, the elongation E'_i of C'_{i-1} by $w'[i..i+\delta-1]$ that is isomorphic to the i-th representative chosen in line 3 is $\mathcal{H}'$-valid.*

Theorem 1. *Given a deterministic finite oritatami system Ξ of delay δ, we can solve Problem 2 for Ξ with the copying ratio $c = 4\delta + 2$.*

5 Lower bounds on copying ratio

Having established a linear upper bound on the copying ratio at least for deterministic oritatami systems, now we examine the lower bound. First, we propose a nondeterministic finite oritatami system Ξ_{nd} such that removing self-attraction from it requires a number of new bead types linearly proportional to the length of its transcript (Theorem 2). Based on it, we will design a deterministic finite oritatami system Ξ_{nd} of delay δ, which requires a copying ratio linear in δ to free itself from self-attraction by bead type modification (Theorem 3). This lower bound asymptotically matches the upper bound established in Theorem 1.

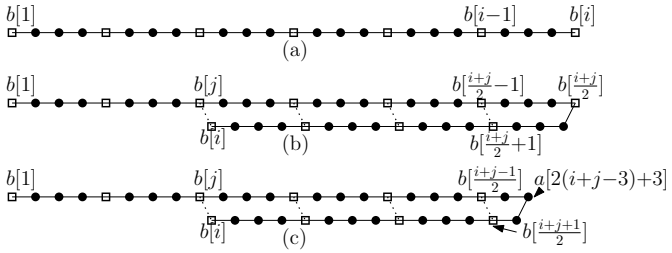

Fig. 4. Three examples of a part of C_j^i for $\delta = 1$. (a) Illustration of a part of C_i^i. (b) The case when $i + j$ is even. (c) The case when $i + j$ is odd.

The transcript of Ξ_{nd} is $w = (b\bullet^{2\delta+1})^t$ for some $t \geq 1$ and the seed of Ξ_{nd} is empty. Its ruleset is a singleton $\{(b, b)\}$, making b-beads self-attractive and $\bullet$-beads inert. By $b[i]$, we denote the i-th b-bead in w. See Fig. 4 for some of the conformations C_j^i foldable by this system, where δ is set to 1, in which $b[i]$ is bound to $b[j]$. Starting from the first bead, $b[1]$, this system can stretch its transcript straight rightward arbitrarily far and switch it back anywhere. The first $\delta + 2$ inert beads after $b[1]$ can be stabilized anyhow because it is not until they are stabilized that the next interactive bead, $b[2]$, is transcribed. Stretching them straight rightward is just one possibility. Being stabilized thus, they keep $b[1]$ out of the event horizon at the transcription of $b[2]$. Otherwise, $b[1]$ can lie in the horizon and pull $b[2]$ next to it and bind. Thus, for arbitrary $i \geq 1$ and $j > i$, this system can fold into a conformation C_j^i in which $b[i]$ is bound to $b[j]$. Consequently, in order to remove the self-attraction (b, b) by bead-type modification, these b-beads must be modified with pairwise-distinct indices, arising the need for the copying ratio $\lfloor |w|/(2\delta + 2) \rfloor$.

Theorem 2. *For a given delay δ and $n \in \mathbb{N}$, there exists a nondeterministic finite oritatami system Ξ_{nd} of delay δ whose transcript is of length n such that any solution Ξ'_{nd} to Problem 2 for Ξ_{nd} requires a copying ratio $c \geq \lfloor |w|/(2\delta+2) \rfloor$.*

The proof of Theorem 2 along with the fact that an oritatami system is only allowed a finite number of unique bead types yield the following result on the impossibility of removing self-attraction from infinite oritatami system.

Corollary 2. *There exists an infinite oritatami system Ξ such that there is no solution to Problem 2 for Ξ.*

Now we give a lower bound for the copying ratio for deterministic systems.

Theorem 3. *For a given delay δ, there exists a deterministic finite oritatami system Ξ_d of delay δ such that any solution Ξ'_d to Problem 2 for Ξ_d requires the copying ratio $c \geq \lfloor \delta/4 \rfloor$.*

We use delay $\delta = 4t$. The transcript of Ξ_d is similar to the transcript of Ξ_{nd} for Theorem 2, having periodic appearance of a self-attractive bead x. Note that conformations in Ξ_{nd} forces the self-attractive bead to have interactions with all other beads of the same type. We want the same phenomenon to happen in Ξ_d. Unlike Ξ_{nd}, a deterministic system stabilizes each bead at a unique point, so it is not possible to force all possible x beads to interact in the final conformation, since the conformation is planar. Instead, we use the most stable elongations during transcription. First, we design elongations to geometrically force all possible x beads to interact. Second, we prove that these elongations are indeed the most stable elongations to stabilize each bead. Third, we prove that all (x, x) interactions in these elongations are necessary for the system.

When we use bead type modification to remove a self-attractive rule (x, x), one (x, x) interaction in the original system may be removed in the resulting system if we modify interacting beads to the same bead type. If we want to maintain the interaction in the resulting system, we need to modify the beads into distinct types, say x_1 and x_2. On the other hand, interactions that are not self-attractive cannot be modified by bead type modification. The fact that we cannot remove unnecessary rules freely makes it challenging to design Ξ_d.

For every $2i - 2$nd bead, we assign an elongation C_i that mimics behaviors of conformations of Ξ_{nd}—proceed straight right, make a single right turn and proceed straight left. In detail, the transcript has two parts: The first part w_h of length δ with the repeated x every fourth bead, and the second part w_t of length $\delta - 5$ with distinct bead types for each bead. Given that x is repeated every fourth bead, it is straightforward to see that the set of these elongations is sufficient to force all possible x beads to interact. We need basic rules between the primary structure and the seed for the primary structure to proceed straight. In addition, we design the system so that the nascent beads for $2i - 2$nd bead in C_i has interaction strength $S_i = 3t + i - 1$. Fig. 5 is an illustration of different elongations C_1 to C_7 when $\delta = 16$ and $t = \delta/4 = 4$. The bead $x[i]$ denotes the ith bead x in the transcript, and the bead $y[i]$ denotes the $i$the bead. To meet the required strength of interactions, we use beads at the very last of each elongation to have special interactions with the seed. Note that a special interaction in one elongation never appears in any other elongations, since the coordinate of a bead

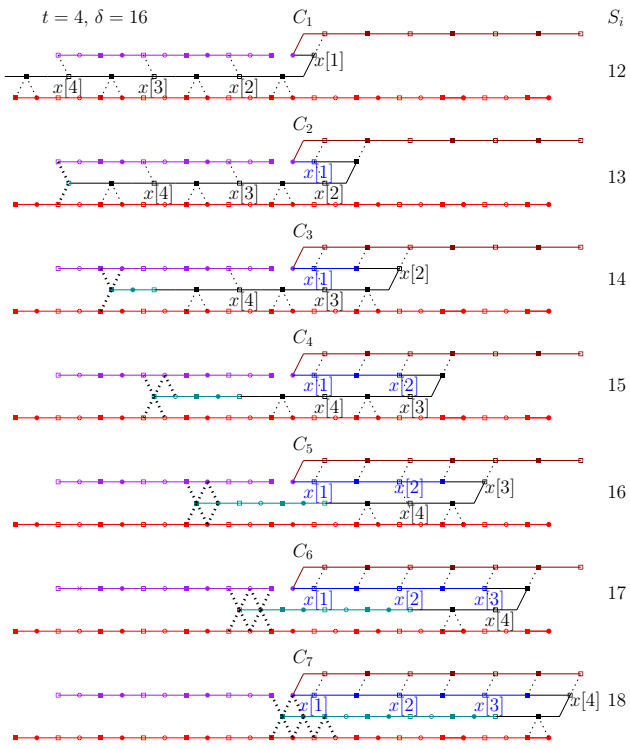

Fig. 5. A list of elongations when $t = 4$ and $\delta = 16$. The seed is given by three conformations $\sigma_b, \sigma_p, \sigma_r$, which are represented by brown, purple and red colors respectively. Two parts w_h and w_t of the transcript are represented by black and cyan colors respectively. The blue line represents beads already stabilized in the transcript. Thick dotted lines represent special interactions.

in w_t is distinct for all elongations. The last elongation C_{2t-1} is also the final conformation of the system.

First, we claim that each C_i is the most stable elongation while stabilizing $y[2i-2]$. If we compare C_i to C_j where $j < i$, S_j is less than S_i and beads after the primary structure of C_j cannot give additional special interactions. Thus, any elongation from C_j is less attractive than C_i. If we compare C_i to C_j where $j > i$, special interactions at the end of the primary structure of C_j makes S_j greater than S_i, but these end beads are too far to allow interactions to stabilize $y[2i-2]$. Thus, the part of C_j until the last bead of C_i is less attractive than C_i, which makes C_i most stable.

Second, we claim that all (x, x) interactions are necessary for Ξ_d to fold into the last elongation C_{2t-1}. Suppose we compare C_i to C_{i+1} while stabilizing $y[2i-2]$. Note that the coordinate of $y[2i-2]$ differs in C_i and C_{i+1}. Since C_{2t-1} is the final conformation, $y[2i-2]$ should be stabilized following the coordinate in C_{i+1}. Now, since S_i and S_{i+1} differs just by 1, if we remove some (x, x) in-

teractions in C_{i+1}, either $y[2i-2]$ is stabilized following the coordinate in C_i, or the system becomes nondeterministic. Thus, all (x, x) interactions are necessary. Since there are t different x beads, the lower bound of the copying ratio is $t = \delta/4$.

References

1. dailymotion.com/video/x3cdj35_oritatami-folding-turing-0-abc_school
2. Aggarwal, G., Goldwasser, M.H., Kao, M.Y., Schweller, R.T.: Complexities for generalized models of self-assembly. In: Proc. SODA 2004. pp. 880–889 (2004)
3. Barish, R.D., Schulman, R., Rothemund, P.W.K., Winfree, E.: An information-bearing seed for nucleating algorithmic self-assembly. PNAS 106(15), 6054–6059 (2009)
4. Cannon, S., Demaine, E.D., Demaine, M.L., Eisenstat, S., Patitz, M.J., Schweller, R.T., Summers, S.M., Winslow, A.: Two hands are better than one (up to constant factors): Self-assembly in the 2ham vs. atam. In: Proc. STACS 2013. LIPIcs, vol. 20, pp. 172–184 (2013)
5. Doty, D., Lutz, J.H., Patitz, M.J., Schweller, R.T., Summers, S.M., Woods, D.: The tile assembly model is intrinsically universal. In: Proc. FOCS 2012. pp. 302–310 (2012)
6. Evans, C.: Crystals that count! Physical principles and experimental investigations of DNA tile self-assembly. Ph.D. thesis, California Institute of Technology (June 2014)
7. Fochtman, T., Hendricks, J., Padilla, J.E., Patitz, M.J., Rogers, T.A.: Signal transmission across tile assemblies: 3d static tiles simulate active self-assembly by 2d signal-passing tiles. Nat. Comp. 14(2), 251–264 (2015)
8. Geary, C., Meunier, P.E., Schabanel, N., Seki, S.: Programming Biomolecules That Fold Greedily During Transcription. In: Proc. MFCS 2016. LIPIcs, vol. 58, pp. 43:1–43:14 (2016)
9. Geary, C., Rothemund, P.W.K., Andersen, E.S.: A single-stranded architecture for cotranscriptional folding of RNA nanostructures. Science 345(6198), 799–804 (2014)
10. Hendricks, J., Patitz, M.J., Rogers, T.A.: Universal simulation of directed systems in the abstract tile assembly model requires undirectedness. In: Proc. FOCS 2016. pp. 800–809 (2016)
11. Miyazono, A.G.E., Faraon, A., Rothemund, P.W.K.: Engineering and mapping nanocavity emission via precision placement of DNA origami. Nature 535(7612), 401–405 (2016)
12. Padilla, J.E., Patitz, M.J., Schweller, R.T., Seeman, N.C., Summers, S.M., Zhong, X.: Asynchronous signal passing for tile self-assembly: Fuel efficient computation and efficient assembly of shapes. Int. J. Found. Comput. S. 25(4), 459–488 (2014)
13. Rothemund, P.W.K., Papadakis, N., Winfree, E.: Algorithmic self-assembly of DNA Sierpinski triangles. PLoS Biol 2(12), e424 (12 2004)
14. Schulman, R., Yurke, B., Winfree, E.: Robust self-replication of combinatorial information via crystal growth and scission. PNAS 109(17), 6405–10 (2012)
15. Winfree, E.: Algorithmic Self-Assembly of DNA. Ph.D. thesis, California Institute of Technology (June 1998)
16. Woods, D.: Intrinsic universality and the computational power of self-assembly. In: Proc. MCU 2013. vol. 128, pp. 16–22 (2013)