

# Constrained RNA mutations as a SAT-problem

Grzegorz Łach (Faculty of Physics, University of Warsaw)

together with Gaja Klaudel, Katarzyna Tarasiewicz

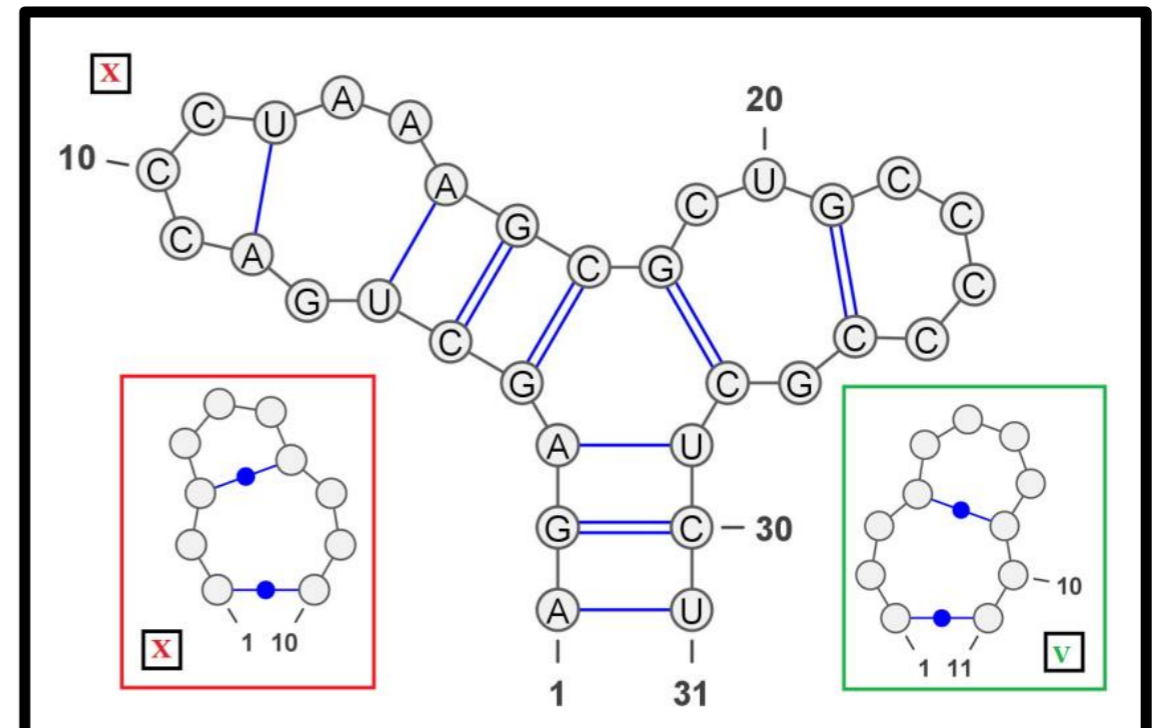
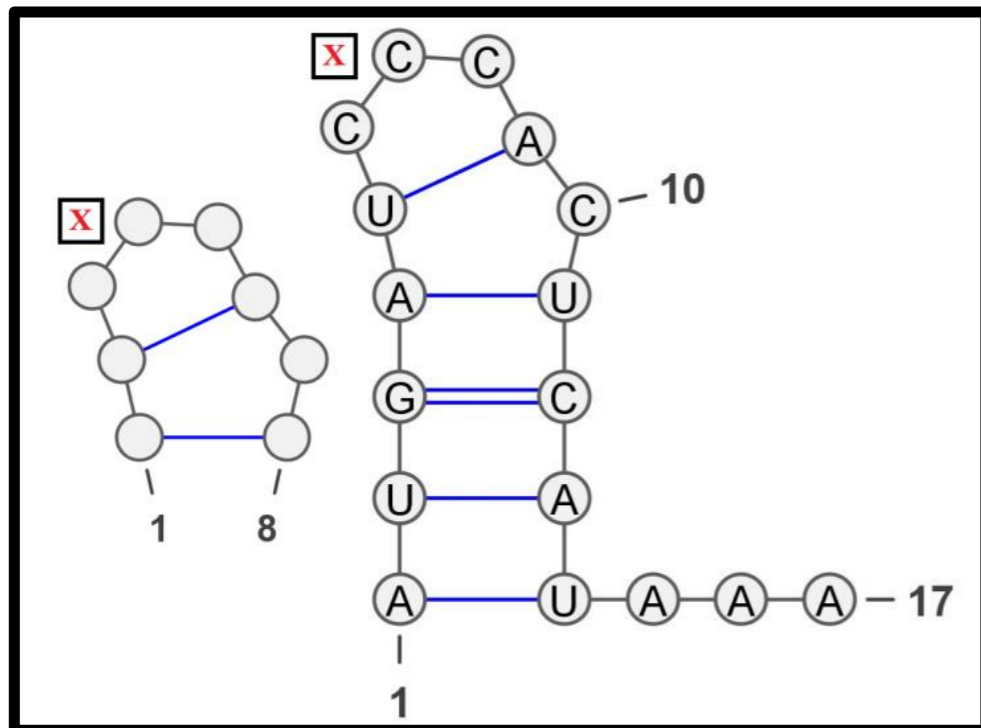
[Benasque 2024](#)



# ~~Forbidden motives in RNA design~~

Grzegorz Łach (Faculty of Physics, University of Warsaw)

together with Katarzyna Tarasiewicz



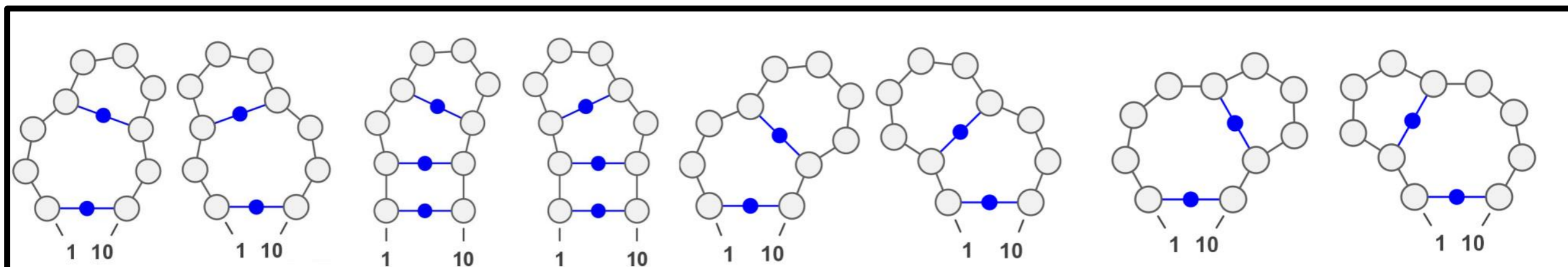
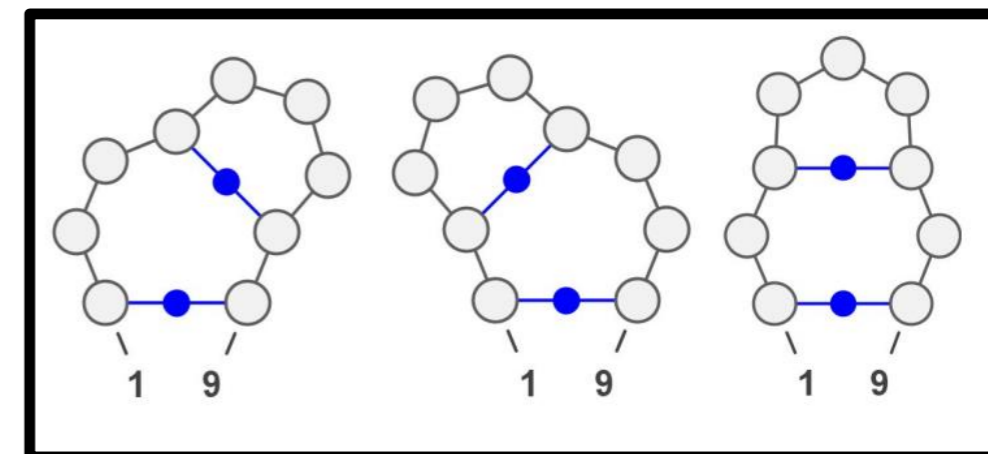
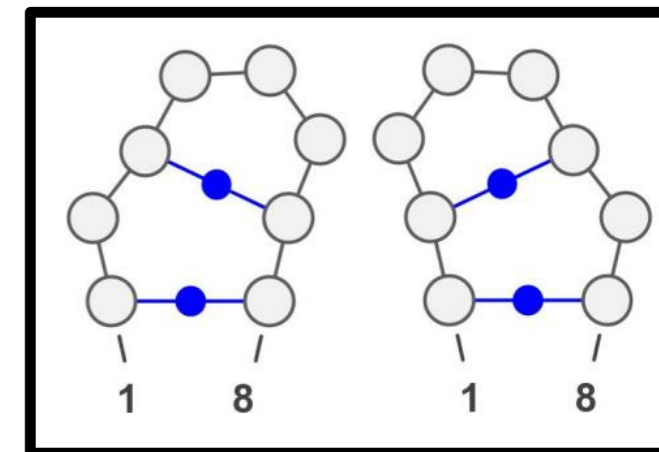
# ~~Forbidden motifs in RNA design~~

Grzegorz Łach (Faculty of Physics, University of Warsaw)

together with Katarzyna Tarasiewicz

hairpin loop

n	prohibited motifs	Turner 1998	Turner 2004	Langdon 2018
8	((...))	X	X	X
	(. (...))	X	X	X
9	((...))	X	X	X
	(. (...))	X	X	X
	(. (...))	-	-	X
10	((...))	X	X	X
	((...))	X	X	X
	(. (...))	X	X	X
	(. (...))	X	X	X
	(. (...))	X	X	X
	(. (...))	X	X	X
	((...))	X	X	X
	((...))	X	X	X



N= 0

M=4

((.....).....) : 9216

((.....).....) : 9216

((.....).....) : 9216

N= 1

M=3

(.(.....)....) : 4252

(.(.....)....) : 3548

(.(.....)....) : 3676

((.....)...) : 2408

((.....)...) : 2669

((.....)...) : 2600

((.....)).. : 2223

((.....)).. : 2560

((.....)).. : 2517

((.....).). : 333

((.....).). : 439

((.....).). : 423

N= 2

M=2

(..(.....)..) : 3861

(..(.....)..) : 3690

(..(.....)..) : 4131

((.(.....).)) : 1849

((.(.....).)) : 2088

((.(.....).)) : 1911

(.((.....)).) : 1849

(.((.....)).) : 2088

(.((.....)).) : 1911

((((.....)))) : 1296

((((.....)))) : 1296

((((.....)))) : 1221

..... : 221

(.(.....).) : 46

(.(.....).) : 16

(.((.....).)) : 72

((.(.....).) : 8

((.(.....).) : 26

((.(.....).) : 68c

N= 3

M=1

(.(.....)....) : 4252

(.(.....)....) : 3548

(.(.....)....) : 3676

((.....)...) : 2408

((.....)...) : 2669

((.....)...) : 2600

((.....)).. : 2223

((.....)).. : 2560

((.....)).. : 2517

((.....).). : 333

((.....).). : 439

((.....).). : 423

N= 4

M=0

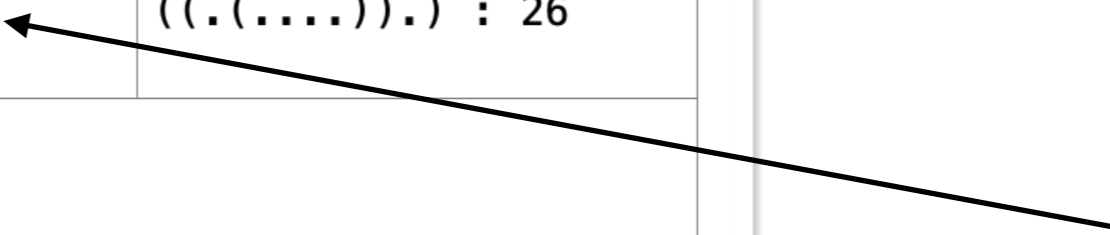
((.....).....) : 9216

((.....).....) : 9216

((.....).....) : 9216

barely  
designable

...possibly  
hard to design



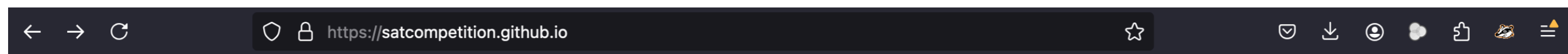
# SAT problems

example:

$(a \text{ or } b \text{ or } (\text{not } c)) \text{ and } (b \text{ or } c \text{ or } (\text{not } d)) \text{ and } (a \text{ or } c \text{ or } d) \text{ and}$

$((\text{not } a) \text{ or } b \text{ or } d) \text{ and } ((\text{not } a) \text{ or } (\text{not } b) \text{ or } c) \text{ and } ((\text{not } b) \text{ or } (\text{not } c) \text{ or } d) \text{ and}$

$((\text{not } a) \text{ or } (\text{not } c) \text{ or } (\text{not } d)) \text{ and } (a \text{ or } (\text{not } b) \text{ or } (\text{not } d))$



## The International SAT Competition Web Page

### Current Competition

#### SAT 2024 Competition

Organizers [Marijn Heule](#), [Markus Iser](#), [Matti Järvisalo](#), [Martin Suda](#)

### Past Competitions, Races and Evaluations

#### SAT 2023 Competition

Organizers [Marijn Heule](#), [Matti Järvisalo](#), [Martin Suda](#), [Markus Iser](#), [Tomáš Balyo](#)

#### SAT 2022 Competition

Organizers [Marijn Heule](#), [Matti Järvisalo](#), [Martin Suda](#), [Markus Iser](#), [Tomáš Balyo](#)

#### SAT 2021 Competition

Organizers [Marijn Heule](#), [Matti Järvisalo](#), [Martin Suda](#), [Markus Iser](#), [Tomáš Balyo](#), [Nils Froleyks](#)

#### SAT 2020 Competition

Organizers [Marijn Heule](#), [Matti Järvisalo](#), [Martin Suda](#), [Markus Iser](#), [Tomáš Balyo](#), [Nils Froleyks](#)

# RNA sequence/structure as SAT

U U C G G G U A C C G A A C  
 [ 0 , 1 , 0 , 1 , 1 , 1 , 1 , 0 , 1 , 0 , 1 , 0 , 0 , 1 , 0 , 0 , 1 , 1 , 1 , 1 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 1 , 1 ]

- 'A' : ( 0 , 0 )
- 'U' : ( 0 , 1 )
- 'G' : ( 1 , 0 )
- 'C' : ( 1 , 1 )

## Encoding of base-pairings

( . . . . . )

[ 0 , 1 , 0 , 1 , 1 , 1 , x , y , 1 , 0 , 1 , 0 , 0 , 1 , 0 , 0 , 1 , 1 , x , y , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 1 , 1 ]  
i j

$$(\neg x_i \vee x_j \vee \neg y_i) \wedge (x_i \vee \neg x_j \vee \neg y_j) \wedge (\neg y_i \vee \neg y_j) \wedge (y_i \vee y_j)$$

# Encoding (local) sequence constrains

U U C **G** G G U A C C G A A C  
 [0,1,0,1,1,1,1,0,1,0,1,0,0,1,0,0,1,1,1,1,1,0,0,0,0,0,1,1]

'A' : (0, 0)  
 'U' : (0, 1)  
 'G' : (1, 0)  
 'C' : (1, 1)

IUPAC code	residues	logic clause
'A'	A	$\neg x_i \wedge \neg y_i$
'U'	U	$\neg x_i \wedge y_i$
'G'	G	$x_i \wedge \neg y_i$
'C'	C	$x_i \wedge y_i$
'W'	A or U	$\neg x_i$
'S'	C or G	$x_i$
'Y'	C or U	$y_i$
'R'	A or G	$\neg y_i$
'M'	A or C	$(\neg x_i \vee y_i) \wedge (x_i \vee \neg y_i)$
'K'	G or U	$(\neg x_i \vee \neg y_i) \wedge (x_i \vee y_i)$
'B'	C or G or U	$x_i \vee y_i$
'D'	A or G or U	$\neg x_i \vee \neg y_i$
'H'	A or C or U	$\neg x_i \vee y_i$
'V'	A or C or G	$x_i \vee \neg y_i$

# Global sequence constrains (eg. forbidden patterns)

forbidden pattern at given position

6 7 8 9

[0,1,0,1,1,1,1,0,1,0,1,0,0,1,0,0,1,1,1,1,1,0,0,0,0,0,1,1]

G U A C

$$\neg x_6 \vee y_6 \vee x_7 \vee \neg y_7 \vee x_8 \vee y_8 \vee \neg x_9 \vee \neg y_9$$

'A' : (0, 0)

'U' : (0, 1)

'G' : (1, 0)

'C' : (1, 1)








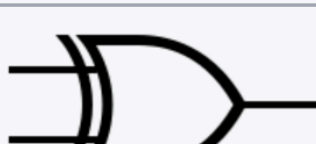
# Global sequence constraints (eg. forbidden patterns)

forbidden pattern at any position

1	2	3	4	5	6	7	8	9																
[0,1,0,1,1,1,1,0,1,0,1,0,0,1,0,0,1,1,1,1,1,0,0,0,0,0,1,1]																								
G	U	A	C																					
	G	U	A	C																				
		G	U	A	C																			
			G	U	A	C																		
				G	U	A	C																	
					G	U	A	C																
						G	U	A	C															
							G	U	A	C														
								G	U	A	C													
									G	U	A	C												
										G	U	A	C											
											G	U	A	C										
												G	U	A	C									
													G	U	A	C								

*not efficient for long sequences*

# Tseitin transform

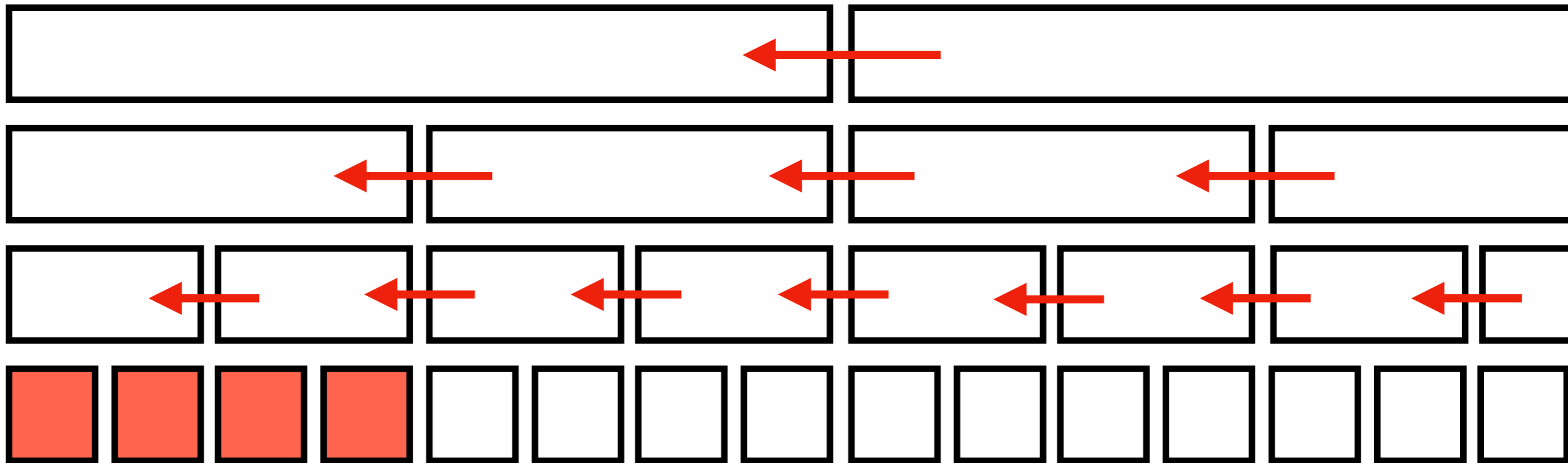
Type	Operation	CNF sub-expression
 <b>AND</b>	$C = A \cdot B$	$(\bar{A} \vee \bar{B} \vee C) \wedge (A \vee \bar{C}) \wedge (B \vee \bar{C})$
 <b>NAND</b>	$C = \overline{A \cdot B}$	$(\bar{A} \vee \bar{B} \vee \bar{C}) \wedge (A \vee C) \wedge (B \vee C)$
 <b>OR</b>	$C = A + B$	$(A \vee B \vee \bar{C}) \wedge (\bar{A} \vee C) \wedge (\bar{B} \vee C)$
 <b>NOR</b>	$C = \overline{A + B}$	$(A \vee B \vee C) \wedge (\bar{A} \vee \bar{C}) \wedge (\bar{B} \vee \bar{C})$
 <b>NOT</b>	$C = \bar{A}$	$(\bar{A} \vee \bar{C}) \wedge (A \vee C)$
 <b>XOR</b>	$C = A \oplus B$	$(\bar{A} \vee \bar{B} \vee \bar{C}) \wedge (A \vee B \vee \bar{C}) \wedge (A \vee \bar{B} \vee C) \wedge (\bar{A} \vee B \vee C)$

# Global sequence constrains (eg. forbidden patterns)

forbidden pattern at any position

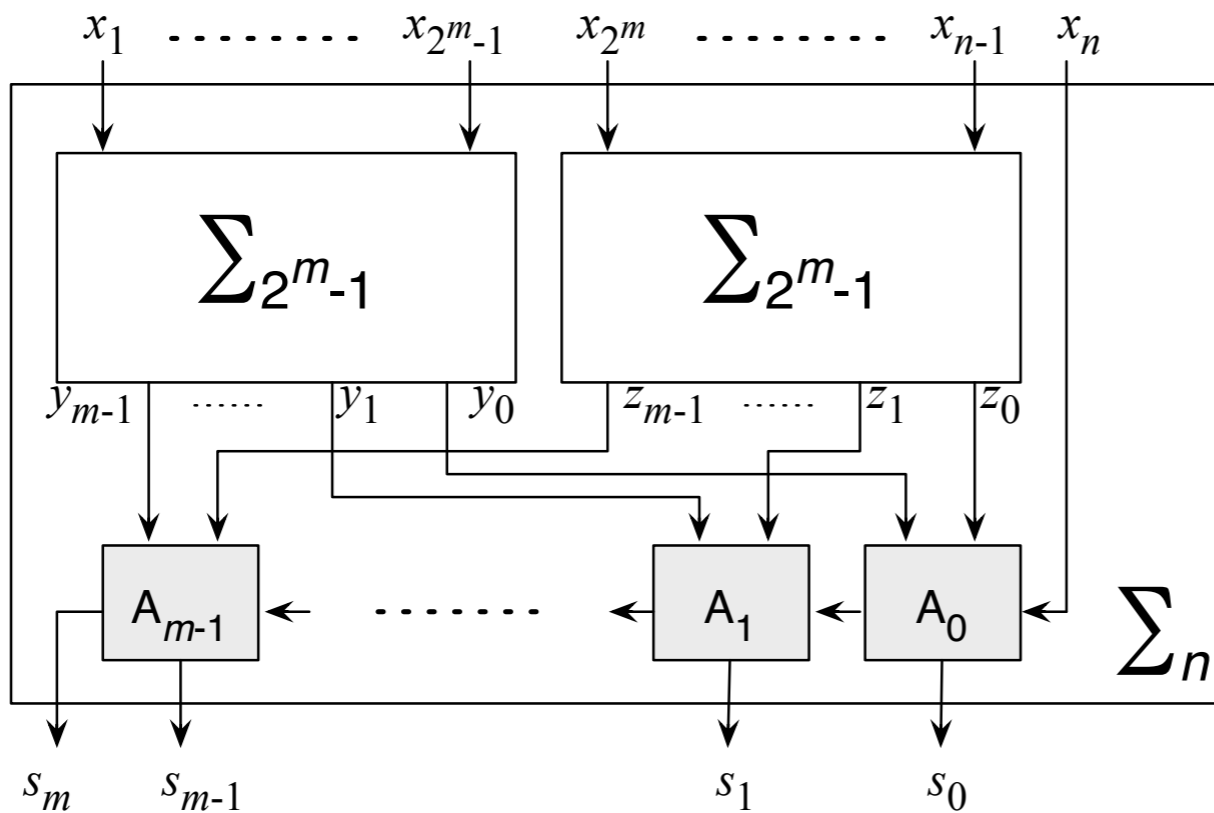
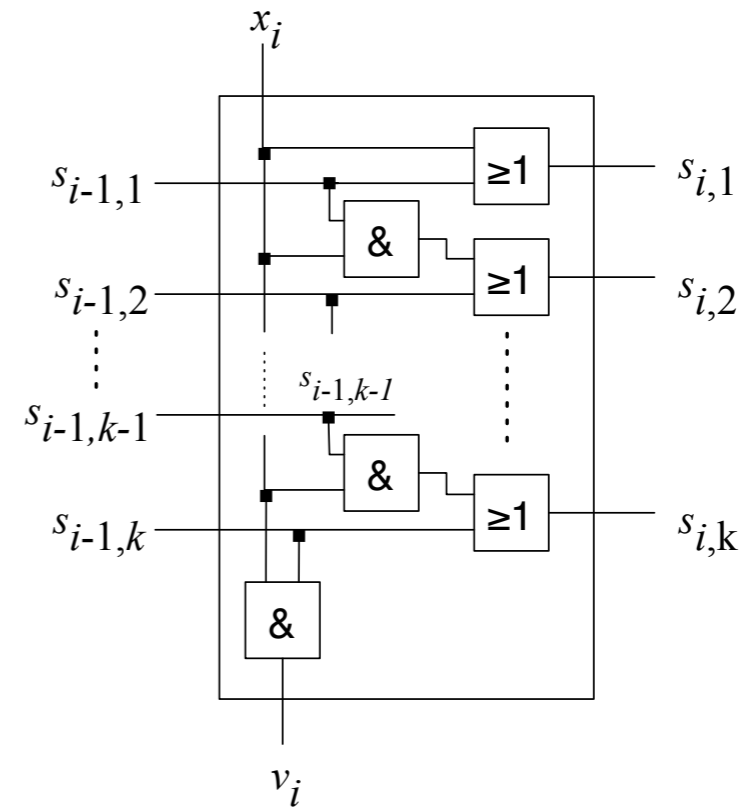
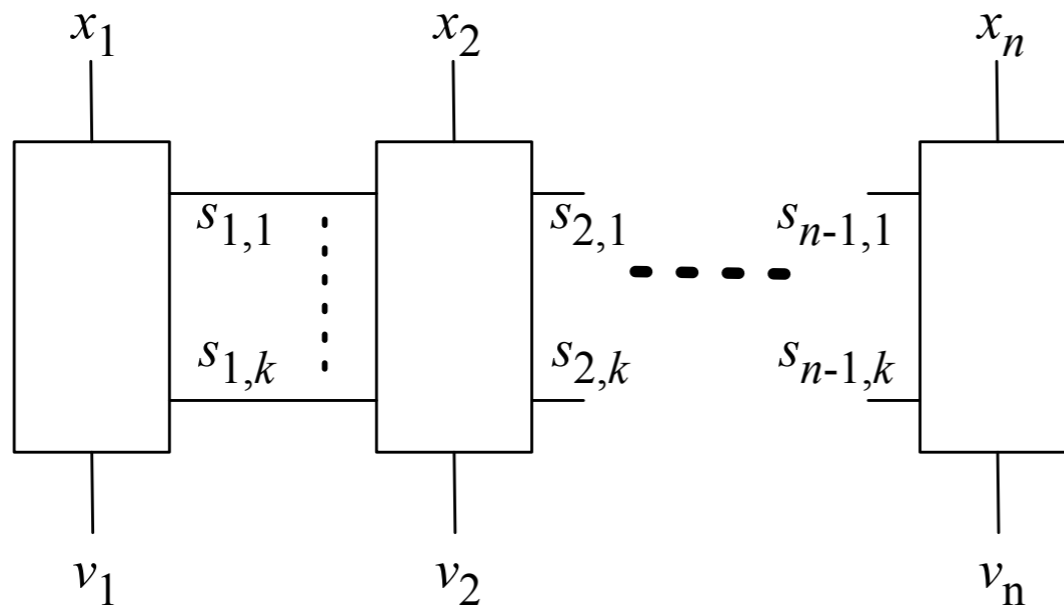
1 2 3 4 5 6 7 8 9

[0,1,0,1,1,1,1,0,1,0,1,0,0,1,0,0,1,1,1,1,1,0,0,0,0,0,1,1]

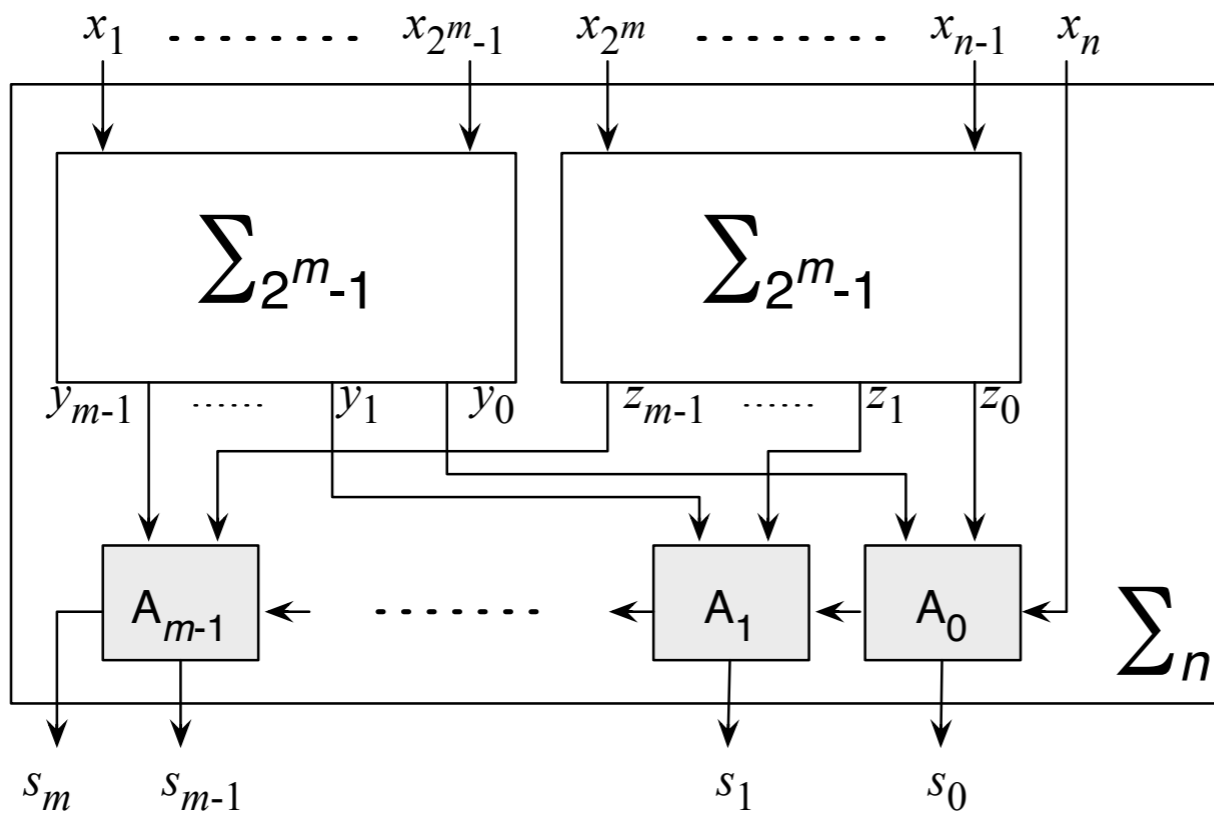
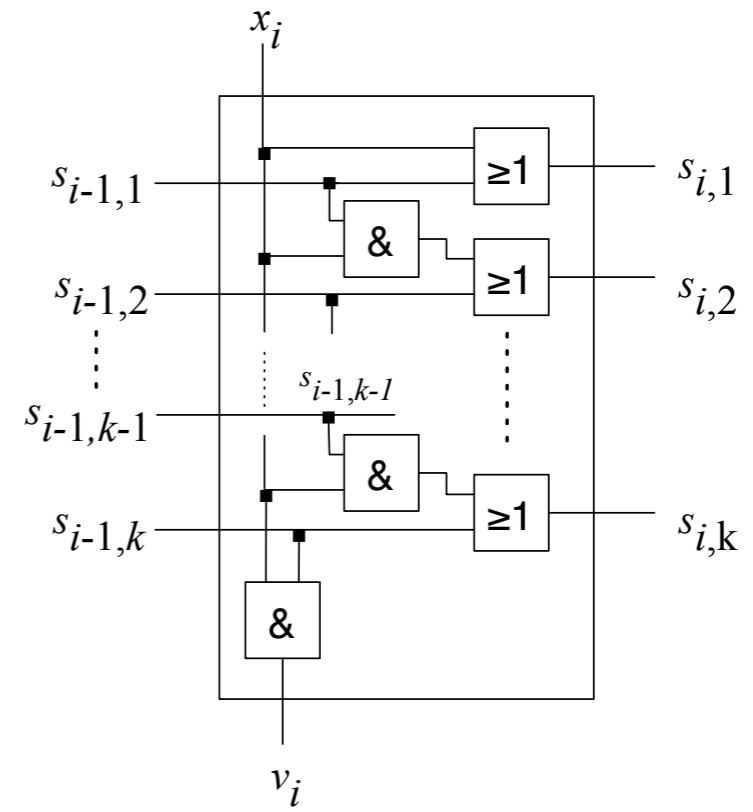
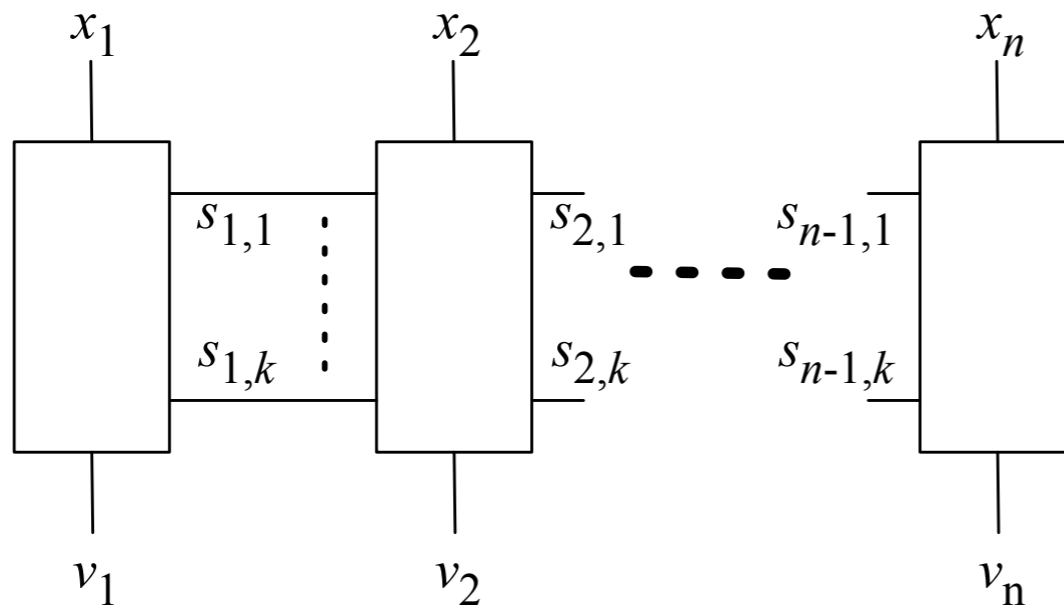


$$\neg x_1 \vee y_1 \vee x_2 \vee \neg y_2 \vee x_3 \vee y_3 \vee \neg x_4 \vee \neg y_4$$

# Cardinality constraints (eg. on GC-content)



# Cardinality constraints (eg. on GC-content)



**Timing**