Transcript Design Problem of Oritatami Systems

Yo-Sub Han[1] Hwee Kim[2*] Shinnosuke Seki[3]

[1]Yonsei University

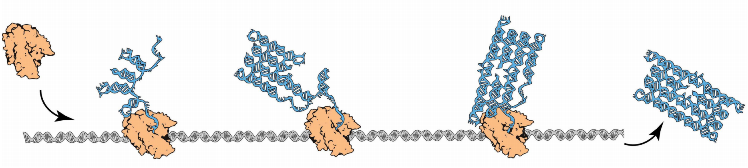
[2]University of South Florida

[3]University of Electro-Communications

The 24th International Conference on DNA Computing and Molecular Programming

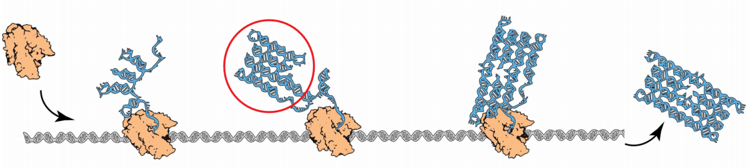
RNA Origami to Oritatami System (OS)

RNA Origami (Geary et al. (2014))



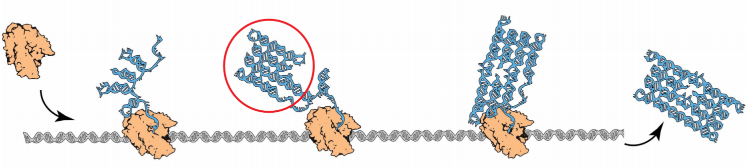
RNA Origami to Oritatami System (OS)

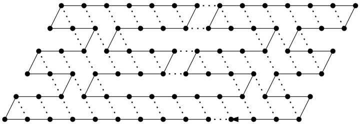
Cotranscriptional folding—folding occurs during transcription



RNA Origami to Oritatami System (OS)

Oritatami System is a mathematical model of computation by cotranscriptional folding (*oritatami* means folding in Japanese).

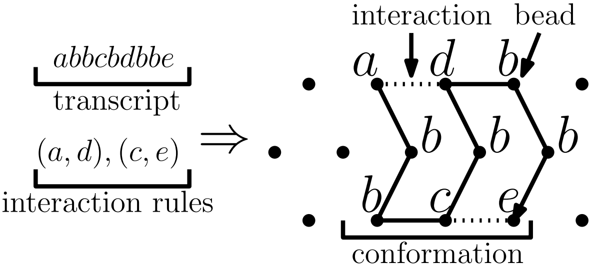




(Left) 3D Image of a tile generated by RNA origami (Right) Conformation that represents the tile

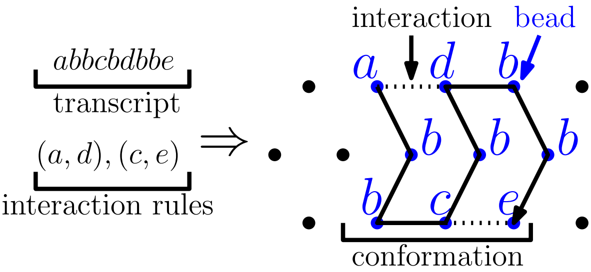
RNA Origami to Oritatami System (OS)

RNA Origami	Oritatami System
Nucleotides	Beads
Transcript	Sequence of beads connected by a line
h-bonds between nucleotides	Interactions
Cotranscriptional folding rate	Delay
Resulting secondary structure	Conformation



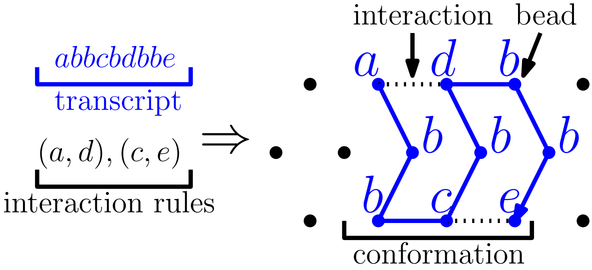
RNA Origami to Oritatami System (OS)

RNA Origami	Oritatami System
Nucleotides	Beads
Transcript	Sequence of beads connected by a line
h-bonds between nucleotides	Interactions
Cotranscriptional folding rate	Delay
Resulting secondary structure	Conformation



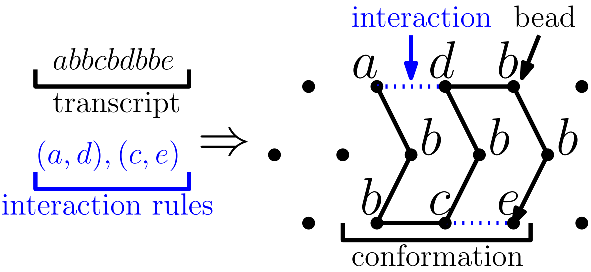
RNA Origami to Oritatami System (OS)

RNA Origami	Oritatami System
Nucleotides	Beads
Transcript	Sequence of beads connected by a line
h-bonds between nucleotides	Interactions
Cotranscriptional folding rate	Delay
Resulting secondary structure	Conformation



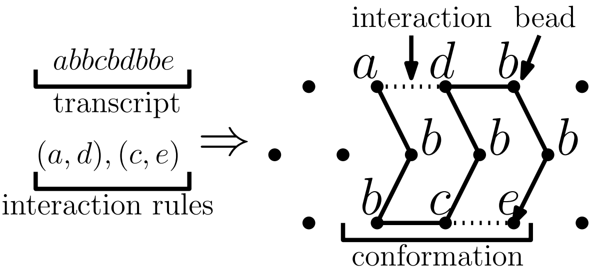
RNA Origami to Oritatami System (OS)

RNA Origami	Oritatami System
Nucleotides	Beads
Transcript	Sequence of beads connected by a line
h-bonds between nucleotides	Interactions
Cotranscriptional folding rate	Delay
Resulting secondary structure	Conformation



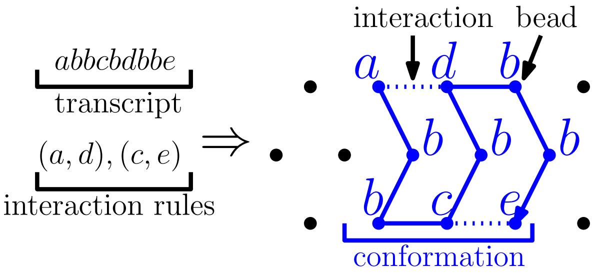
RNA Origami to Oritatami System (OS)

RNA Origami	Oritatami System
Nucleotides	Beads
Transcript	Sequence of beads connected by a line
h-bonds between nucleotides	Interactions
Cotranscriptional folding rate	Delay
Resulting secondary structure	Conformation



RNA Origami to Oritatami System (OS)

RNA Origami	Oritatami System
Nucleotides	Beads
Transcript	Sequence of beads connected by a line
h-bonds between nucleotides	Interactions
Cotranscriptional folding rate	Delay
Resulting secondary structure	Conformation



Dynamics of OS (Geary et al. (2015))

The seed C_σ is the initial conformation. We stabilize each bead of the transcript w as follows:

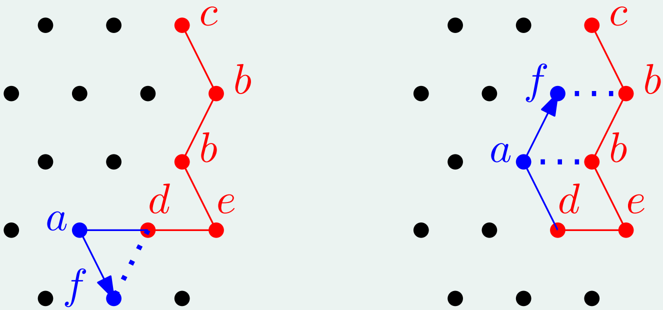
1. The bead look ahead up to next δ (delay) beads
2. Following rules in the ruleset $\mathcal{H}$, each pair of adjacent beads can form an interaction
3. The arity α denotes the maximum number of interactions that a bead can form
4. The first bead stabilizes as to maximize the number of interactions that the lookahead forms

Dynamics of OS (Geary et al. (2015))

Example

The delay of the system is 2. The seed is given as the red line. The transcript is *afe*. The ruleset is $\{(a, b), (b, f), (d, f), (d, e)\}$.

$$w = \underline{af}e$$

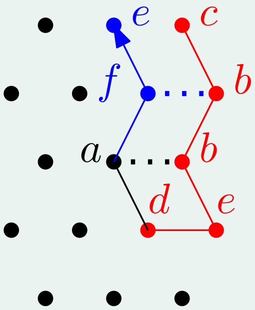


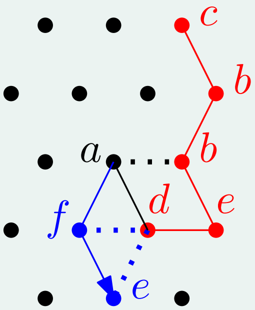
Dynamics of OS (Geary et al. (2015))

Example

The delay of the system is 2. The seed is given as the red line. The transcript is *afe*. The ruleset is $\{(a, b), (b, f), (d, f), (d, e)\}$.

$$w = a\underline{fe}$$



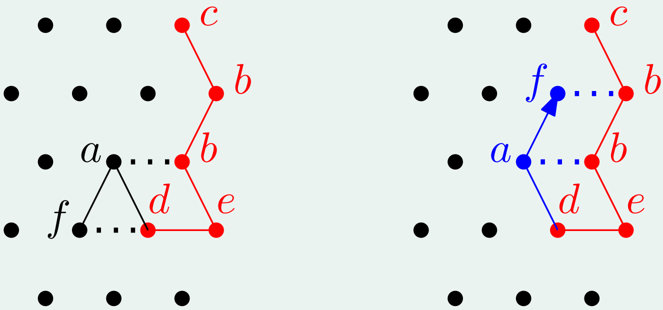


Dynamics of OS (Geary et al. (2015))

Example

The delay of the system is 2. The seed is given as the red line. The transcript is *afe*. The ruleset is $\{(a,b),(b,f),(d,f),(d,e)\}$.

$$w = a\underline{fe}$$

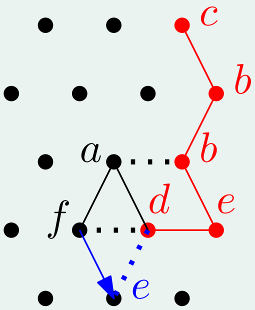


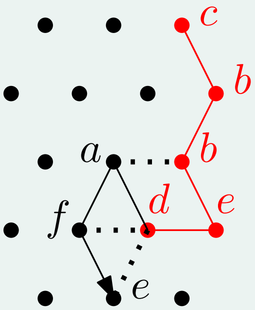
Dynamics of OS (Geary et al. (2015))

Example

The delay of the system is 2. The seed is given as the red line. The transcript is *afe*. The ruleset is $\{(a,b),(b,f),(d,f),(d,e)\}$.

$$w = af\underline{e}$$





Previous Works

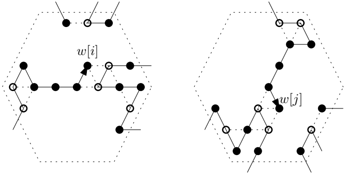
- Design
  - OS is Turing complete (Geary et al. (2015))
  - We can construct a binary counter (Geary et al. (2016))
  - We can construct a tautology checker (and a SAT solver) (Han et al. (2018))
- Hardness and Complexity
  - OS equivalence problem is coNP-hard (Han et al. (2018))
  - In general, it is NP-hard to retrieve a ruleset to fold the given conformation (Ota and Seki (2017))
  - Self-attraction removal by bead type copying (Han et al. (2017))
  - Ruleset optimization problem (Han and Kim (2017))
- Geometric Construction
  - Construction of the dragon curve (Masuda et al. (2018))
  - Hardness of constructing geometric shapes using certain delays (Rogers and Seki (2017))

Transcript Design Problem

- Transcript Design Problem (TDP)
  - Given a target structure, we want to find a sequence of beads that folds into the given structure
  - We assume that all other information are given—alphabet Σ, ruleset $\mathcal{H}$, delay δ, arity α
- Our contribution
  - Generalized parameterized algorithm for TDP
  - CTDP ($\mathcal{H}$ is complementary) is NP-hard
  - CTDP is NP-complete when $\delta = 3$ and $|\mathcal{H}| = 3$
  - CTDP can be solved in linear time when $\delta = 1$, $|\mathcal{H}| = 1$, $\alpha = 1$ or $\alpha \geq 4$
  - While solving CTDP of $\delta = 1$ and $\alpha = 1$, at most 26 other beads affect stabilization of a bead
  - There is no lower bound for the size of the ruleset where we can always find a transcript for the CTDP

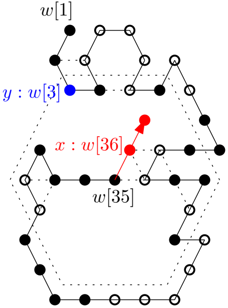
Event Horizon

- The stabilization of a bead $w[i]$ is only dependent to
  - the lookahead δ beads ($w[i]$ to $w[i+\delta-1]$)
  - already stabilized beads, which are geometrically at most $\delta + 1$ away from $w[i-1]$
- The event horizon denotes the hexagon, whose edges are $\delta + 1$ away from $w[i-1]$
- The event horizon context denotes all beads and interactions within the horizon
- The same event horizon context and lookahead yields the same stabilization



Dependence Distance

The dependence distance denotes the maximum range of indices for a bead x that effects stabilization of x.



$\delta = 2, d_{36} = 36 + 2 - 1 - 3 = 34$
($w[3]$ to $w[37]$)

Dependence distance: $\max(d_i)$

General Algorithm

Theorem

Given a TDP instance $(\Sigma, \mathcal{H}, \delta, \alpha, C_\sigma, P, H)$*, we can solve the TDP in* $O(|\Sigma|^t \times |P|)$*, where* t *is the dependence distance of the TDP instance.*

Proof.

Idea: For each bead x, at most t consecutive beads affect stabilization of x. □

Note that this general algorithm is fixed parameter linear.

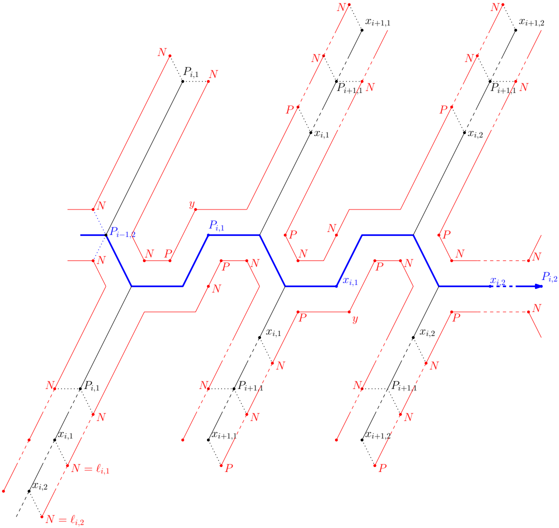
NP-hardness of CTDP

- CTDP: TDP when the ruleset is complementary (e.g. DNA)
- Inspired by Ota and Seki (2017), we reduce the problem from 1-IN-3-SAT problem

Definition

Given n Boolean variables $v_1, v_2, \ldots, v_n$ and m clauses $C_1, C_2, \ldots, C_m$ with exactly three positive variables in each clause, 1-IN-3-SAT problem finds an assignment that makes exactly one variable true in each clause.

NP-hardness of CTDP



"Chamber Gun" construction for a clause

NP-hardness of CTDP

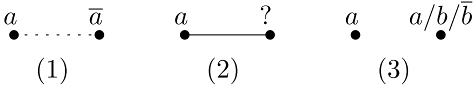
- Each unit of the transcript represents variable assignment for one clause
- Each unit of the seed represents variables in the cause
- Each unit transcript should represent the same value assignment not to fall into the wrong chamber
- In each unit, exactly one variable should be true not to fall into the wrong chambers
- Conclusion: We have the transcript that fold properly if and only if the original 1-IN-3-SAT problem has a solution

Theorem

For all $\alpha \geq 1$, the complementary transcript design problem (CTDP) at arity α is NP-hard. It remains NP-hard even if an input ruleset is restricted to be of size at most 2.

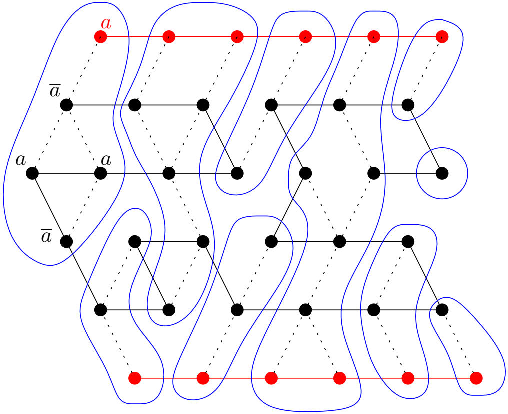
Graph-theoretic approach to CTDP

- When the path and the set of interactions are given, we can retrieve necessary dependence conditions between two adjacent beads (static dependence, or s-dependence):
  1. Connected with an interaction: Two beads should be complementary.
  2. Connected with a path: There is no necessary condition between two beads.
  3. No relationship: Two beads should not be complementary.



- Based on s-dependence, we may retrieve "dependent" sets of beads, where one bead in a set determines the rest
- We also have dynamic dependence, or d-dependence, that occurs during stabilization

Graph-theoretic approach to CTDP



An example of dependent sets. The seed is colored in red.

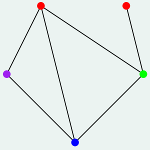
Graph-theoretic approach to CTDP

Theorem

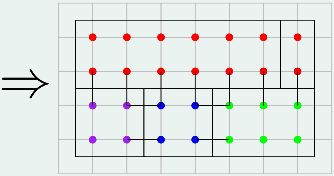
The CTDP is NP-complete when $\delta = 3$ and $|\mathcal{H}| = 3$.

Proof.

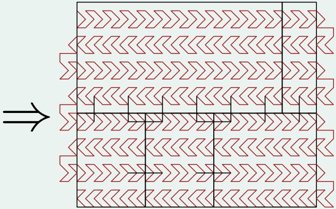
Idea: We may reduce the problem from the planar 3-coloring problem, using dependent sets to represent different vertices.
Planar graph→Square grid graph→dependent bead sets



Planar graph

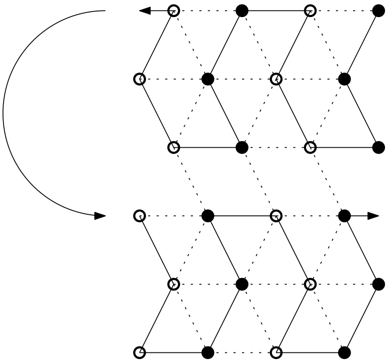


Square grid graph



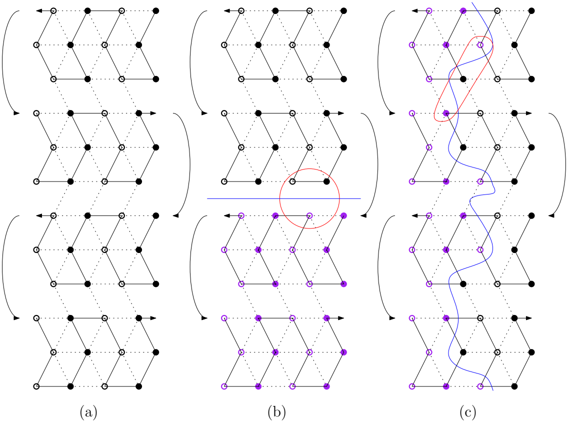
Dependent bead sets

Graph-theoretic approach to CTDP



A module that represent a vertex of the square grid

Graph-theoretic approach to CTDP



(a) The module that represents the lack of a vertical edge. (b) The module representing the presence of a vertical edge. (c) The module representing the presence of a horizontal edge.

Graph-theoretic approach to CTDP

- In this design, s-dependence is sufficient to satisfy d-dependence
- Conclusion: With three complementary rules, we have the transcript that fold properly if and only if the original graph is 3-colorable.

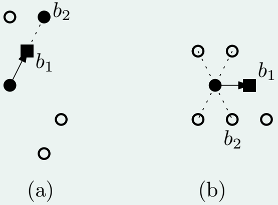
Linear Solvable Conditions of CTDP

Lemma

We can solve the CTDP in $O(|w|)$ time when $\delta = 1$, $|\mathcal{H}| = 1$ and $\alpha \geq 4$.

Proof.

Idea: There are two cases in stabilization of a bead. In all cases, we can uniquely determine the current bead type b_1 by one of the neighboring bead type b_2.



(a) Stabilization by interactions (b) Stabilization by geometry

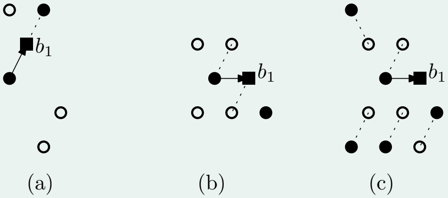
Linear Solvable Conditions of CTDP

Lemma

We can solve the CTDP in $O(|w|)$ time when $\delta = 1$, $|\mathcal{H}| = 1$ and $\alpha = 1$.

Proof.

When $\alpha = 1$, once a bead forms an interaction with another, these two beads become inactive and cannot form an interaction anymore. We call beads that are not binded as active beads. There are three cases.



(a) Stabilization by an interaction (b) Stabilization by geometry, having an active neighbor (c) Stabilization by geometry, not having an active neighbor

Linear Solvable Conditions of CTDP

Lemma

We can solve the CTDP in $O(|w|)$ time when $\delta = 1$, $|\mathcal{H}| = 1$ and $\alpha = 1$.

Proof.

Idea: For (a) and (b), we can determine the current bead type uniquely (same as when $\alpha \geq 4$). For (c), we may assign new bead type "variable" at the moment. Once that "variable" doesn't make conflict until the last bead, it is safe to assign arbitrary bead type to the variable. □

Conclusions

- Oritatami System (OS) is a mathematical model of computation by cotranscriptional folding
- We proposed the transcript design problem (TDP) to fold into the desired conformation.
- Our contribution
  - Generalized parameterized algorithm for TDP
  - CTDP ($\mathcal{H}$ is complementary) is NP-hard
  - CTDP is NP-complete when $\delta = 3$ and $|\mathcal{H}| = 3$
  - CTDP can be solved in linear time when $\delta = 1$, $|\mathcal{H}| = 1$, $\alpha = 1$ or $\alpha \geq 4$
- Future work: Exact boundary of NP-hardness for various conditions

Thank You!

References I

Geary, C., Meunier, P., Schabanel, N., and Seki, S. (2015). Efficient universal computation by greedy molecular folding. *CoRR*, abs/1508.00510.

Geary, C., Rothemund, P. W. K., and Andersen, E. S. (2014). A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science*, 345:799–804.

Geary, C. W., Meunier, P., Schabanel, N., and Seki, S. (2016). Programming biomolecules that fold greedily during transcription. In *Proceedings of the 41st International Symposium on Mathematical Foundations of Computer Science*, pages 43:1–43:14.

Han, Y. and Kim, H. (2017). Ruleset optimization on isomorphic oritatami systems. In *Proceedings of the 23rd International Conference on DNA Computing and Molecular Programming*, pages 33–45.

References II

Han, Y., Kim, H., Ota, M., and Seki, S. (2018). Nondeterministic seedless oritatami systems and hardness of testing their equivalence. *Natural Computing*, 17(1):67–79.

Han, Y., Kim, H., Rogers, T. A., and Seki, S. (2017). Self-attraction removal from oritatami systems. In *Proceedings of the 19th International Conference on Descriptional Complexity of Formal Systems*, pages 164–176.

Masuda, Y., Seki, S., and Ubukata, Y. (2018). Towards the algorithmic molecular self-assembly of fractals by cotranscriptional folding. In *Proceedings of the 23rd International Conference on Implementation and Application of Automata*, pages 261–273.

Ota, M. and Seki, S. (2017). Rule set design problems for oritatami system. *Theoretical Computer Science*, 671:16–35.

References III

Rogers, T. A. and Seki, S. (2017). Oritatami system; a survey and the impossibility of simple simulation at small delays. *Fundamenta Informaticae*, 154(1-4):359–372.