Oritatami, a model of cotranscriptional folding

Shinnosuke Seki

TUCS seminar
March 4th, 2015

Is folding hard?

Most structures in the nature, however complex and intricate, are obtained from a linear genetic code by folding, e.g.,

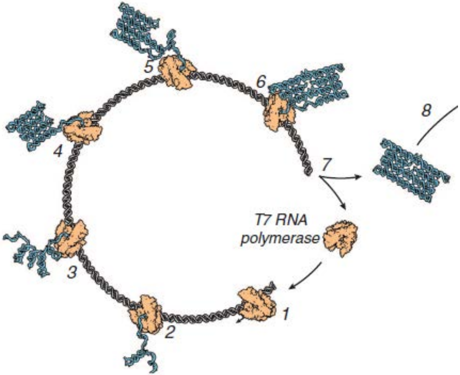
RNA sequences $\rightarrow$ large chain of amino-acids $\rightarrow$ proteins.

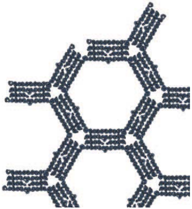
Predicting the most likely folding of an input sequence is known to be **NP**-hard.

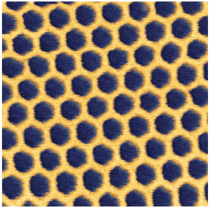
Is the nature stubbornly solving such hard problems?

Cotranscriptional folding

RNA origami

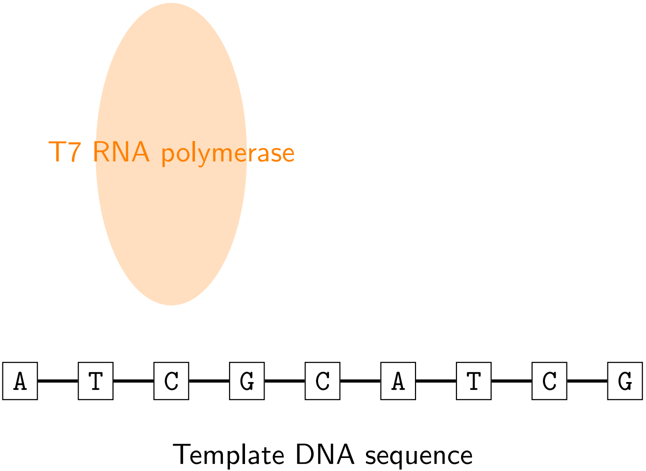




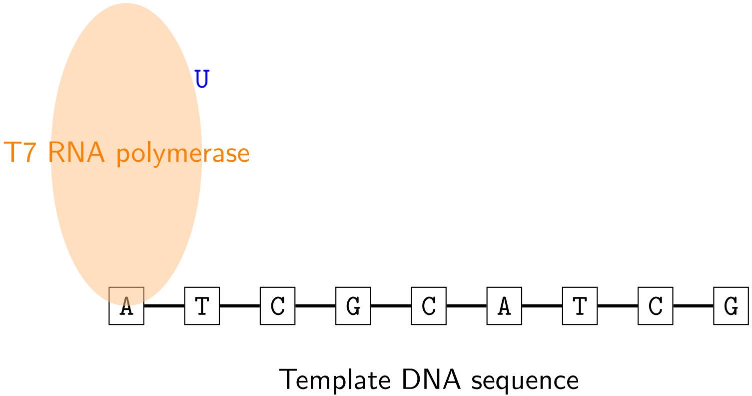


Design of ssRNA origami that self-assembles RNA tiles
[Geary, Rothemund, Andersen 2014]

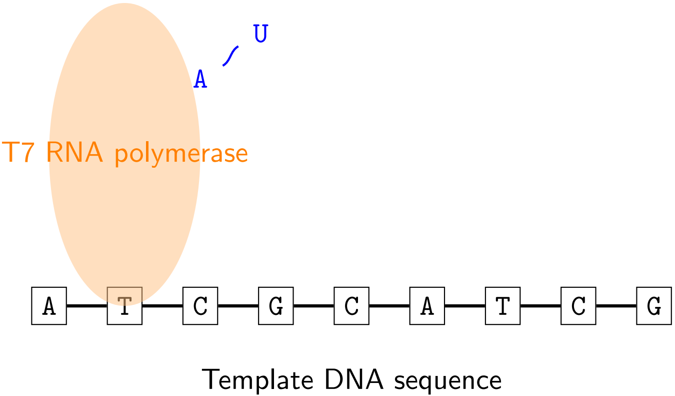
Cotranscriptional folding



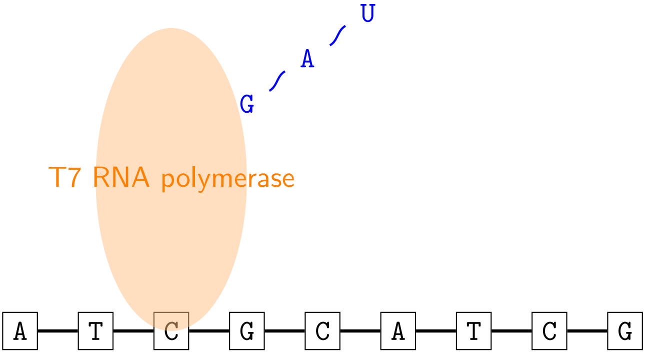
Cotranscriptional folding



Cotranscriptional folding

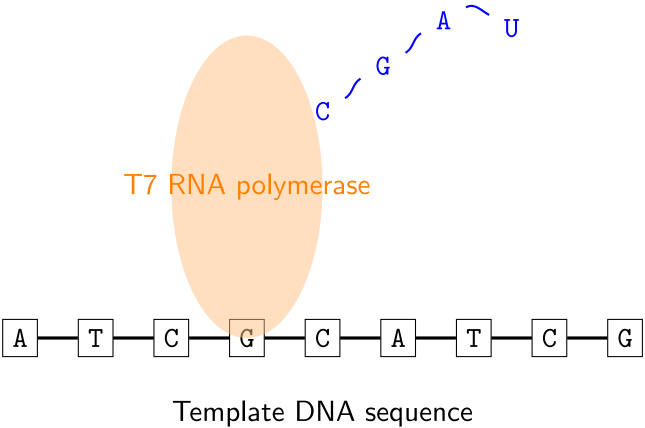


Cotranscriptional folding

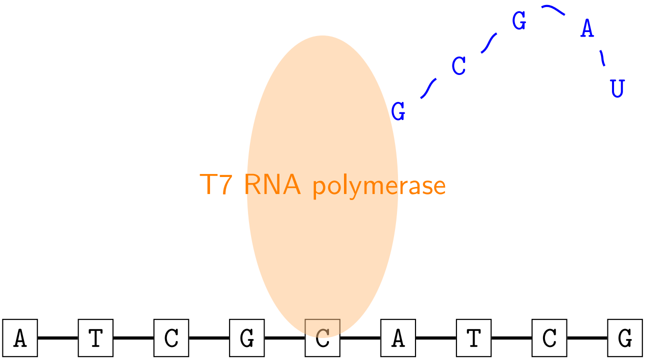


Template DNA sequence

Cotranscriptional folding

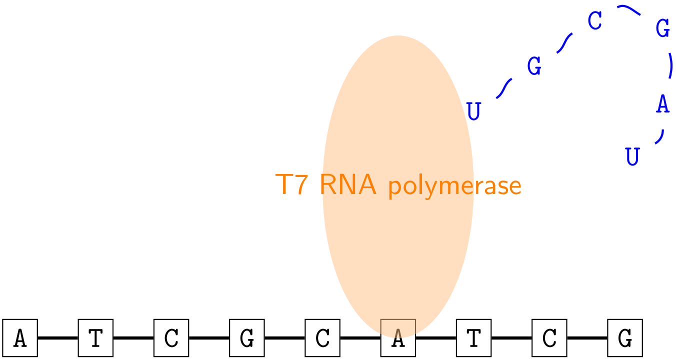


Cotranscriptional folding



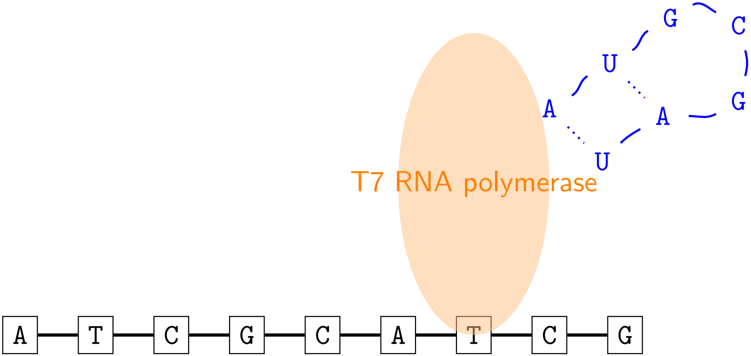
Template DNA sequence

Cotranscriptional folding



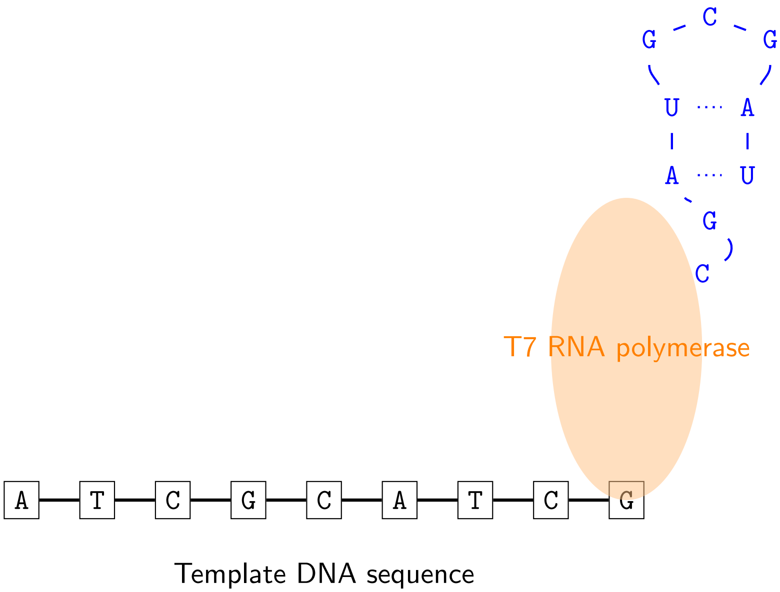
Template DNA sequence

Cotranscriptional folding

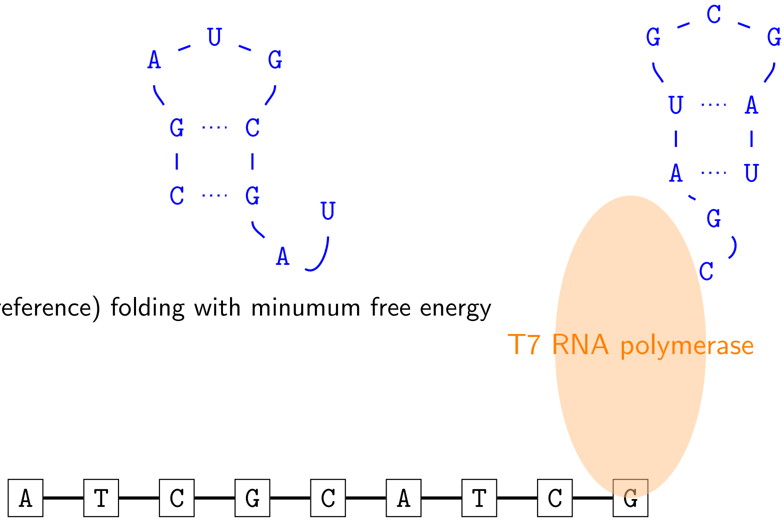


Template DNA sequence

Cotranscriptional folding



Cotranscriptional folding



Template DNA sequence

Oritatami system

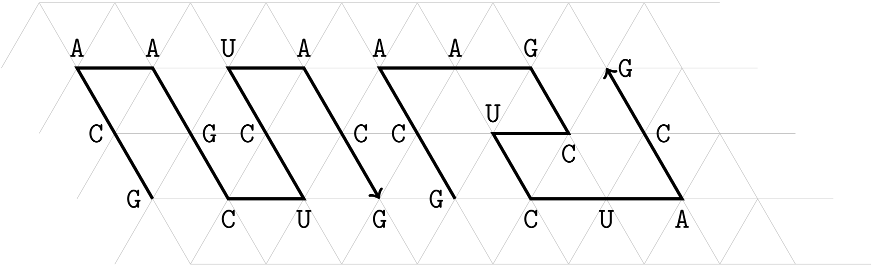
RNA primary structure is modeled as a sequence over Σ.

RNA secondary structure (conformation) is modeled as a pair of

p a non-self-crossing directed path on the triangular grid that is labeled by a primary structure

R A subset of $\{(i,j) \mid$ The i-th and j-th vertices on p with $j \geq i+2$ that are adjecent to each other on the grid$\}$.

Ex.) Two conformations a primary structure `GCAAGCUCUACG` may take.



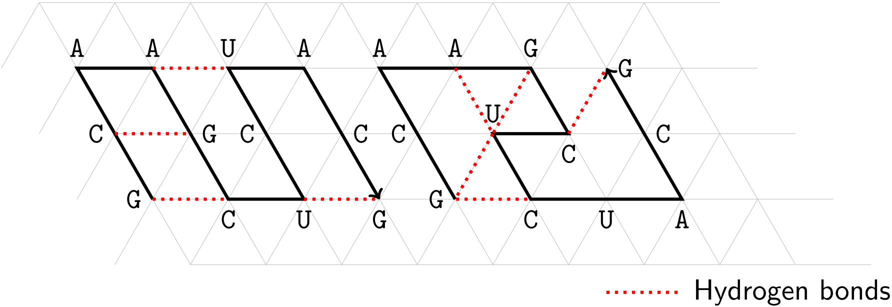
Oritatami system

RNA primary structure is modeled as a sequence over Σ.

RNA secondary structure (conformation) is modeled as a pair of

- p — a non-self-crossing directed path on the triangular grid that is labeled by a primary structure
- R — A subset of $\{(i,j) \mid$ The i-th and j-th vertices on p with $j \geq i+2$ that are adjecent to each other on the grid$\}$.

Ex.) Two conformations a primary structure `GCAAGCUCUACG` may take.



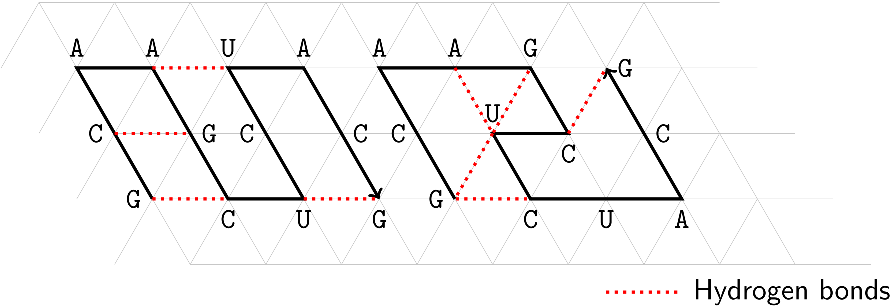
Oritatami system

Free energy

Principle of thermodynamics

Secondary structures with smaller free energy are more stable. Hence, among all possible secondary structures, a primary structure folds into the one(s) with smallest free energy.

Ex.) The right conformation has more hydrogen bonds, and hence, more stable, that is, has smaller free energy.



Oritatami system

Definition

An oritatami system is a 6-tuple $\Xi = (\Sigma, R, \alpha, w, \sigma, \delta_t)$, where

$R \subseteq \Sigma \times \Sigma$	aRb means that a bead of type $a \in \Sigma$ can form a hydrogen bond with a bead of type $b \in \Sigma$.
$\alpha \in \mathbb{N}$	Beads can form at most α bonds.
$w \in \Sigma^* \cup \Sigma^\omega$	a primary structure
σ	an initial conformation called seed.
$\delta_t \in \mathbb{N}$	delay time

If w is periodic, we say that the oritatami system is cyclic.

Oritatami system

Folding process and nondeterminism

Delay time $\delta_t = 2$

Primary structure $w = \texttt{GCUCUACG}$

Oritatami system

Folding process and nondeterminism

There are various ways to elongate the current conformation (blue) by transcribing the next 2 letters GC of the primary structure.

Delay time $\delta_t = 2$

Primary structure $w = \texttt{GCUCUACG}$

Oritatami system

Folding process and nondeterminism

There are various ways to elongate the current conformation (blue) by transcribing the next 2 letters GC of the primary structure.

Delay time $\delta_t = 2$

Primary structure $w = \texttt{GCUCUACG}$

Oritatami system

Folding process and nondeterminism

There are various ways to elongate the current conformation (blue) by transcribing the next 2 letters GC of the primary structure.

Delay time $\delta_t = 2$

Primary structure $w = \mathtt{GCUCUACG}$

Oritatami system

Folding process and nondeterminism

There are various ways to elongate the current conformation (blue) by transcribing the next 2 letters GC of the primary structure.

Delay time $\delta_t = 2$

Primary structure $w = \texttt{GCUCUACG}$

Oritatami system

Folding process and nondeterminism

There are various ways to elongate the current conformation (blue) by transcribing the next 2 letters GC of the primary structure.

Delay time $\delta_t = 2$

Primary structure $w = \texttt{GCUCUACG}$

Oritatami system

Folding process and nondeterminism

As such, it folds like this.

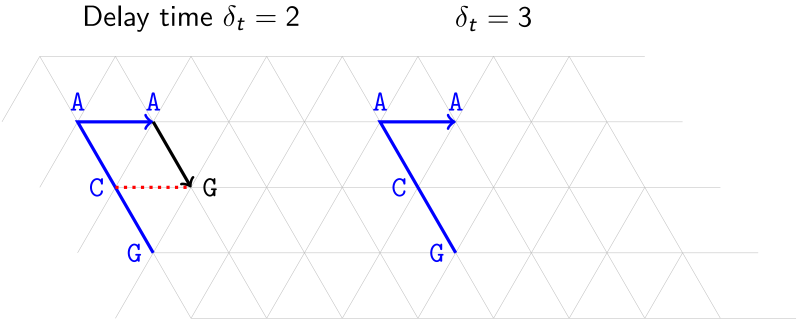
Delay time $\delta_t = 2$

Primary structure $w = \texttt{GCUCUACG}$

Oritatami system

Folding process and nondeterminism

When the delay time becomes 3, the determination of how to fold refers to the next 3 letters GCU instead.



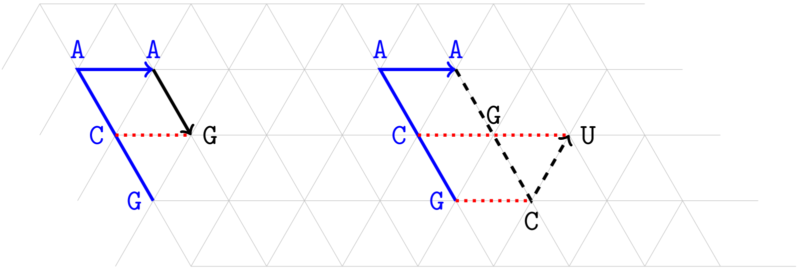
Primary structure $w =$ GCUCUACG

Oritatami system

Folding process and nondeterminism

When the delay time becomes 3, the determination of how to fold refers to the next 3 letters GCU instead.

Delay time $\delta_t = 2$ $\delta_t = 3$



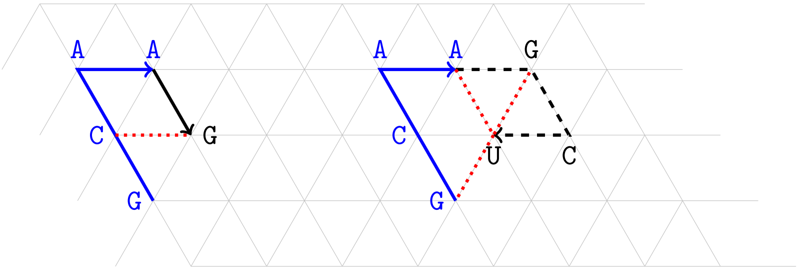
Primary structure $w = \texttt{GCUCUACG}$

Oritatami system

Folding process and nondeterminism

When the delay time becomes 3, the determination of how to fold refers to the next 3 letters GCU instead.

Delay time $\delta_t = 2$ $\delta_t = 3$

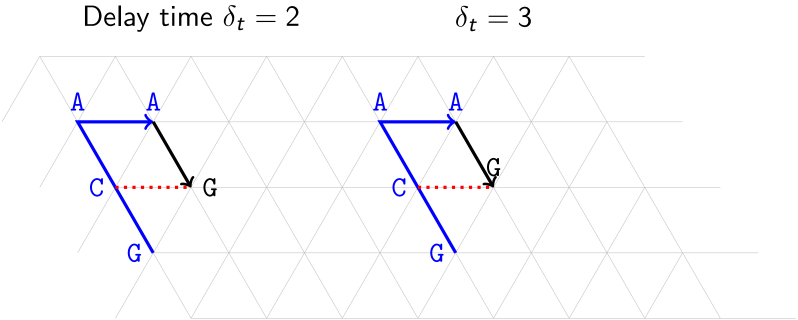


Primary structure $w = \texttt{GCUCUACG}$

Oritatami system

Folding process and nondeterminism

Hence, it folds nondeterministically in the two ways.



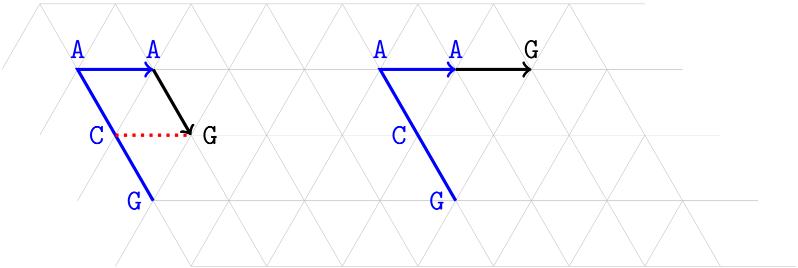
Primary structure $w = \texttt{GCUCUACG}$

Oritatami system

Folding process and nondeterminism

Hence, it folds nondeterministically in the two ways.

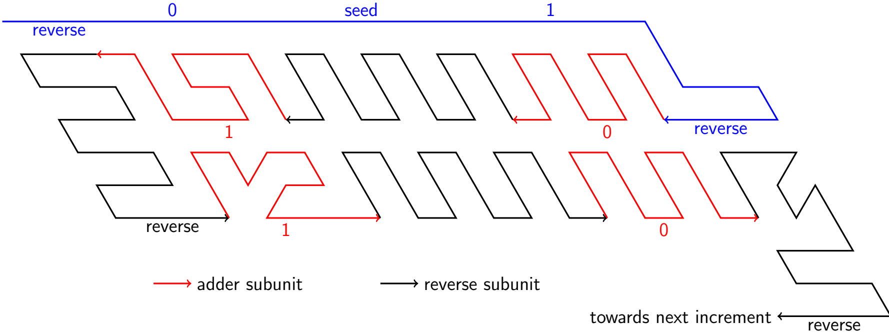
Delay time $\delta_t = 2$ $\delta_t = 3$



Primary structure $w = \mathtt{GCUCUACG}$

Cyclic oritatami system

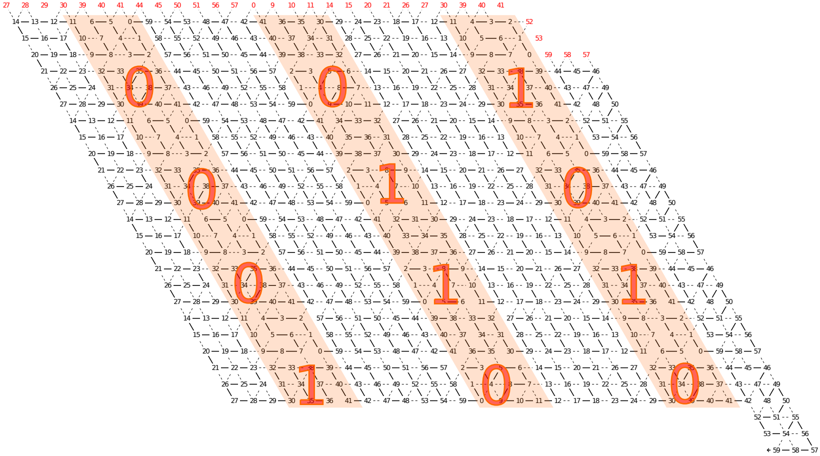
Binary counter



This primary structure is a repetition of adder subunit and reverse subunit. Repetitive primary structures can be transcribed easily from a cyclic DNA sequence.

Cyclic oritatami system

Binary counter (screenshot)



Cyclic oritatami system is Turing complete

Theorem

The class of cyclic oritatami system with delay time 3 is Turing complete.

Proof.

We design an oritatami system to emulate a cyclic tag system. $\square$

Cyclic oritatami system is Turing complete

Modified cyclic tag system

A cyclic tag system (cts) is a Turing-complete binary-string rewriting system which consists of an initial word $u \in \{0,1\}^*$ and a list of productions $v_1, v_2, \ldots, v_n \in \{0,1\}^*$ considered sequentially in this order, cycling back to v_1 after v_n being considered.
Its rewriting proceeds as follows:

1. Examine the leftmost letter of the current word.
2. If it is 1, then append the current production at the end of the word.
3. Delete the examined (leftmost) letter.

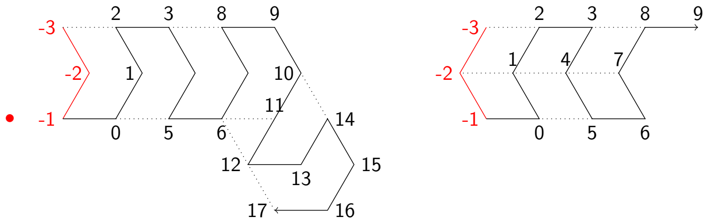
Skipping cts

It skips the next production after appending the current one in the case of the leftmost letter being 1.

Cyclic oritatami system is Turing complete

Tools

Glider



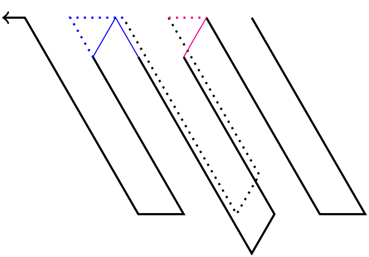
Gliders for oritatami systems with delay time 3.
(Left) A glider proceeds rightward according to the rule $(2, -3), (5, 0), (8, 3) \in R$. It can make a turn, due to some hardcoding in R or collision against environments.
(Right) Forward-swept wing (fsw) glider.

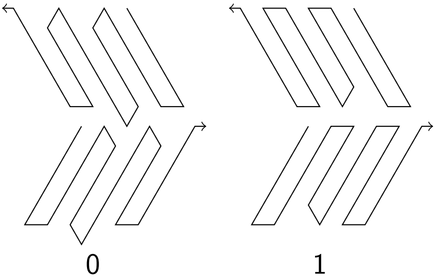
Cyclic oritatami system is Turing complete

Tools

Geometrical encoding

Our cts emulator encodes letters (0/1) of the current word geometrically by bumps (0) and dents (1).





Cyclic oritatami system is Turing complete

Primary structure

Assume that the skipping cts to be emulated has n productions $v_1, v_2, \ldots, v_n \in \{0, 1^*$.

The primary structure of our cts emulator consists of the subunits of the following two kinds:

Production
: It encodes one of the productions and plays the role of appending it to the current word.

Reversal-read-copy (r2c)
: As the name suggests, it plays three roles: reversal, reading a letter, and copying letters.

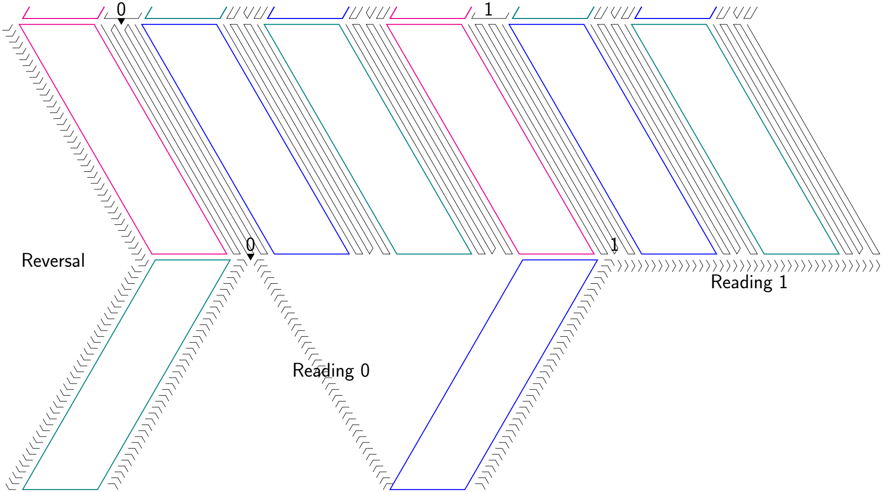
The primary structure is of the form:

$$\alpha_1 x_1 \alpha_2 x_2 \cdots \alpha_n x_n,$$

where $x_1, \ldots, x_n$ are r2c and α_i encodes the i-th production v_i.

Cyclic oritatami system is Turing complete

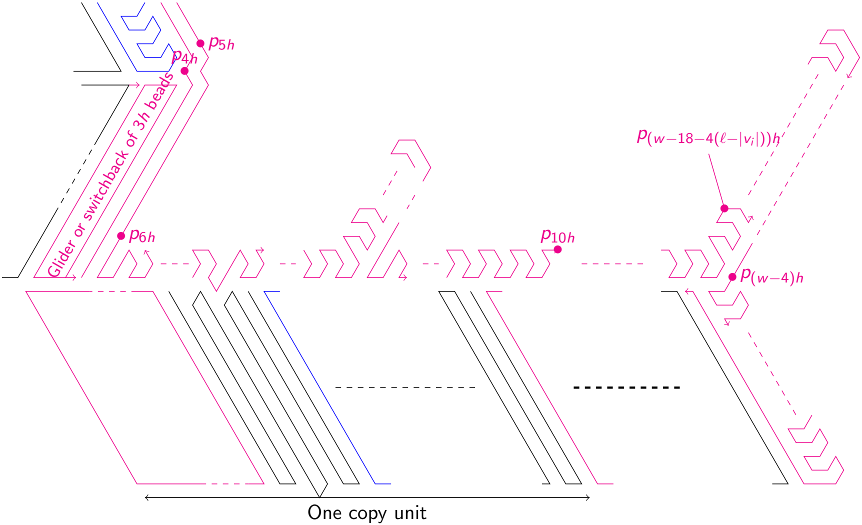
Reading the prefix of the form 0^*1



$n = 3$ production subunits are colored.

Cyclic oritatami system is Turing complete

Appending the encoded current production at the end



Cyclic oritatami system is Turing complete

Related open problems

Arity-1

Our cts emulator requires multiple arity. Is the class of oritatami systems with arity 1 still Turing complete?

Minimum alphabet size

What is the smallest alphabet Σ with which oritatami systems can be Turing complete?

Minimum period

What is the shortest period with which cyclic oritatami systems can be Turing complete?

Future projects on oritatami systems

There is a broad unexplored frontier of oritatami systems and cotranscriptional folding.

- Design of oritatami systems to self-assemble structures and mechanisms useful in nano-engineering.
- Algorithms to convert folding designs into oritatami systems and their computational complexity
- Development of oritatami simulator (Pierre-Etienne Meunier made the first one) and oritatami CAD
- Optimization of oritatami systems design
- Intrinsic universality
- Stochastic oritatami systems
- In-vitro implementation of oritatami systems in laboratories
- Oritatami GWAP (game with a purpose).

Oritatamists

Cody Geary
(CalTech)

Pierre-Etienne Meunier
(Aalto Univ.)

Nicolas Schabanel
(LIAFA, Univ. Paris Diderot)

UEC TOKYO

University of Electro-Communications

- National University specialized in computer and physical science, engineering, and technology.
- Notable alumni includes
  - Seinosuke Toda (Toda's theorem)
  - Sumio Iijima (inventor of carbon nanotubes)
  - Ken Kutaragi (father of PlayStation)
- Just 15 mins train ride from the central Tokyo.

University of
Electro-Communications

Hiro Ito

Satoshi Kobayashi

ACADEMY OF FINLAND

JAPAN SOCIETY FOR THE PROMOTION OF SCIENCE · FOUNDED 1932 ·

JSPS Fellowship to Japan

The Academy of Finland has funded researcher mobility with the Japan Society for the Promotion of Science (JSPS) since 1988.

- All scientific disciplines
- Duration 12-24 months
- It covers
  - travel costs between Finland and Japan
  - monthly grant of 362,000 JPY (about 2700 EUR)
  - settling-in allowance of 200,000 JPY

Thank you very much for your attention!

This talk is supported by the Academy of Finland, Postdoctoral Researcher Grant No. 13266670/T30606.

ACADEMY OF FINLAND

References I

C. Geary, P-E. Meunier, N. Schabanel, S. Seki.
Efficient universal computation by molecular co-transcriptional folding.
submitted.

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