State complexity of the multiples of the Thue-Morse set

Célia Cisternino
Joint work with Émilie Charlier and Adeline Massuir

January 24, 2019

Basics

- Alphabet A, letter $a \in A$, word w
- ε, $|w|$, $|w|_a$
- Language

Moreover,

- Automaton (DFA) $\mathcal{A}$
- The language accepted from a state q is denoted by $L(q)$.
- Regular language

- Reduced, accessible, coaccessible

Definition

A DFA is ***minimal*** iff it is ***reduced*** and ***accessible***.

- Trim minimal

Definition

The ***state complexity*** of a regular language is equal to the number of states of its minimal automaton.

Definition

A DFA has ***disjoint states*** if, for distinct states p and q, we have $L(p) \cap L(q) = \emptyset$.

Remark

Any coaccessible DFA having disjoint states is reduced.

Let $b \in \mathbb{N}_{\geq 2}$, $n \in \mathbb{N}$. The ***Greedy b-representation*** $\mathrm{rep}_b(n)$ of n:

$$c_{\ell-1} \cdots c_0$$

$c_i \in A_b := \{0, \ldots, b-1\}$ such that

$$n = \sum_{i=0}^{\ell-1} c_i b^i, \quad c_{\ell-1} \neq 0.$$

- $\mathrm{val}_b(c_{\ell-1} \cdots c_0) = n$
- $\mathrm{rep}_b(0) = \varepsilon$, $\mathrm{val}_b(\varepsilon) = 0$

- $u = u_1 \cdots u_n \in A^*$, $v = v_1 \cdots v_n \in B^*$

$$(u, v) = (u_1, v_1) \cdots (u_n, v_n) \in (A \times B)^*.$$

- Denote $\ell = \max\{|\mathrm{rep}_b(n_1)|, |\mathrm{rep}_b(n_2)|\}$,

$$\mathrm{rep}_b(n_1, n_2) = (0^{\ell - |\mathrm{rep}_b(n_1)|}\mathrm{rep}_b(n_1), 0^{\ell - |\mathrm{rep}_b(n_2)|}\mathrm{rep}_b(n_2)).$$

Definition

For a base b, a subset X of $\mathbb{N}$ is said to be ***b-recognizable*** if the language $\mathrm{rep}_b(X)$ is regular.

Definition

For a base b, a subset X of $\mathbb{N}$ is said to be ***b-recognizable*** if the language $0^*\mathrm{rep}_b(X)$ is regular.

Proposition

Let $b \in \mathbb{N}_{\geq 2}$ and $m \in \mathbb{N}$. If X is b-recognizable, then so is mX.

Multiplicatively independent integers:

$$(p^a = q^b) \Rightarrow (a = b = 0)$$

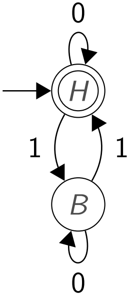
Theorem (Cobham, 1969)

- Let b, b' be two multiplicatively independent bases. Then a subset of $\mathbb{N}$ is both b-recognizable and b'-reconnaissable if and only if it is a finite union of arithmetic progressions.
- Let b, b' be two multiplicatively dependent bases. Then a subset of $\mathbb{N}$ is b-recognizable if and only if it is b'-recognizable.

Thue-Morse set

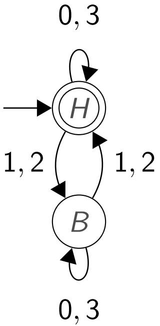
The ***Thue-Morse set*** :

$$\mathcal{T} = \{n \in \mathbb{N} \colon |\mathrm{rep}_2(n)|_1 \in 2\mathbb{N}\}$$



Proposition

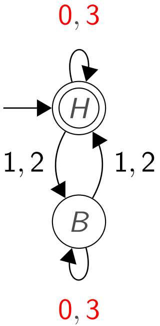
The set $\mathcal{T}$ is 2^p-recognizable for all $p \in \mathbb{N}_{\geq 1}$ and is not b-recognizable for any other base b.



$$A_4 \cap \mathcal{T} = \{0, 3\}$$
$$A_4 \cap (\mathbb{N} \setminus \mathcal{T}) = \{1, 2\}$$

Proposition

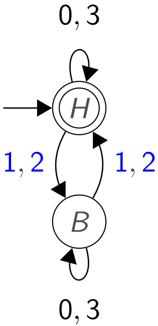
The set $\mathcal{T}$ is 2^p-recognizable for all $p \in \mathbb{N}_{\geq 1}$ and is not b-recognizable for any other base b.



$$A_4 \cap \mathcal{T} = \{0, 3\}$$
$$A_4 \cap (\mathbb{N} \setminus \mathcal{T}) = \{1, 2\}$$

Proposition

The set $\mathcal{T}$ is 2^p-recognizable for all $p \in \mathbb{N}_{\geq 1}$ and is not b-recognizable for any other base b.



$$A_4 \cap \mathcal{T} = \{0, 3\}$$
$$A_4 \cap (\mathbb{N} \setminus \mathcal{T}) = \{1, 2\}$$

For each $p \in \mathbb{N}_{\geq 1}$, the language $0^*\mathrm{rep}_{2^p}(\mathcal{T})$ is accepted by the DFA

$$(\{H, B\}, H, H, A_{2^p}, \delta)$$

where for all $X \in \{H, B\}$ and all $a \in A_{2^p}$,

$$\delta(X, a) := X_a = \begin{cases} X & \text{if } a \in \mathcal{T} \\ \overline{X} & \text{otherwise} \end{cases}$$

where $\overline{H} = B$ and $\overline{B} = H$.

For each $p \in \mathbb{N}_{\geq 1}$, the language $0^*\mathrm{rep}_{2^p}(\mathcal{T})$ is accepted by the DFA

$$(\{H, B\}, H, H, A_{2^p}, \delta)$$

where for all $X \in \{H, B\}$ and all $a \in A_{2^p}$,

$$\delta(X, a) := X_a = \begin{cases} X & \text{if } a \in \mathcal{T} \\ \overline{X} & \text{otherwise} \end{cases}$$

where $\overline{H} = B$ and $\overline{B} = H$.

For each $p \in \mathbb{N}_{\geq 1}$, the language $0^*\mathrm{rep}_{2^p}(\mathcal{T})$ is accepted by the DFA

$$(\{H, B\}, H, H, A_{2^p}, \delta)$$

where for all $X \in \{H, B\}$ and all $a \in A_{2^p}$,

$$\delta(X, a) := X_a = \begin{cases} X & \text{if } a \in \mathcal{T} \\ \overline{X} & \text{otherwise} \end{cases}$$

where $\overline{H} = B$ and $\overline{B} = H$.

Main Theorem

Lemma

For any $m \in \mathbb{N}$ and $p \in \mathbb{N}_{\geq 1}$, the set $m\mathcal{T}$ is 2^p-recognizable.

Theorem

Let $m \in \mathbb{N}$ and $p \in \mathbb{N}_{\geq 1}$. Then the state complexity of the language $0^*\mathrm{rep}_{2^p}(m\mathcal{T})$ is equal to

$$2k + \left\lceil \frac{z}{p} \right\rceil$$

if $m = k2^z$ with k odd.

Method

- Let $\mathcal{A}_{\mathcal{T},2^p}$ the DFA accepting

$$(0,0)^*\{\mathrm{rep}_{2^p}(t,n)\colon t \in \mathcal{T}, n \in \mathbb{N}\}.$$

- Let $\mathcal{A}_{m,2^p}$ the DFA accepting

$$(0,0)^*\{\mathrm{rep}_{2^p}(n,mn)\colon n \in \mathbb{N}\}.$$

Consequently, the DFA $A_{m,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ accepts

$$(0,0)^* \left\{\mathrm{rep}_{2^p}\left(t, mt\right) : t \in \mathcal{T}\right\}$$

and $\Pi_2(\mathcal{A}_{m,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$ accepts

$$0^* \left\{\mathrm{rep}_{2^p}\left(mt\right) : t \in \mathcal{T}\right\}.$$

The state complexity of the multiples of the Thue-Morse set in base 2^p is the number of states of the DFA obtained after the minimisation of $\Pi_2(\mathcal{A}_{m,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$.

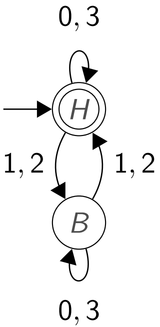
The automaton $\mathcal{A}_{\mathcal{T},2^p}$

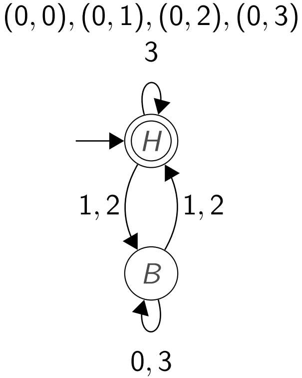
Formally, we have

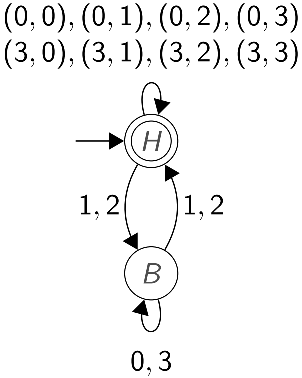
$$\mathcal{A}_{\mathcal{T},2^p} = (\{H, B\}, H, H, A_{2^p} \times A_{2^p}, \delta_{\mathcal{T},2^p})$$

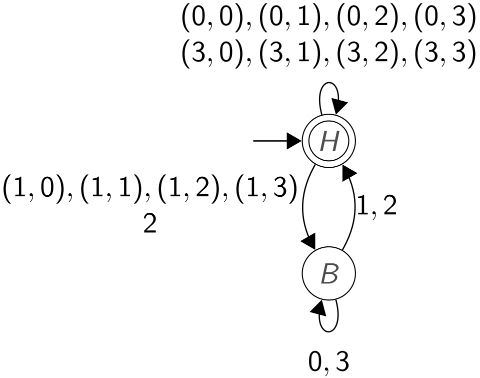
where, for all $X \in \{H, B\}$ and all $d, e \in A_{2^p}$, we have

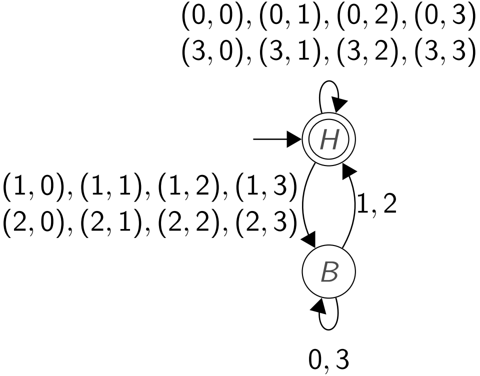
$$\delta_{\mathcal{T},2^p}(X, (d, e)) := X_d.$$

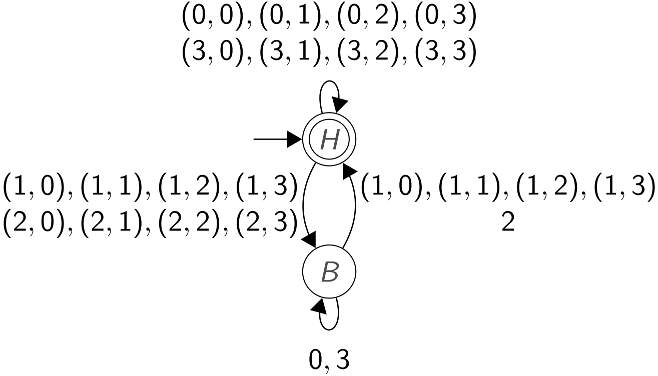

0, 3
H
1, 2
1, 2
B
0, 3

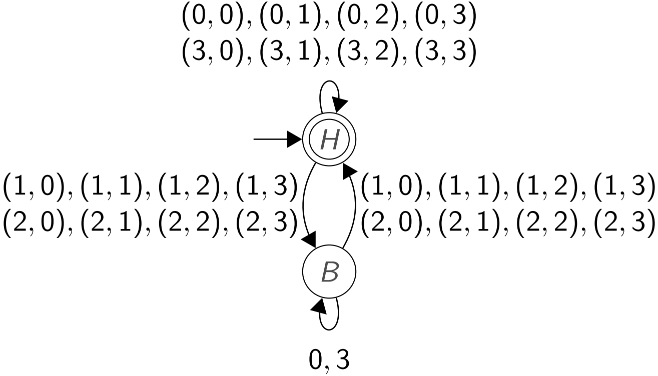

(0, 0), (0, 1), (0, 2), (0, 3)
3
H
1, 2
1, 2
B
0, 3

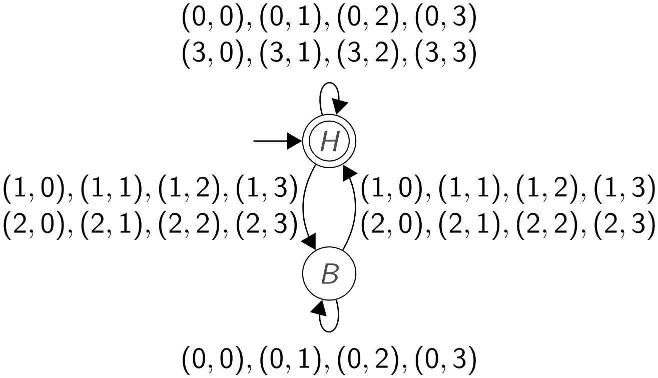

(0, 0), (0, 1), (0, 2), (0, 3)
(3, 0), (3, 1), (3, 2), (3, 3)
H
1, 2
1, 2
B
0, 3


(0, 0), (0, 1), (0, 2), (0, 3)
(3, 0), (3, 1), (3, 2), (3, 3)
H
(1, 0), (1, 1), (1, 2), (1, 3)
2
1, 2
B
0, 3

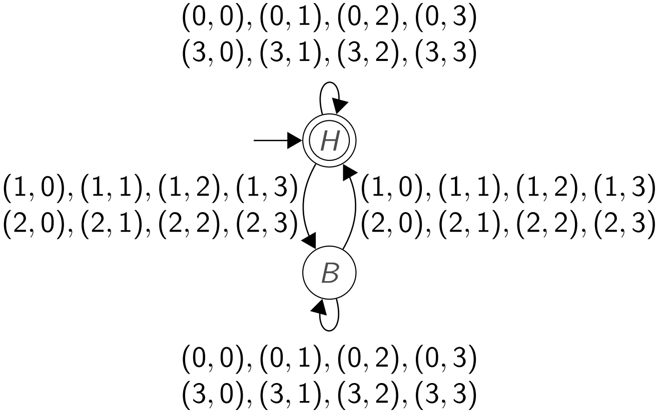

(0, 0), (0, 1), (0, 2), (0, 3)
(3, 0), (3, 1), (3, 2), (3, 3)
H
(1, 0), (1, 1), (1, 2), (1, 3)
(2, 0), (2, 1), (2, 2), (2, 3)
1, 2
B
0, 3


(0, 0), (0, 1), (0, 2), (0, 3)
(3, 0), (3, 1), (3, 2), (3, 3)
H
(1, 0), (1, 1), (1, 2), (1, 3)
(2, 0), (2, 1), (2, 2), (2, 3)
(1, 0), (1, 1), (1, 2), (1, 3)
2
B
0, 3


(0, 0), (0, 1), (0, 2), (0, 3)
(3, 0), (3, 1), (3, 2), (3, 3)
H
(1, 0), (1, 1), (1, 2), (1, 3)
(2, 0), (2, 1), (2, 2), (2, 3)
(1, 0), (1, 1), (1, 2), (1, 3)
(2, 0), (2, 1), (2, 2), (2, 3)
B
0, 3


(0, 0), (0, 1), (0, 2), (0, 3)
(3, 0), (3, 1), (3, 2), (3, 3)
H
(1, 0), (1, 1), (1, 2), (1, 3)
(2, 0), (2, 1), (2, 2), (2, 3)
(1, 0), (1, 1), (1, 2), (1, 3)
(2, 0), (2, 1), (2, 2), (2, 3)
B
(0, 0), (0, 1), (0, 2), (0, 3)

3


(0, 0), (0, 1), (0, 2), (0, 3)
(3, 0), (3, 1), (3, 2), (3, 3)
H
(1, 0), (1, 1), (1, 2), (1, 3)
(2, 0), (2, 1), (2, 2), (2, 3)
(1, 0), (1, 1), (1, 2), (1, 3)
(2, 0), (2, 1), (2, 2), (2, 3)
B
(0, 0), (0, 1), (0, 2), (0, 3)
(3, 0), (3, 1), (3, 2), (3, 3)

Lemma

For all $X, Y \in \{H, B\}$ and $(u, v) \in (A_{2^p} \times A_{2^p})^*$, we have

$$\delta_{\mathcal{T},2^p}(X, (u, v)) = Y \quad \Longleftrightarrow \quad Y = X_{\mathrm{val}_{2^p}(u)}.$$

Lemma

The automaton $\mathcal{A}_{\mathcal{T},2^p}$

- accepts $(0,0)^*\{\mathrm{rep}_{2^p}(t,n) \colon t \in \mathcal{T}, n \in \mathbb{N}\}$
- is accessible
- is coaccessible
- has disjoint states
- is trim minimal
- is complete

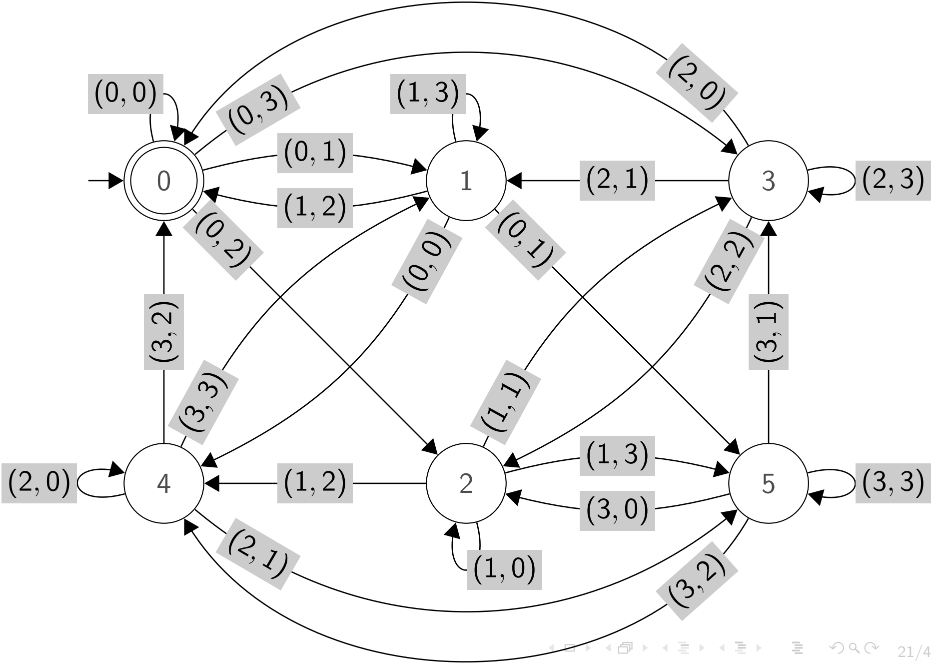
The automaton $\mathcal{A}_{m,b}$

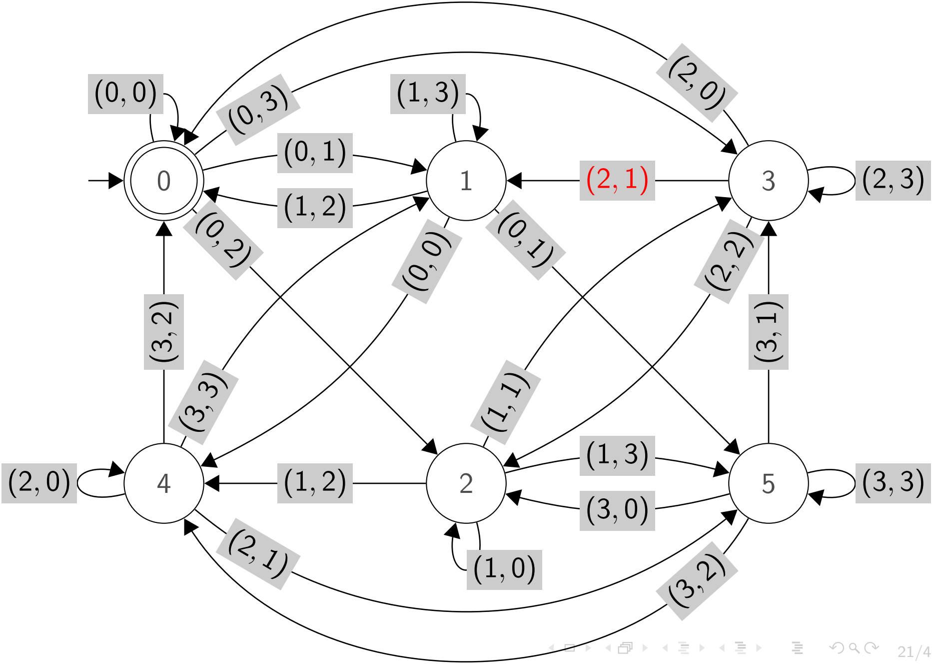
Formally, we have

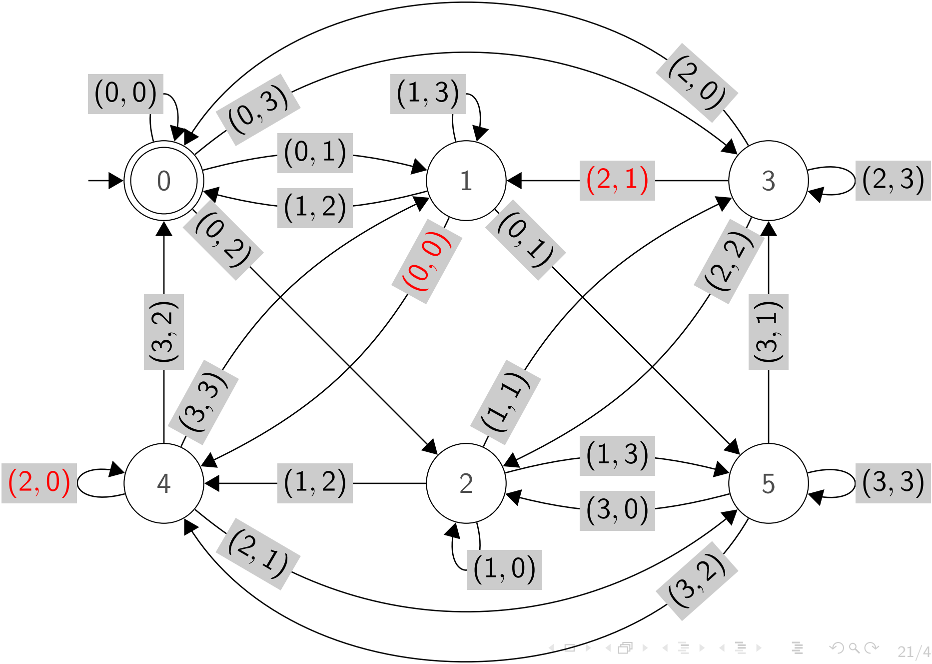
$$\mathcal{A}_{m,b} = (\{0,\ldots,m-1\}, 0, 0, A_b \times A_b, \delta_{m,b})$$

where, for each $i, j \in \{0,\ldots,m-1\}$ and each $d, e \in A_b$,

$$\delta_{m,b}(i,(d,e)) = j \quad \Longleftrightarrow \quad bi + e = md + j.$$


(0, 0)
(0, 3)
(0, 1)
(1, 2)
(1, 3)
(2, 0)
(2, 1)
(2, 3)
(0, 2)
(0, 0)
(0, 1)
(2, 2)
(3, 2)
(3, 1)
(3, 3)
(1, 1)
(2, 0)
(1, 2)
(1, 3)
(3, 0)
(3, 3)
(2, 1)
(1, 0)
(3, 2)
0
1
2
3
4
5


0
1
2
3
4
5
(0, 0)
(1, 3)
(0, 1)
(1, 2)
(0, 3)
(2, 0)
(2, 1)
(2, 3)
(0, 2)
(0, 0)
(0, 1)
(2, 2)
(3, 2)
(3, 3)
(1, 1)
(3, 1)
(2, 0)
(1, 2)
(1, 3)
(3, 0)
(3, 3)
(2, 1)
(1, 0)
(3, 2)


0
1
2
3
4
5
(0, 0)
(0, 3)
(0, 1)
(1, 2)
(1, 3)
(2, 0)
(2, 1)
(2, 3)
(0, 2)
(0, 0)
(0, 1)
(2, 2)
(3, 2)
(3, 3)
(1, 1)
(3, 1)
(2, 0)
(1, 2)
(1, 3)
(3, 0)
(3, 3)
(2, 1)
(1, 0)
(3, 2)

Lemma

For $i, j \in \{0, \ldots, m-1\}$ and $(u, v) \in (A_b \times A_b)^*$, we have

$$\delta_{m,b}(i, (u, v)) = j \iff b^{|(u,v)|}\, i + \mathrm{val}_b(v) = m\, \mathrm{val}_b(u) + j.$$

For instance, we have

$$\delta_{6,4}(3, (202, 100)) = 4$$

because

$$\begin{aligned} 4^3.3 + \mathrm{val}_4(100) &= 208 \\ &= 6.34 + 4 \\ &= 6.\mathrm{val}_4(202) + 4. \end{aligned}$$

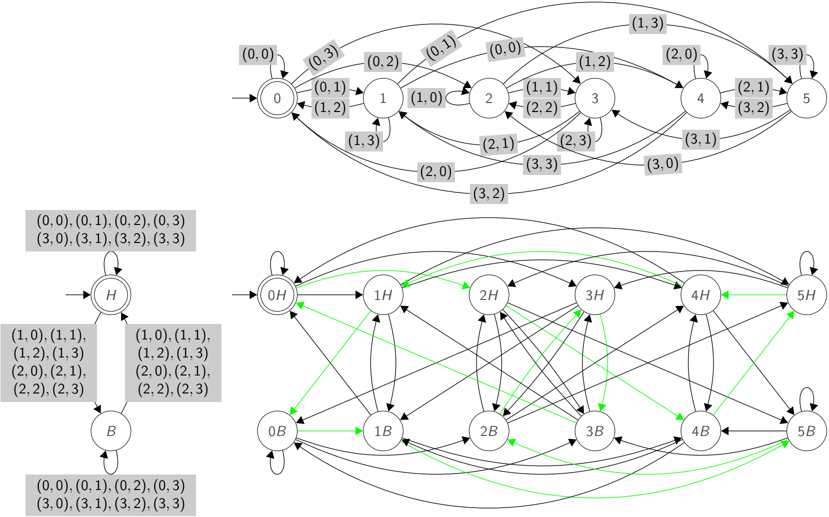
Lemma

The automaton $\mathcal{A}_{m,b}$

- accepts $(0,0)^*\{\mathrm{rep}_b(n, mn) \colon n \in \mathbb{N}\}$
- is accessible
- is coaccessible
- has disjoint states
- is trim minimal

Remark : The automaton $\mathcal{A}_{m,b}$ is not complete.

The product automaton $\mathcal{A}_{m,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$



Let

$$Q = \{(0,H), \ldots, (m-1,H), (0,B), \ldots, (m-1,B)\}.$$

We have

$$\mathcal{A}_{m,2^p} \times \mathcal{A}_{\mathcal{T},2^p} = (Q, (0,H), (0,H), A_{2^p} \times A_{2^p}, \delta_\times),$$

where, for each $i, j \in \{0, \ldots, m-1\}$, $X, Y \in \{H, B\}$ and each $d, e \in A_{2^p}$,

$$\delta_\times((i,X),(d,e)) = (j,Y)$$
$$\iff$$
$$2^p i + e = md + j \quad \text{and} \quad Y = X_d.$$

Lemma

For $i, j \in \{0, \ldots, m-1\}$, $X, Y \in \{H, B\}$ and $(u, v) \in (A_{2^p} \times A_{2^p})^*$, we have

$$\delta_\times((i, X), (u, v)) = (j, Y)$$
$$\Longleftrightarrow$$
$$2^{p\,|(u,v)|}\, i + \mathrm{val}_{2^p}(v) = m\,\mathrm{val}_{2^p}(u) + j \quad \text{and} \quad Y = X_{\mathrm{val}_{2^p}(u)}.$$

Lemma

The automaton $\mathcal{A}_{m,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$

- accepts $(0,0)^* \{\mathrm{rep}_{2^p}(t, mt) : t \in \mathcal{T}\}$
- is accessible
- is coaccessible
- has disjoint states
- is trim minimal

Remark : The automaton $\mathcal{A}_{m,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$ is not complete.

Projection of $\mathcal{A}_{m,2^p} \times \mathcal{A}_{\mathcal{T},2^p}$

Let

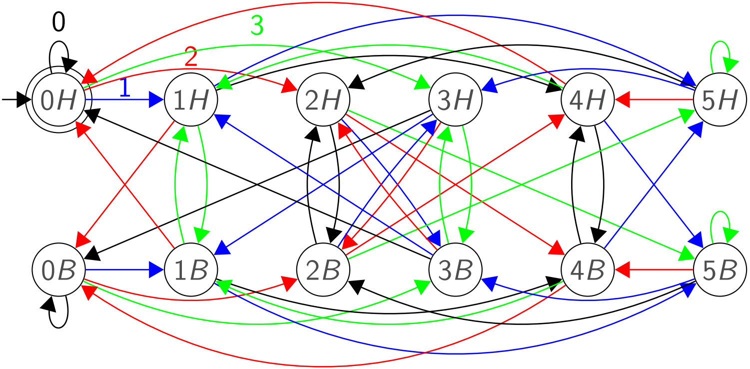
$$Q = \{(0,H),\ldots,(m-1,H),(0,B),\ldots,(m-1,B)\}.$$

We have

$$\Pi_2(\mathcal{A}_{m,2^p} \times \mathcal{A}_{\mathcal{T},2^p}) = (Q,(0,H),(0,H),A_{2^p},\delta_\Pi),$$

where, for each $i,j \in \{0,\ldots,m-1\}$, $X,Y \in \{H,B\}$ and each $e \in A_{2^p}$,

$$\delta_\Pi((i,X),e) = (j,Y)$$
$$\Longleftrightarrow$$
$$\exists d \in A_{2^p} \quad : \quad 2^p i + e = md + j \quad \text{and} \quad Y = X_d.$$


0
3
2
1
0H
1H
2H
3H
4H
5H
0B
1B
2B
3B
4B
5B

Lemma

The automaton $\Pi_2(\mathcal{A}_{m,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$

- accepts $0^* \{\mathrm{rep}_{2^p}(mt) : t \in \mathcal{T}\}$
- is deterministic
- is accessible
- is coaccessible
- has disjoint states **if** m **is odd**
- is trim minimal **if** m **is odd**

Corollary

The state complexity of $m\mathcal{T}$ in base 2^p is $2m$ **if** m **is odd**.

In that case,

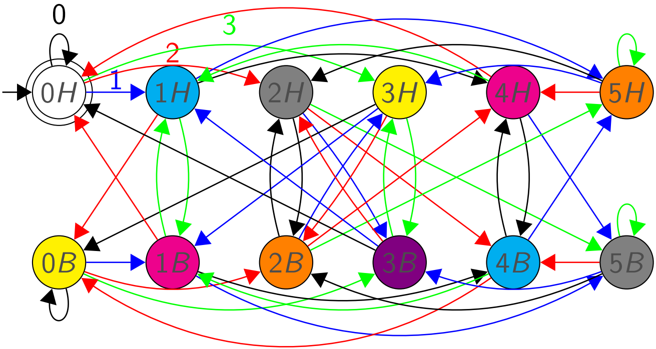
$$m = k \quad \text{and} \;\; z = 0$$

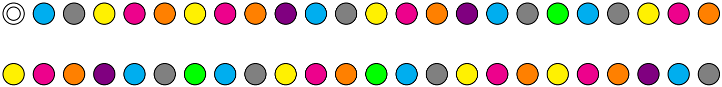
so

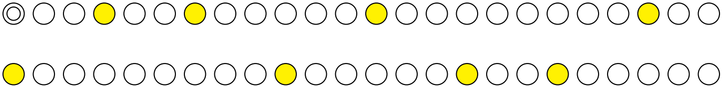
$$2m = 2k + \left\lceil \frac{z}{p} \right\rceil .$$

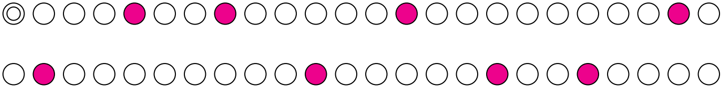
The question will be solved for even m's after the minimisation of the DFA $\Pi_2(\mathcal{A}_{m,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$.

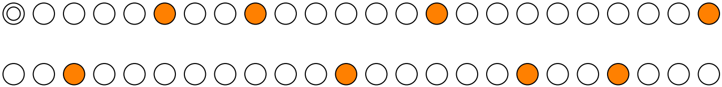
Minimisation of $\Pi_2(\mathcal{A}_{m,2^p} \times \mathcal{A}_{\mathcal{T},2^p})$

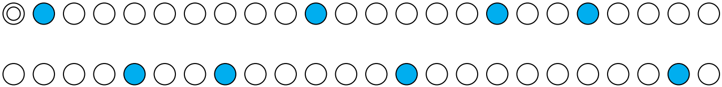


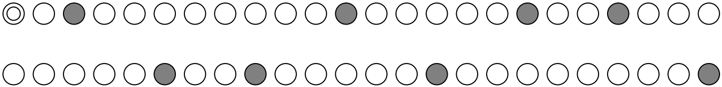


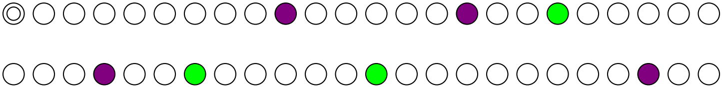


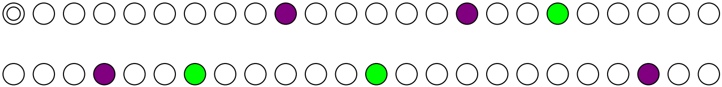












Description of the classes

Let $m = k.2^z$ where $k, z \in \mathbb{N}$ with k odd.

For $(j, X) \in (\{1, \dots, k-1\} \times \{H, B\}) \cup \{(0, B)\}$, the classe of (j, X) is

$$[(j, X)] = \{(j + k\ell, X_\ell) : 0 \leq \ell \leq 2^z - 1\}.$$

Moreover, the classe of $(0, H)$ is

$$[(0, H)] = \{(0, H)\}.$$

For $\alpha \in \{0, \ldots, z-1\}$, we define a *pre-classe* C_α of size 2^α:

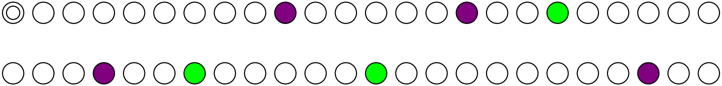
$$C_\alpha := \left[\left(k2^{z-\alpha-1}, B\right)\right] = \left\{\left(k2^{z-\alpha-1} + k2^{z-\alpha}\ell, B_\ell\right) : 0 \leq \ell \leq 2^\alpha - 1\right\}$$

Then, for all $\beta \in \{0, \ldots, \left\lceil \frac{z}{p} \right\rceil - 2\}$, we define a *classe* Γ_β as follows:

$$\Gamma_\beta := \bigcup_{\alpha \in \{\beta p, \ldots, (\beta+1)p-1\}} C_\alpha$$

and we set

$$\Gamma_{\left\lceil \frac{z}{p} \right\rceil - 1} := \bigcup_{\alpha \in \left\{\left(\left\lceil \frac{z}{p} \right\rceil - 1\right)p, \ldots, z-1\right\}} C_\alpha$$



In this example $m = 3.2^3$ and $b = 4$.
So, $k = 3, z = 3, p = 2$, and $\left\lceil \frac{z}{p} \right\rceil = 2$. We obtain

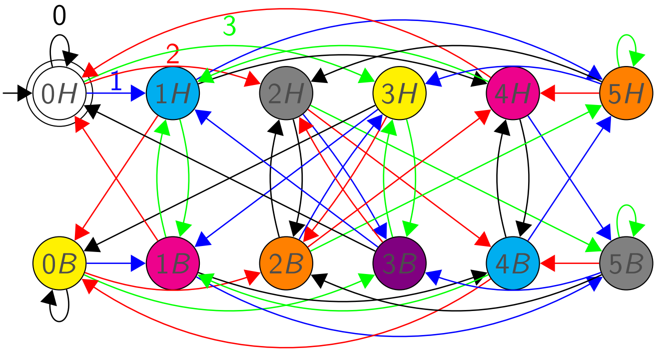
$$C_0 = \{(12, B)\}$$
$$C_1 = \{(6, B), (18, H)\}$$
$$C_2 = \{(3, B), (9, H), (15, H), (21, B)\}$$

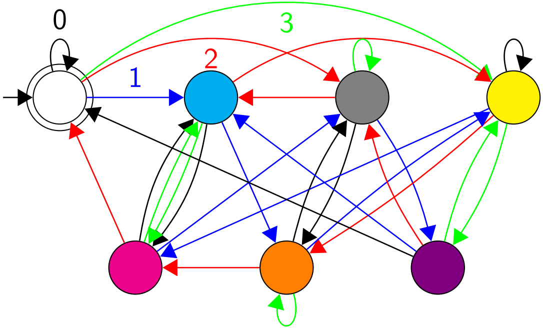
and

$$\Gamma_1 = C_0 \cup C_1 = \{(6, B), (12, B), (18, H)\}$$
$$\Gamma_2 = C_2 = \{(3, B), (9, H), (15, H), (21, B)\}$$

Proof:

(1) The classes consist in indistinguishable states
(2) States belonging to different classes are distinguishable


0
1
2
3
0H
1H
2H
3H
4H
5H
0B
1B
2B
3B
4B
5B


0
1
2
3

Counting and Conclusion

Classes	Number of such classes
$[(j, X)]$ for $(j, X) \in (\{1, \ldots, k-1\} \times \{H, B\})$	$2(k-1)$
$[(0, B)]$	1
$[(0, H)]$	1
Γ_β for $\beta \in \{0, \ldots, \left\lceil \frac{z}{p} \right\rceil - 2\}$	$\left\lceil \frac{z}{p} \right\rceil - 1$
$\Gamma_{\left\lceil \frac{z}{p} \right\rceil - 1}$	1

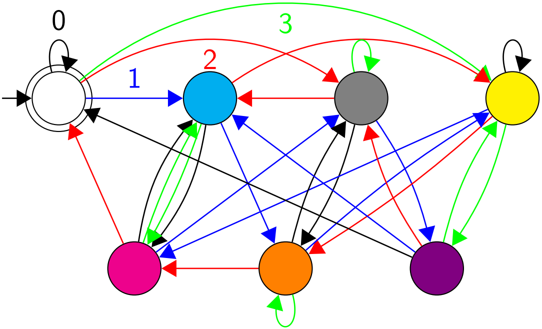
Total $= 2k + \left\lceil \frac{z}{p} \right\rceil$

Theorem

Let $m \in \mathbb{N}$ and $p \in \mathbb{N}_{\geq 1}$. Then the state complexity of the language $0^*\mathrm{rep}_{2^p}(m\mathcal{T})$ is equal to

$$2k + \left\lceil \frac{z}{p} \right\rceil$$

if $m = k2^z$ with k odd.



The state complexity of $6\mathcal{T}$ in base 4 is equal to

$$2.3 + \left\lceil \frac{1}{2} \right\rceil .$$

Thank you!