Factor complexity and critical exponent of words in a Thue-Morse family

Savinien Kreczman

Joint work with Ľubomíra Dvořáková and Edita Pelantová

January 20, 2026

Sponsors

fnrs
LA LIBERTÉ DE CHERCHER

LIÈGE université

Wallonie - Bruxelles International.be

FÉDÉRATION WALLONIE-BRUXELLES

ČVUT
ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

A generalization of the Thue-Morse word

The binary numeration system is based on the linear recurrence sequence given by $U_0 = 1$ and $U_n = U_{n-1} + U_{n-1}$ for all $n \geq 1$. Any natural number can be written uniquely as a sum $U_{i_{\ell-1}} + U_{i_{\ell-2}} \ldots + U_0$ where ℓ depends on n and $i_j \geq i_{j-1} + 1$ for all j.
The Thue-Morse word **t** is given by $t_n = \ell(n) \bmod 2$.

	1	2		4				8								16	
0	1	1	0	1	0	0	1	1	0	0	1	0	1	1	0	1	···

A generalization of the Thue-Morse word

The Zeckendorf numeration system is based on the linear recurrence sequence given by $U_0 = 1$, $U_1 = 2$ and $U_n = U_{n-1} + U_{n-2}$ for all $n \geq 2$. Any natural number can be written uniquely as a sum $U_{i_{\ell-1}} + U_{i_{\ell-2}} \ldots + U_0$ where ℓ depends on n and $i_j \geq i_{j-1} + 2$ for all j.
The Fibonacci-Thue-Morse word $\mathbf{f}$ is given by $f_n = \ell(n) \bmod 2$.

	1	2	3		5			8					13				
0	1	1	1	0	1	0	0	1	0	0	0	1	1	0	0	0	⋯

A generalization of the Thue-Morse word

The Narayana numeration system is based on the linear recurrence sequence given by $U_0 = 1$, $U_1 = 2$, $U_2 = 3$ and $U_n = U_{n-1} + U_{n-3}$ for all $n \geq 3$. Any natural number can be written uniquely as a sum $U_{i_{\ell-1}} + U_{i_{\ell-2}} \ldots + U_0$ where ℓ depends on n and $i_j \geq i_{j-1} + 3$ for all j.
The Allouche-Johnson ('96) word **a** is given by $a_n = \ell(n) \bmod 2$.

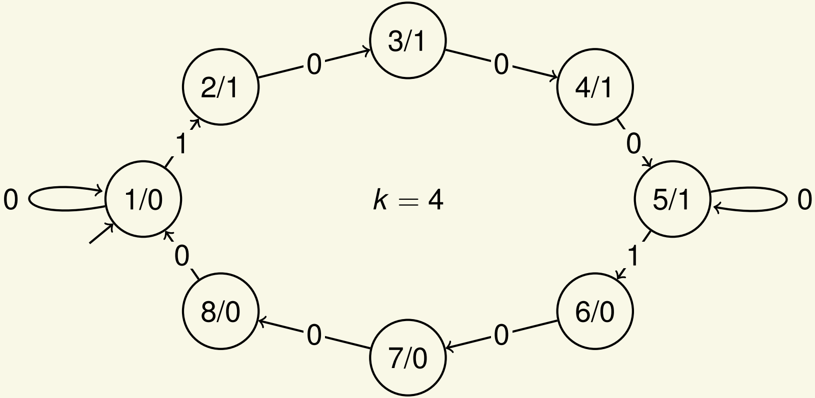
	1	2	3	4		6			9				13				
0	1	1	1	1	0	1	0	0	1	0	0	0	1	0	0	0	···

The word $\mathbf{x}_k$

Consider the numeration system based on the recurrence relation $U_n = U_{n-1} + U_{n-k}$ and $U_0, \ldots, U_{k-1} = 1, \ldots, k$. Define $\mathbf{x}_k$ by

$$(\mathbf{x}_k)_n = |\operatorname{rep}_U(n)|_1 \bmod 2.$$

This word is U-automatic, in the sense that there exists an automaton that outputs $(\mathbf{x}_k)_n$ when reading $\operatorname{rep}_U(n)$.



The word $\mathbf{x}_k$

We know that U-automatic words in exotic numeration systems are morphic (Shallit '88 in this case, later Rigo '00).
We may define $\mathbf{x}_k$ as the projection $\pi(\mathbf{u}_k)$ where $\mathbf{u}_k$ is the fixed point $\xi_k^\omega(0)$, with

$$\xi_k : \begin{array}{rcl} 0 & \to & 01 \\ 1 & \to & 2 \\ 2 & \to & 3 \\ & \vdots & \\ (k-2) & \to & (k-1) \\ (k-1) & \to & 0' \end{array} \quad , \quad \begin{array}{rcl} 0' & \to & 0'1' \\ 1' & \to & 2' \\ 2' & \to & 3' \\ & \vdots & \\ (k-2)' & \to & (k-1)' \\ (k-1)' & \to & 0. \end{array} \tag{1}$$

and

$$\pi(a) = \begin{cases} 0, & \text{if } a = 0 \text{ or } a \in \{1', 2', \dots, (k-1)'\}; \\ 1, & \text{if } a = 0' \text{ or } a \in \{1, 2, \dots, k-1\}. \end{cases} \tag{2}$$

The word $\mathbf{x}_4$

0	1	2	3	$0'$	$0'$	$1'$	$0'$	$1'$	$2'$	$0'$	$1'$	$2'$	$3'$	$0'$	$1'$	$2'$	$3'$	0	$0'$	$1'$	$2'$	$3'$	0	0	1	$\cdots \mathbf{u}_4$
0	1	1	1	1	1	0	1	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	$\cdots \mathbf{x}_4$

In a March 2025 preprint, J. Shallit used `Walnut` to investigate properties of the sequence $\mathbf{x}_3$, as well as the sequence obtained from $\mathbf{u}_3$ by deleting the $'$s (This sequence is linked to Hofstadter functions, see the March 11 One World talk by P. Letouzey). He formulated two conjectures on words in the family $\mathbf{x}_k$, but these cannot be verified for all k using just `Walnut` as the underlying automaton depends on k.

Two conjectures

The *critical exponent* of a sequence $\mathbf{x}$ is defined as

$$E(\mathbf{x}) = \sup\{r \in \mathbb{Q} : u^r \text{ is a non-empty factor of } \mathbf{x}\}.$$

For instance, since the Thue-Morse word $\mathbf{t}$ contains squares but no overlaps, $E(\mathbf{t}) = 2$.
The asymptotic version is called the *asymptotic critical exponent*, denoted $E^*(\mathbf{x})$. If $E(\mathbf{x}) = \infty$, then $E^*(\mathbf{x}) = \infty$.
Otherwise, it is defined as

$$E^*(\mathbf{x}) = \lim_{n\to\infty} \sup\{r \in \mathbb{Q} : u^r \text{ is a non-empty factor of period} \geq n \text{ of } \mathbf{x}\}.$$

Conjecture (Shallit '25)

Let $k \geq 1$. The sequence $\mathbf{x}_k$ has critical exponent $k+1$, which is attained by the words 0^{k+1} and 1^{k+1}. It contains no factor of length $2n+k$ and period n, and therefore has asymptotic critical exponent equal to 2.

Two conjectures

The factor complexity of $\mathbf{x}$ is given by $p_{\mathbf{x}}(n) = |\,\mathrm{Fac}(\mathbf{x}) \cap A^n|$.
Its first difference is $\Delta p_{\mathbf{x}}(n) = p_{\mathbf{x}}(n+1) - p_{\mathbf{x}}(n)$.

Conjecture (Shallit ’25)

Let $k \geq 1$. The first difference of factor complexity of $\mathbf{x}_k$, for n large enough, takes the values $4k-2$ and $4k$ only.

Main result (Dvořáková, K., Pelantová ’25)

Both conjectures are true.

Elementary considerations

0	1	2	3	0′	0′	1′	0′	1′	2′	0′	1′	2′	3′	0′	1′	2′	3′	0	0′	1′	2′	3′	0	0	1

We can make some elementary remarks about $\mathbf{u}_k$ by observing its Rauzy graphs of small order.
If w is a factor of $\mathbf{u}_k$, we call *twin* of w the word obtained by swapping j and j' for all j in $\{0, \dots, k-1\}$ (e.g. $3'00' \leftrightarrow 30'0$).
The twin of w is also a factor of $\mathbf{u}_k$.

The letter $j \in \{1, \dots, k-1\}$ only appears in $\mathbf{u}_k$ as a suffix of the word $01 \cdots j$, and $0'$ when followed by 0 or $0'$ can only occur as a suffix of $01 \cdots (k-1)0'$. The projections of these k words and their twins form the set $\{01^\ell : 1 \le \ell \le k\} \cup \{10^\ell : 1 \le \ell \le k\}$, which is a suffix code.

Bispecial factors

The study of *bispecial factors* will be our angle of attack for both conjectures.

Definition

Let $\mathbf{w} \in A^\omega$. A factor y of $\mathbf{w}$ is

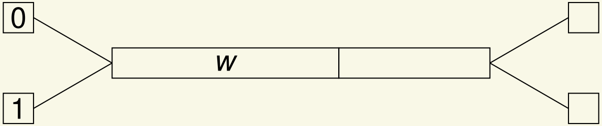
- *left-special* if there exist $a \neq b \in A$ such that ay and by are both factors of $\mathbf{w}$.
- *right-special* if there exist $a \neq b \in A$ such that ya and yb are both factors of $\mathbf{w}$.
- *bispecial* if it is both left- and right-special.

0	1	2	3	$0'$	$0'$	$1'$	$0'$	$1'$	$2'$	$0'$	$1'$	$2'$	$3'$	$0'$	$1'$	$2'$	$3'$	0	$0'$	$1'$	$2'$	$3'$	0	0	1	$\cdots \mathbf{u}_4$
0	1	1	1	1	1	0	1	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	$\cdots \mathbf{x}_4$

Bispecial factors for our purposes

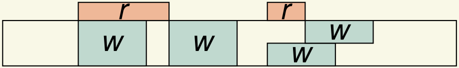
Proposition

- *In a recurrent binary sequence, $\Delta p_{\mathbf{x}}(n)$ is the number of left-special factors of length n of $\mathbf{x}$.*
- *In an aperiodic sequence, every left-special factor of $\mathbf{x}$ is a prefix of exactly one shortest bispecial factor of $\mathbf{x}$.*



Bispecial factors for our purposes

Recall that a *return word* to w in $\mathbf{x}$ is a word r such that rw starts and ends with w and contains no other occurrence of w.



Proposition (Dolce, Dvořáková, Pelantová '23)

Let $\mathbf{u}$ *be a uniformly recurrent aperiodic sequence. Let* $(w_n)_{n\in\mathbb{N}}$ *be the sequence of all bispecial factors in* $\mathbf{u}$ *ordered by length. For every* $n \in \mathbb{N}$*, let* r_n *be the shortest return word to the bispecial factor* w_n *in* $\mathbf{u}$*. Then*

$$E(\mathbf{u}) = 1+\sup\left\{\frac{|w_n|}{|r_n|} \ : \ n \in \mathbb{N}\right\} \quad \textit{and} \quad E^*(\mathbf{u}) = 1+\limsup_{n\to\infty}\frac{|w_n|}{|r_n|}\,.$$

Klouda's method for bispecial factors

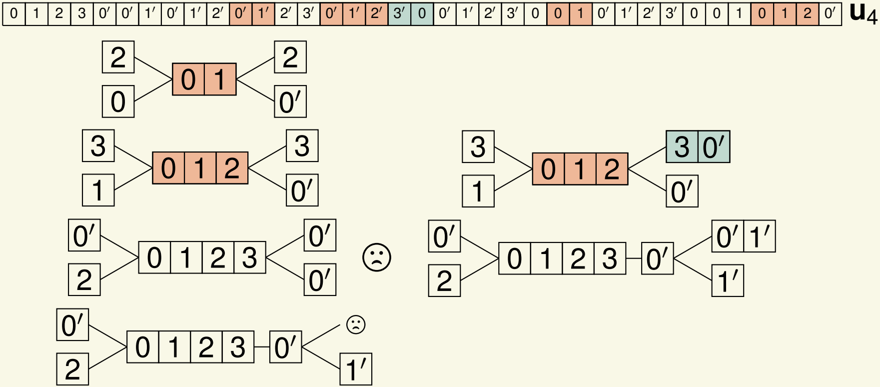
In a 2012 article, Klouda describes a method to find the bispecial factors within a circular non-pushy morphic word. The word $\mathbf{u}_k$ fits this framework. We illustrate this method with $\mathbf{u}_4$.

Idea: we aim to derive from ξ_4 a transformation Ξ_4 that maps bispecial factors to longer bispecial factors. However, we must work with *bispecial triplets*: a bispecial factor together with a pair of left- and right-extensions.

Every bispecial factor will then be obtained within a triplet $(\Xi_4)^n(t)$ where t is in a set of well-chosen initial triplets and n is a natural number.

Method sketch

$$\xi_4 : 0 \mapsto 01,\ 1 \mapsto 2,\ 2 \mapsto 3,\ 3 \mapsto 0',\ \ldots$$



We must carefully select pairs of left- and right-extensions. This will be formalized by the notion of (left-/right-)*forky sets*.

What is a forky set?

Definition 18. *Let (w_1, w_2) and (v_1, v_2) be unordered pairs of words. We say that*

(i) (w_1, w_2) *is a* prefix *(*suffix*) of* (v_1, v_2) *if either* w_1 *is a prefix (suffix) of* v_1 *and* w_2 *of* v_2*, or* w_1 *is a prefix (suffix) of* v_2 *and* w_2 *of* v_1*;*

(ii) (w_1, w_2) *and* (v_1, v_2) *are* L-aligned *if*

$$(v_1 = uw_1 \text{ or } w_1 = uv_1) \quad \text{and} \quad (v_2 = u'w_2 \text{ or } w_2 = u'v_2)$$

or

$$(v_1 = uw_2 \text{ or } w_2 = uv_1) \quad \text{and} \quad (v_2 = u'w_1 \text{ or } w_1 = u'v_2)$$

for some words u, u'*.*

Definition 20. *Let* φ *be a morphism with a fixed point* $\mathbf{u}$*. A finite set* $\mathcal{B}_L$ *of unordered pairs* (w_1, w_2) *of nonempty factors of* $\mathbf{u}$ *is called* L-forky *if all the following conditions are satisfied:*

(i) the last letters of w_1 *and* w_2 *are different for all* $(w_1, w_2) \in \mathcal{B}_L$*,*

(ii) no distinct pairs (w_1, w_2) *and* (w_1', w_2') *from* $\mathcal{B}_L$ *are L-aligned,*

(iii) for any $v_1, v_2 \in \mathcal{L}(\mathbf{u}) \setminus \{\epsilon\}$ *with distinct last letters there exists* $(w_1, w_2) \in \mathcal{B}_L$ *such that* (w_1, w_2) *and* (v_1, v_2) *are L-aligned,*

(iv) for any $(w_1, w_2) \in \mathcal{B}_L$ *there exists* $(w_1', w_2') \in \mathcal{B}_L$ *such that*

$$(w_1' f_L(w_1, w_2), w_2' f_L(w_1, w_2))$$

is a suffix of $(\varphi(w_1), \varphi(w_2))$*.*

Analogously we define an R-forky *set.*

(Klouda '12)

Forky sets

A *right forky set* is a set of pairs of words that represent right-extensions of bispecial factors. We want:

- All possible pairs of extensions to be represented
- With no duplicates. (For instance, $\{10, 01\}$ and $\{1, 012\}$ cannot both be in a right forky set. In which category would we put $\{10, 012\}$?)
- When we apply ξ then factor out the longest common prefix, the pair we obtain is also seen in the forky set.

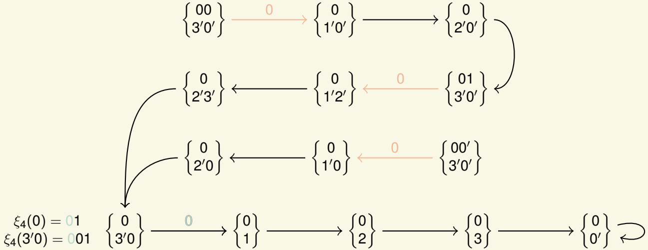
Left forky sets are defined similarly. For all circular non-pushy morphic words, left and right forky sets exist.

A right forky set for $\mathbf{u}_4$

$$\xi_4 \colon 0 \mapsto 01,\ 1 \mapsto 2,\ 2 \mapsto 3,\ 3 \mapsto 0',\ldots$$

0	1	2	3	0′	0′	1′	0′	1′	2′	0′	1′	2′	3′	0′	1′	2′	3′	0	0′	1′	2′	3′	0	0	1	0′	1′	2′	3′	0	0	1	0	1	2	0′	$\mathbf{u}_4$

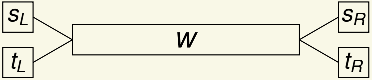
Pictured is the subset of a right forky set for $\mathbf{u}_4$ with all pairs that have a 0 as prefix of one element. Red arrows indicate moving to the corresponding twin pair. Arrows are labeled with the longest common prefix that must be factored out.



The left forky set is simply given by all the pairs of different letters.

Bispecial triplets

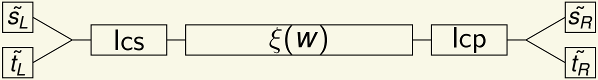
A *bispecial triplet* of **u** is given by a bispecial factor w, and pairs of extensions $\{s_L, t_L\}$ and $\{s_R, t_R\}$ that are in the chosen left (resp. right) forky set and such that $s_L w s_R$ and $t_L w t_R$ are factors of w (or maybe $s_L w t_R$ and $t_L w s_R$).



The bispecial triplet $(\{s_L, t_L\}, w, \{s_R, t_R\})$ can be transformed into the triplet

$$\left(\{\tilde{s}_L, \tilde{t}_L\},\ \mathrm{lcs}(\xi(s_L), \xi(t_L))\xi(w)\,\mathrm{lcp}(\xi(s_R), \xi(t_R)),\ \{\tilde{s}_R, \tilde{t}_R\}\right)\}$$

where $\{\tilde{s}_L, \tilde{t}_L\}$ is the pair in the left forky set that is reached after factoring out the longest common suffix of the images.



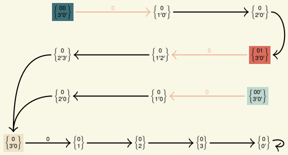
Initial triplets

Naming this transformation Ξ, all bispecial factors can be found within a bispecial triplet obtained by iterating Ξ from a well-chosen initial set.

A *synchronizing point* is a point in a factor of $\mathbf{u}$ that always separates images of letters when desubstituting ξ. For instance, in the Thue-Morse word, $0 \cdot 0$ is a synchronizing point but $0 \cdot 1$ is not.
The *initial bispecial factors* are those that contain no synchronizing point.

In the case of $\mathbf{u}_4$, there is a synchronizing point to the left of every letter except 1 and $1'$, and to the right of every letter except 0 and $0'$. Therefore, the only initial bispecial triplets are those for which $w = \varepsilon$.

Application to $\mathbf{u}_4$



Things only get interesting
when we depart from triplets that have
ε as their factors. Our four starting points
are (from bottom to top on the figure,
a star means any letter can be used):

$T_1 = (\{*, 2'\},\ \varepsilon,\ \{0, 3'0\})$

$T_3 = (\{3', 2'\},\ \varepsilon,\ \{00', 3'0'\})$

$T_2 = (\{*, 2'\},\ \varepsilon,\ \{01, 3'0'\})$

$T_4 = (\{3', 2'\},\ \varepsilon,\ \{00, 3'0'\})$.

From there, we can obtain all bispecial factors of $\mathbf{u}_k$.

$$\xi_4 : 0 \mapsto 01,\ 1 \mapsto 2,\ 2 \mapsto 3,\ 3 \mapsto 0',\ldots$$

As an example, here are all the bispecial triplets arising from T_1:

$$\begin{aligned}
T_1 &= (\{*,2'\},\ \varepsilon,\ \{0,3'0\}) \\
\Xi_4(T_1) &= \big(\{*,3'\},\ 0,\ \{1,0\}\big) \\
\Xi_4^2(T_1) &= \big(\{*,0\},\ \xi_4(0),\ \{2,0\}\big) \\
\Xi_4^3(T_1) &= \big(\{*,1\},\ \xi_4^2(0),\ \{3,0\}\big) \\
\Xi_4^4(T_1) &= \big(\{*,2\},\ \xi_4^3(0),\ \{0',0\}\big) \\
&\ \vdots
\end{aligned}$$

Therefore, every $\xi_4^n(0)$ is bispecial in $\mathbf{u}_4$, with every letter as a left extension and two known right extensions.

Bispecial factors of $\mathbf{u}_k$

Proposition

The bispecial factors within $\mathbf{u}_k$ are the $\xi_k^n(0)$ for any n and the $\xi_k^{n+2k-1}(0)\xi_k^n(0)$ for any n, plus their twins. Moreover, $\xi_k^n(0)$ can be extended to the left by any of the $2k$ letters, while $\xi_k^{n+2k-1}(0)\xi_k^n(0)$ only has two left-extensions.

n	0	1	2	...
Left-extensions of $\xi_4^{n+7}(0)\xi_4^n(0)$	$1', 2'$	$2', 3'$	$3', 0$	...

Through a study of the return words to these bispecial factors, we obtain the following corollary.

Corollary

The word $\mathbf{u}_k$ is overlap-free.

Projecting to $\mathbf{x}_k$

Bispecial factors in $\mathbf{u}_k$ and $\mathbf{x}_k$ do not necessarily match. The following lemmas bridge the gap between $\mathbf{u}_k$ and $\mathbf{x}_k$.

Lemma (1)

Let $u, v \in \mathrm{Fac}(\mathbf{u}_k)$. If $\pi_k(u) = \pi_k(v)$ and u, v disagree on every letter, then $|u| \leq k + 1$. If furthermore $u0, v0 \in \mathrm{Fac}(\mathbf{u}_k)$, or $u0', v0' \in \mathrm{Fac}(\mathbf{u}_k)$, then $|u| \leq k - 1$ and $\pi_k(u) = 0^{|u|}$ or $1^{|u|}$.

Lemma (2)

Let $w \neq \varepsilon$ be a bispecial factor of $\mathbf{u}_k$. Given two right-extensions of w, their projections can agree on at most

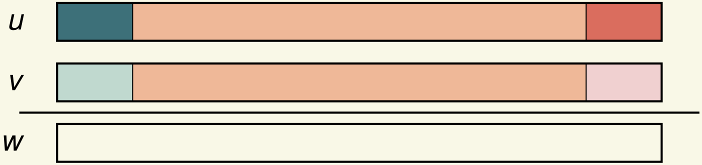
- *the first two letters, if $w = \xi_k^{\ell}(0)$ with $0 \leq \ell \leq k - 2$.*
- *the first letter, if $w = \xi_k^{\ell}(0)$ with $k \leq \ell \leq 2k - 2$.*
- *no letters, otherwise.*

Factor complexity

To recall, the first difference of factor complexity $\Delta p_{\mathbf{x}_k}(n)$ is given by the number of left-special factors of length n in $\mathbf{x}_k$.

Lemma (3)

Let w be a left special factor of $\mathbf{x}_k$ such that $|w|$ is at least $|\xi^{2k-2}(0)| + k + 2$. Then there exist $i \in \mathbb{N}, 1 \leq i \leq k$, and a left special factor f of $\mathbf{u}_k$ such that w or its twin equals $w = 0^{i-1}\pi(f)$.



Factor complexity

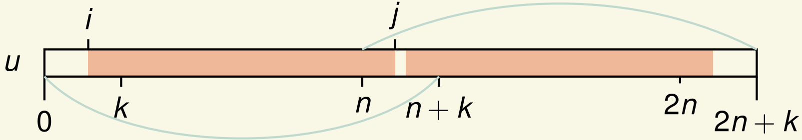
- If f is a prefix of $\xi_k^n(0)$, it can be extended to the left by any letter. Examining further extensions to the left, we obtain that all words $0^\ell\pi(f)$ and $1^\ell\pi(f)$ are left-special in $\mathbf{x}_k$ for $\ell = 0, \ldots, k-1$. Counting words related to the twin of f, we find $4k-2$ left-special words for any sufficiently large length.
- If f is instead a prefix of $\xi_k^{n+2k-1}(0)\xi_k^n(0)$ but not of $\xi_k^n(0)$, it has only two left extensions. This and its twin sporadically give an extra 2 left-special factors, with no overlap between different values of n. Thus we have proven our first conjecture.

Theorem (Dvořáková, K., Pelantová '25)

Let $k \geq 1$. The first difference of factor complexity of $\mathbf{x}_k$, for n large enough, takes the values $4k-2$ and $4k$ only.

Overlaps and asymptotic critical exponent

Assume on the contrary that an overlap exists, i.e. a word of length $2n + k$ and period n, and consider one of its preimages.



The shaded areas represent equal words. The leftmost and rightmost nonshaded areas must be small due to our lemmas, but the shaded areas cannot overlap since there is no overlap in $\mathbf{u}_k$. This gives strong enough conditions on i and j to exclude all possibilities by casework.

Theorem (Dvořáková, K., Pelantová '25)

Let $k \geq 1$. The sequence $\mathbf{x}_k$ contains no factor of length $2n + k$ and period n, and therefore has asymptotic critical exponent equal to 2. Since 0^{k+1} and 1^{k+1} are factors of $\mathbf{x}_k$, the critical exponent is $k + 1$.

Conclusion

- Klouda’s method can identify bispecial factors, and is especially tractable in words where synchronization points are abundant.
- (some manual casework might still be required)

Thank you for your attention!