# A conjecture on the 2-abelian complexity of the Thue-Morse word (Work in progress)

Aline Parreau, Michel Rigo, *Elise Vandomme*

Representing streams II 2014 – Leiden

GRENOBLE 1
UNIVERSITE JOSEPH FOURIER
SCIENCES.TECHNOLOGIE.SANTÉ

Université de Liège

# Thue-Morse word

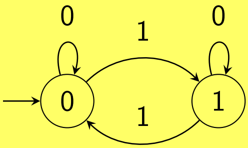
The Thue-Morse word $\mathbf{t} = t_0 t_1 t_2 \cdots$ is the infinite word $\lim_{n\to+\infty} \varphi^n(0)$ where

$$\varphi : 0 \mapsto 01, \qquad 1 \mapsto 10,$$

$$\mathbf{t} = 0110100110010110100101100110 1001 \cdots$$

**The Thue-Morse word t is 2-automatic.**

| $n$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | ... |
|---|---|---|---|---|---|---|---|---|---|---|
| $\mathrm{rep}_2(n)$ | 0 | 1 | 10 | 11 | 100 | 101 | 110 | 111 | 1000 | ... |
| $t_n$ | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | ... |



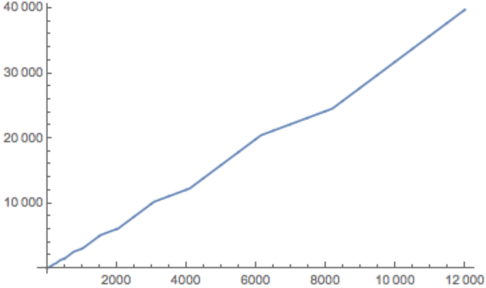
# Thue-Morse word

The factor complexity of the Thue-Morse word

$$p_{\mathbf{t}}(n) = \#\{\text{factors of length } n \text{ of } \mathbf{t}\}$$

is well-known : $p_{\mathbf{t}}(0) = 1, \quad p_{\mathbf{t}}(1) = 2, \quad p_{\mathbf{t}}(2) = 4,$

$$p_{\mathbf{t}}(n) = \begin{cases} 4n - 2 \cdot 2^m - 4 & \text{if } 2 \cdot 2^m < n \leq 3 \cdot 2^m, \\ 2n + 4 \cdot 2^m - 2 & \text{if } 3 \cdot 2^m < n \leq 4 \cdot 2^m. \end{cases}$$





S. Brlek, Enumeration of factors in the Thue-Morse word, **DAM**'89

A. de Luca, S. Varricchio, On the factors of the Thue-Morse word on three symbols, **IPL**'88

# Thue-Morse word

The factor complexity of the Thue-Morse word

$$p_{\mathbf{t}}(n) = \#\{\text{factors of length } n \text{ of } \mathbf{t}\}$$

is well-known : $p_{\mathbf{t}}(0) = 1, \quad p_{\mathbf{t}}(1) = 2, \quad p_{\mathbf{t}}(2) = 4,$

$$p_{\mathbf{t}}(n) = \begin{cases} 4n - 2 \cdot 2^m - 4 & \text{if } 2 \cdot 2^m < n \leq 3 \cdot 2^m, \\ 2n + 4 \cdot 2^m - 2 & \text{if } 3 \cdot 2^m < n \leq 4 \cdot 2^m. \end{cases}$$

**Definition**

Two words $u$ and $v$ are abelian equivalent if $|u|_\sigma = |v|_\sigma$ for any letter $\sigma$.

The abelian complexity of $\mathbf{t}$ takes only two values

$$\mathcal{P}_{\mathbf{t}}(2n) = 3 \text{ and } \mathcal{P}_{\mathbf{t}}(2n+1) = 2.$$

# $k$-abelian equivalence

Let $k \geq 1$ be an integer. Two words $u$ and $v$ in $A^+$ are $k$-abelian equivalent, denoted by $u \equiv_k v$, if

- $\text{pref}_{k-1}(u) = \text{pref}_{k-1}(v)$,
- $\text{suf}_{k-1}(u) = \text{suf}_{k-1}(v)$,
- for all $w \in A^k$, the number of occurences of $w$ in $u$ and in $v$ coincide, $|u|_w = |v|_w$.

**Example**

$A = \{a, b\}$, $u = abbabaabb$, $v = aabbabbab$,

- $u \equiv_2 v$ because $\text{pref}_1(u) = a = \text{pref}_1(v), \ldots,$ and $|u|_{aa} = 1 = |v|_{aa}, |u|_{ab} = 3 = |v|_{ab}, \ldots$
- $u \not\equiv_3 v$ because $\text{suf}_2(u) = bb \neq ab = \text{suf}_2(v)$
- $abcababb \equiv_3 ababcabb$

# $k$-abelian equivalence

Let $k \geq 1$ be an integer. Two words $u$ and $v$ in $A^+$ are $k$-abelian equivalent, denoted by $u \equiv_k v$, if

- $\text{pref}_{k-1}(u) = \text{pref}_{k-1}(v)$,
- $\text{suf}_{k-1}(u) = \text{suf}_{k-1}(v)$,
- for all $w \in A^k$, the number of occurences of $w$ in $u$ and in $v$ coincide, $|u|_w = |v|_w$.

**Remark**

- $\equiv_k$ is an equivalence relation
- $u \equiv_k v \Rightarrow u \equiv_{k-1} v$, $\forall k \geq 1$
- $u = v \Leftrightarrow u \equiv_k v, \forall k \geq 1$

# 2-abelian complexity of $\mathbf{t}$

The first values of the 2-abelian complexity of the Thue-Morse word

$$\mathcal{P}_{\mathbf{t}}^{(2)}(n) = \#\{\text{factors of length } n \text{ of } \mathbf{t}\}/_{\equiv_2}$$

are

$$\begin{aligned}(\mathcal{P}_{\mathbf{t}}^{(2)}(n))_{n\geq 0} = (&1, 2, 4, 6, 8, 6, 8, 10, 8, 6, 8, 8, 10, 10,\\ &10, 8, 8, 6, 8, 10, 10, 8, 10, 12, 12, 10, 12, 12, 10, 8, 10, 10,\\ &8, 6, 8, 8, 10, 10, 12, 12, 10, 8, 10, 12, 14, 12, 12, 12, 12, 10,\\ &12, 12, 12, 12, 14, 12, 10, 8, 10, 12, 12, 10, 10, 8, 8, 6, 8, 10,\\ &10, 8, 10, 12, 12, 10, 12, 12, 12, 12, 14, 12, 10, 8, 10, 12, 14,\\ &12, 14, 16, 14, 12, 14, 14, 14, 12, 12, 12, 12, 10, 12, 12, \ldots\end{aligned}$$

# 2-abelian complexity of $\mathbf{t}$

The first values of the 2-abelian complexity of the Thue-Morse word

$$\mathcal{P}^{(2)}_{\mathbf{t}}(n) = \#\{\text{factors of length } n \text{ of } \mathbf{t}\}/_{\equiv_2}$$

are

$$\begin{aligned}(\mathcal{P}^{(2)}_{\mathbf{t}}(n))_{n\geq 0} = (&1, 2, 4, 6, 8, 6, 8, 10, 8, 6, 8, 8, 10, 10,\\ &10, 8, 8, 6, 8, 10, 10, 8, 10, 12, 12, 10, 12, 12, 10, 8, 10, 10,\\ &8, 6, 8, 8, 10, 10, 12, 12, 10, 8, 10, 12, 14, 12, 12, 12, 12, 10,\\ &12, 12, 12, 12, 14, 12, 10, 8, 10, 12, 12, 10, 10, 8, 8, 6, 8, 10,\\ &10, 8, 10, 12, 12, 10, 12, 12, 12, 12, 14, 12, 10, 8, 10, 12, 14,\\ &12, 14, 16, 14, 12, 14, 14, 14, 12, 12, 12, 12, 10, 12, 12, \ldots\end{aligned}$$

## Questions

- Is the sequence $(\mathcal{P}^{(2)}_{\mathbf{t}}(n))_{n\geq 0}$ bounded ?
- Is the sequence "regular" ?

# 2-abelian complexity of $\mathbf{t}$

A sequence $(x_n)_{n\geq 0}$ (over $\mathbb{Z}$) is *k-regular* of its $\mathbb{Z}$-module generated by its $k$-kernel

$$\mathcal{K} = \{(x_{k^e n+r})_{n\geq 0} \mid e \geq 0, r < k^e\}$$

is finitely generated.

J.-P. Allouche, J. Shallit, The ring of $k$-regular sequences, *Theoret. Comput. Sci.* **98** (1992)

**Example**

The 2-kernel of $\mathbf{t}$ is

$$\begin{aligned}\mathcal{K} &= \{(t_{2^e n+r})_{n\geq 0} \mid e \geq 0, r < 2^e\} \\ &= \{\mathbf{t}, \bar{\mathbf{t}}\}\end{aligned}$$

where $\bar{\mathbf{t}} = (1 - t_n)_{n\geq 0}$.

# 2-abelian complexity of $\mathbf{t}$

A sequence $(x_n)_{n\geq 0}$ (over $\mathbb{Z}$) is *k-regular* of its $\mathbb{Z}$-module generated by its $k$-kernel

$$\mathcal{K} = \{(x_{k^e n+r})_{n\geq 0} \mid e \geq 0, r < k^e\}$$

is finitely generated.

J.-P. Allouche, J. Shallit, The ring of $k$-regular sequences, *Theoret. Comput. Sci.* **98** (1992)

**Theorem (Eilenberg)**

A sequence $(x_n)_{n\geq 0}$ is $k$-automatic iff its $k$-kernel is finite.

# Related work

**Theorem (Madill, Rampersad)**

The abelian complexity of the paperfolding word

$$0010011000110110001001110011011\cdots$$

is a 2-regular sequence.

**Proposition (Karhumäki, Saarela, Zamboni)**

The abelian complexity of the period doubling word, obtained as the fixed point of $\mu : 0 \mapsto 01, 1 \mapsto 00$, is a 2-regular sequence.

**Question**

Is the abelian complexity of a $k$-automatic sequence always $k$-regular ?

## Conjecture

The 2-abelian complexity of $\mathbf{t}$ is 2-regular.

**Notation** : $\mathbf{x}_{2^e+r} = (\mathcal{P}_{\mathbf{t}}^{(2)}(2^e n + r))_{n\geq 0}$.
We conjecture the following relations (`Mathematica` experiments)

$$
\begin{aligned}
\mathbf{x}_5 &= \mathbf{x}_3\\
\mathbf{x}_9 &= \mathbf{x}_3\\
\mathbf{x}_{12} &= -\mathbf{x}_6 + \mathbf{x}_7 + \mathbf{x}_{11}\\
\mathbf{x}_{13} &= \mathbf{x}_7\\
\mathbf{x}_{16} &= \mathbf{x}_8\\
\mathbf{x}_{17} &= \mathbf{x}_3\\
\mathbf{x}_{18} &= \mathbf{x}_{10}\\
\mathbf{x}_{20} &= -\mathbf{x}_{10} + \mathbf{x}_{11} + \mathbf{x}_{19}\\
\mathbf{x}_{21} &= \mathbf{x}_{11}\\
\mathbf{x}_{22} &= -\mathbf{x}_3 - 2\mathbf{x}_6 + \mathbf{x}_7 + 3\mathbf{x}_{10} + \mathbf{x}_{11} - \mathbf{x}_{19}\\
\mathbf{x}_{23} &= -\mathbf{x}_3 - 3\mathbf{x}_6 + 2\mathbf{x}_7 + 3\mathbf{x}_{10} + \mathbf{x}_{11} - \mathbf{x}_{19}\\
\mathbf{x}_{24} &= -\mathbf{x}_3 + \mathbf{x}_7 + \mathbf{x}_{10}\\
\mathbf{x}_{25} &= \mathbf{x}_7\\
\mathbf{x}_{26} &= -\mathbf{x}_3 + \mathbf{x}_7 + \mathbf{x}_{10}\\
\mathbf{x}_{27} &= -2\mathbf{x}_3 + \mathbf{x}_7 + 3\mathbf{x}_{10} - \mathbf{x}_{19}\\
\mathbf{x}_{28} &= -2\mathbf{x}_3 + \mathbf{x}_7 + 3\mathbf{x}_{10} - \mathbf{x}_{14} + \mathbf{x}_{15} - \mathbf{x}_{19}\\
\mathbf{x}_{29} &= \mathbf{x}_{15}\\
\mathbf{x}_{30} &= -\mathbf{x}_3 + 3\mathbf{x}_6 - \mathbf{x}_7 - \mathbf{x}_{10} - \mathbf{x}_{11} + \mathbf{x}_{15} + \mathbf{x}_{19}\\
\mathbf{x}_{31} &= -3\mathbf{x}_3 + 6\mathbf{x}_6 - 2\mathbf{x}_{11} - 3\mathbf{x}_{14} + 2\mathbf{x}_{15} + \mathbf{x}_{19}
\end{aligned}
$$

We also conjecture the following relations

$$
\begin{aligned}
\mathbf{x}_{32} &= \mathbf{x}_8\\
\mathbf{x}_{33} &= \mathbf{x}_3\\
\mathbf{x}_{34} &= \mathbf{x}_{10}\\
\mathbf{x}_{35} &= \mathbf{x}_{11}\\
\mathbf{x}_{36} &= -\mathbf{x}_{10} + \mathbf{x}_{11} + \mathbf{x}_{19}\\
\mathbf{x}_{37} &= \mathbf{x}_{19}\\
\mathbf{x}_{38} &= -\mathbf{x}_3 + \mathbf{x}_{10} + \mathbf{x}_{19}\\
\mathbf{x}_{39} &= -\mathbf{x}_3 + \mathbf{x}_{11} + \mathbf{x}_{19}\\
\mathbf{x}_{40} &= -\mathbf{x}_3 + \mathbf{x}_{10} + \mathbf{x}_{11}\\
\mathbf{x}_{41} &= \mathbf{x}_{11}\\
\mathbf{x}_{42} &= -\mathbf{x}_3 + \mathbf{x}_{10} + \mathbf{x}_{11}\\
\mathbf{x}_{43} &= -2\mathbf{x}_3 + 3\mathbf{x}_{10}\\
\mathbf{x}_{44} &= -2\mathbf{x}_3 - \mathbf{x}_6 + \mathbf{x}_7 + 3\mathbf{x}_{10}\\
\mathbf{x}_{45} &= -\mathbf{x}_3 - 3\mathbf{x}_6 + 2\mathbf{x}_7 + 3\mathbf{x}_{10} + \mathbf{x}_{11} - \mathbf{x}_{19}\\
\mathbf{x}_{46} &= -2\mathbf{x}_3 - 3\mathbf{x}_6 + 2\mathbf{x}_7 + 5\mathbf{x}_{10} + \mathbf{x}_{11} - 2\mathbf{x}_{19}\\
\mathbf{x}_{47} &= -2\mathbf{x}_3 + \mathbf{x}_7 + 3\mathbf{x}_{10} - \mathbf{x}_{19}\\
\mathbf{x}_{48} &= -\mathbf{x}_3 + \mathbf{x}_7 + \mathbf{x}_{10}\\
\mathbf{x}_{49} &= \mathbf{x}_7\\
\mathbf{x}_{50} &= -\mathbf{x}_3 + \mathbf{x}_7 + \mathbf{x}_{10}\\
\mathbf{x}_{51} &= -\mathbf{x}_3 - 3\mathbf{x}_6 + 2\mathbf{x}_7 + 3\mathbf{x}_{10} + \mathbf{x}_{11} - \mathbf{x}_{19}\\
\mathbf{x}_{52} &= -2\mathbf{x}_3 - 3\mathbf{x}_6 + 2\mathbf{x}_7 + 5\mathbf{x}_{10} + \mathbf{x}_{11} - 2\mathbf{x}_{19}\\
\mathbf{x}_{53} &= -2\mathbf{x}_3 + \mathbf{x}_7 + 3\mathbf{x}_{10} - \mathbf{x}_{19}\\
\mathbf{x}_{54} &= -4\mathbf{x}_3 + 3\mathbf{x}_6 + \mathbf{x}_7 + 3\mathbf{x}_{10} - \mathbf{x}_{11} - 2\mathbf{x}_{14} + \mathbf{x}_{15}\\
\mathbf{x}_{55} &= -4\mathbf{x}_3 + 3\mathbf{x}_6 + \mathbf{x}_7 + 3\mathbf{x}_{10} - \mathbf{x}_{11} - 3\mathbf{x}_{14} + 2\mathbf{x}_{15}\\
\mathbf{x}_{56} &= -\mathbf{x}_3 + \mathbf{x}_{10} + \mathbf{x}_{15}\\
\mathbf{x}_{57} &= \mathbf{x}_{15}\\
\mathbf{x}_{58} &= -\mathbf{x}_3 + \mathbf{x}_{10} + \mathbf{x}_{15}\\
\mathbf{x}_{59} &= -2\mathbf{x}_3 + 3\mathbf{x}_6 - \mathbf{x}_7 - \mathbf{x}_{11} + \mathbf{x}_{15} + \mathbf{x}_{19}\\
\mathbf{x}_{60} &= -4\mathbf{x}_3 + 6\mathbf{x}_6 + \mathbf{x}_{10} - 2\mathbf{x}_{11} - 3\mathbf{x}_{14} + 2\mathbf{x}_{15} + \mathbf{x}_{19}\\
\mathbf{x}_{61} &= -3\mathbf{x}_3 + 6\mathbf{x}_6 - 2\mathbf{x}_{11} - 3\mathbf{x}_{14} + 2\mathbf{x}_{15} + \mathbf{x}_{19}\\
\mathbf{x}_{62} &= -\mathbf{x}_3 + 3\mathbf{x}_6 - \mathbf{x}_7 - \mathbf{x}_{10} - \mathbf{x}_{11} + \mathbf{x}_{15} + \mathbf{x}_{19}\\
\mathbf{x}_{63} &= \mathbf{x}_{15}
\end{aligned}
$$

If the conjecture is true, then any sequence $\mathbf{x}_n$ for $n \geq 32$ is a linear combination of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{19}$.

## Proposition

For all $n \geq 0$, $\mathcal{P}_{\mathbf{t}}^{(2)}(2n+1) = \mathcal{P}_{\mathbf{t}}^{(2)}(4n+1)$.

$$
\begin{aligned}
x_{32} &= x_8 \\
x_{33} &= x_3 \\
x_{34} &= x_{10} \\
x_{35} &= x_{11} \\
x_{36} &= -x_{10} + x_{11} + x_{19} \\
x_{37} &= x_{19} \\
x_{38} &= -x_3 + x_{10} + x_{19} \\
x_{39} &= -x_3 + x_{11} + x_{19} \\
x_{40} &= -x_3 + x_{10} + x_{11} \\
x_{41} &= x_{11} \\
x_{42} &= -x_3 + x_{10} + x_{11} \\
x_{43} &= -2x_3 + 3x_{10} \\
x_{44} &= -2x_3 - x_6 + x_7 + 3x_{10} \\
x_{45} &= x_{23} \\
x_{46} &= -2x_3 - 3x_6 + 2x_7 + 5x_{10} + x_{11} - 2x_{19} \\
x_{47} &= -2x_3 + x_7 + 3x_{10} - x_{19} \\
x_{48} &= -x_3 + x_7 + x_{10} \\
x_{49} &= x_7 \\
x_{50} &= -x_3 + x_7 + x_{10} \\
x_{51} &= -x_3 - 3x_6 + 2x_7 + 3x_{10} + x_{11} - x_{19} \\
x_{52} &= -2x_3 - 3x_6 + 2x_7 + 5x_{10} + x_{11} - 2x_{19} \\
x_{53} &= x_{27} \\
x_{54} &= -4x_3 + 3x_6 + x_7 + 3x_{10} - x_{11} - 2x_{14} + x_{15} \\
x_{55} &= -4x_3 + 3x_6 + x_7 + 3x_{10} - x_{11} - 3x_{14} + 2x_{15} \\
x_{56} &= -x_3 + x_{10} + x_{15} \\
x_{57} &= x_{15} \\
x_{58} &= -x_3 + x_{10} + x_{15} \\
x_{59} &= -2x_3 + 3x_6 - x_7 - x_{11} + x_{15} + x_{19} \\
x_{60} &= -4x_3 + 6x_6 + x_{10} - 2x_{11} - 3x_{14} + 2x_{15} + x_{19} \\
x_{61} &= x_{31} \\
x_{62} &= -x_3 + 3x_6 - x_7 - x_{10} - x_{11} + x_{15} + x_{19} \\
x_{63} &= x_{15}
\end{aligned}
$$

# Another approach

Consider the function

$$f : \mathbb{N} \to \mathbb{N}^4, n \mapsto \begin{pmatrix} |p_n|_{00} \\ |p_n|_{01} \\ |p_n|_{10} \\ |p_n|_{11} \end{pmatrix}$$

where $p_n$ is the prefix of length $n$ of the Thue-Morse word.

## Properties

- $f(3 \cdot 2^i + 1) = (2^{i-1}, 2^i, 2^i, 2^{i-1})$
- $f(3 \cdot 2^i) = \begin{cases} (2^{i-1} - 1, 2^i, 2^i, 2^{i-1}) & \text{if } i \text{ is odd} \\ (2^{i-1}, 2^i, 2^i - 1, 2^{i-1}) & \text{if } i \text{ is even} \end{cases}$

## Property

The function $f_{01} : n \mapsto |p_n|_{01}$ is 2-regular.

| $\mathbf{t}$ | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | $\cdots$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $(a_n)$ | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | $\cdots$ |
| $(b_n)$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | $\cdots$ |
| $(f_{01}(n))$ | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 4 | 4 | 5 | 5 | 5 | $\cdots$ |

## Remark

The convolution of two $k$-regular sequences $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$

$$(a_n)_{n\geq 0} \star (b_n)_{n\geq 0} = \left( \sum_{i+j=n} a(i)b(j) \right)_{n\geq 0}$$

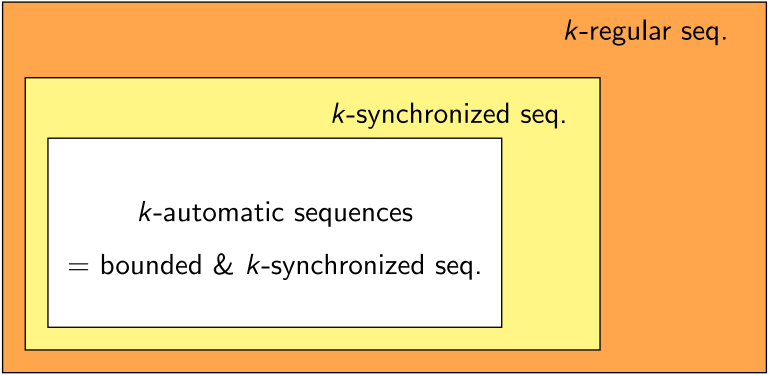
is a $k$-regular sequence.

## Question

Can we find a nice and useful property of the function $f_{01}$ ?

For example, is the sequence $(f_{01}(n))$ 2-synchronized ?

$$\{(\mathrm{rep}_2(n), \mathrm{rep}_2(f_{01}(n))) : n \in \mathbb{N}\} \text{ is accepted by a DFA ?}$$



# Why such a property would be useful ?

If $(f_{01}(n))$ is 2-synchronized,

- $\{(\text{rep}_2(n), \text{rep}_2(f_{01}(n))) : n \in \mathbb{N}\}$ is accepted by a DFA.
- $L = \{(\text{rep}_2(\ell), \text{rep}_2(f_{01}(n+\ell) - f_{01}(n))) : \ell, n \in \mathbb{N}\}$ is accepted by a DFA.
- $\ell \mapsto \#\{(\text{rep}_2(\ell), \_) \in L\}$ forms a 2-regular sequence.

**Theorem (Charlier, Rampersad, Shallit)**

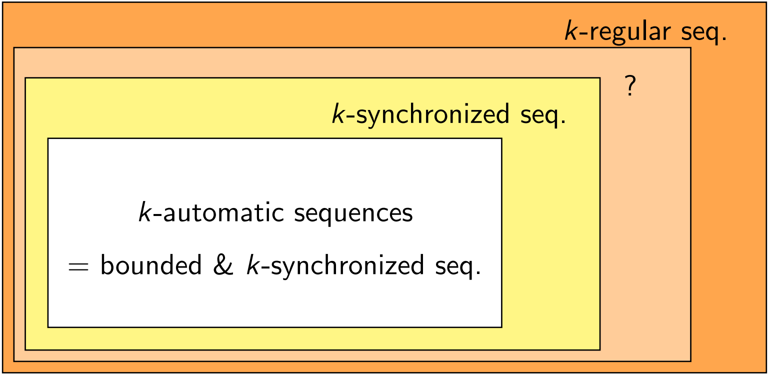
Let $A, B \subset \mathbb{N}$. If the language

$$\{(\text{rep}_k(n), \text{rep}_k(m)) : (n, m) \in A \times B\}$$

is accepted by a DFA, then $n \mapsto \#\{(\text{rep}_k(n), \_) \in L\}$ forms a $k$-regular sequence.

# $(f_{01}(n))$ is not 2-synchronized

- Assume $(f_{01}(n))$ is 2-synchronized.
- Then $(f_{01}(n) - \frac{n}{3})$ is 2-synchronized.
- For $n$ with $\mathrm{rep}_2(n) = (10)^{4\ell}$, $f_{01}(n) - \frac{n}{3} = -\frac{2\ell}{3}$.
- For such n, the subsequence has logarithmic growth and is 2-synchronized.
- Any non-increasing $k$-synchronized sequence is either constant or linear.
- So $(f_{01}(n))$ is not 2-synchronized.



$k$-regular seq.
?
$k$-synchronized seq.
$k$-automatic sequences
$=$ bounded & $k$-synchronized seq.