# String attractors for bi-infinite words

Pierre Béaur, France Gheeraert, Benjamin Hellouin de Menibus

LISN, Université Paris-Saclay

One World Combinatorics on Words Seminar
November 5th, 2024

# Definitions

**Factor of a word $w$**

A contiguous subword of $w$.

$ada \preceq_f abrac\textcolor{red}{ada}bra \qquad aaa \not\preceq_f abracadabra$

# Definitions

### Factor of a word $w$

A contiguous subword of $w$.

$$ada \preceq_f abrac\textcolor{red}{ada}bra \qquad aaa \not\preceq_f abracadabra$$

### String attractor of a word $w$

Set of marked positions $\Gamma$ such that any factor of $w$ has an occurrence crossing it.

$$a\underline{\textcolor{red}{b}}r\,a\,c\,a\underline{\textcolor{red}{d}}abra$$

# Motivation and existing work

**Kempa, Prezza 17**

Dictionary compression algorithms are related to the task of finding the smallest string attractor of the input. This is a NP-complete problem.

# Motivation and existing work

**Kempa, Prezza 17**

Dictionary compression algorithms are related to the task of finding the smallest string attractor of the input. This is a NP-complete problem.

- Size of the smallest attractor as a measure of repetitiveness / complexity.
- Profile function of an infinite word: size of the smallest attractor for the prefix of length $n$.
- String attractors for prefixes of classical words: Thue-Morse, $k$-bonacci, etc.

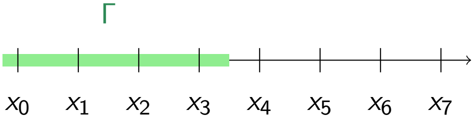
The **span** $\mathrm{span}(\Gamma)$ of an attractor is its diameter.

# The one-sided infinite case

**Restivo, Romana, Sciortino 21**

$w \in \mathcal{A}^{\mathbb{N}}$ admits a finite string attractor if and only if $w$ is ultimately periodic.

$(\Rightarrow)$

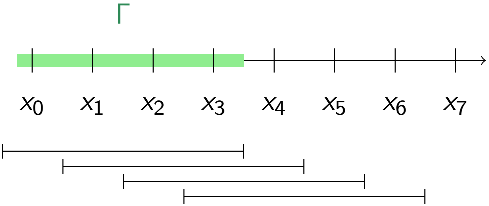


# The one-sided infinite case

**Restivo, Romana, Sciortino 21**

$w \in \mathcal{A}^{\mathbb{N}}$ admits a finite string attractor if and only if $w$ is ultimately periodic.

$(\Rightarrow)$

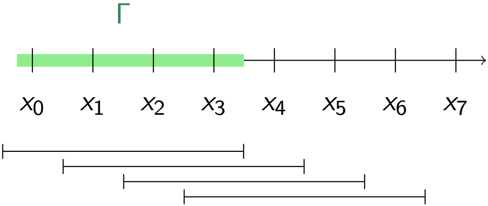


# The one-sided infinite case

**Restivo, Romana, Sciortino 21**

$w \in \mathcal{A}^{\mathbb{N}}$ admits a finite string attractor if and only if $w$ is ultimately periodic.

$(\Rightarrow)$



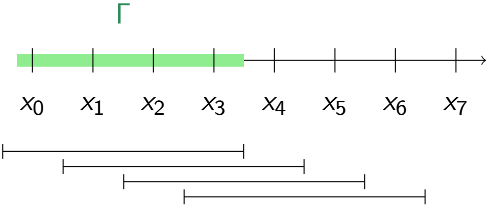

At most $\mathrm{span}(\Gamma) + 1$ factors of length $n \Rightarrow$ ultimately periodic.

# The one-sided infinite case

## Restivo, Romana, Sciortino 21

$w \in \mathcal{A}^{\mathbb{N}}$ admits a finite string attractor if and only if $w$ is ultimately periodic.

$(\Rightarrow)$




At most $\mathrm{span}(\Gamma) + 1$ factors of length $n \Rightarrow$ ultimately periodic.

## Infinite string attractors

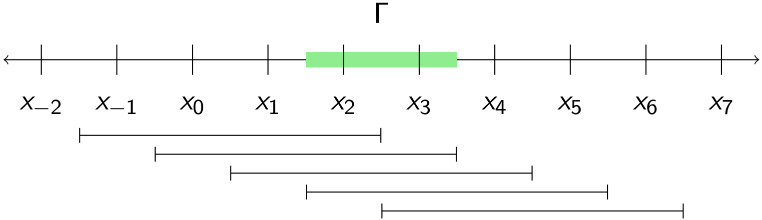
No good notion of "size of the smallest infinite string attractor" (sparsest, lowest density...).

# The two-sided infinite case

## Proposition

If $w \in \mathcal{A}^{\mathbb{Z}}$ admits a finite string attractor $\Gamma$, then

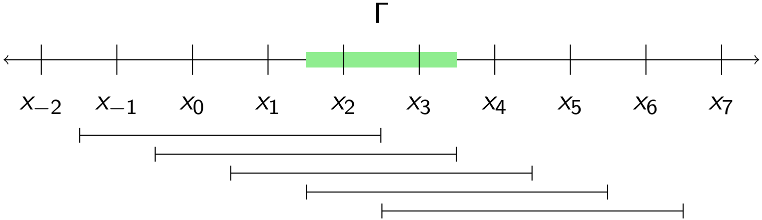
.



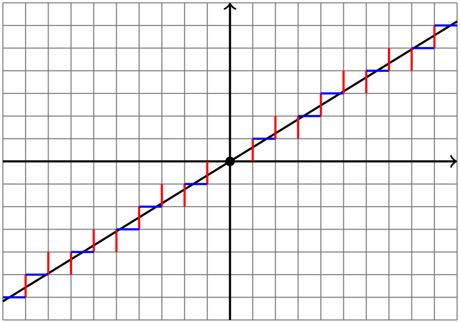
# The two-sided infinite case

**Proposition**

If $w \in \mathcal{A}^{\mathbb{Z}}$ admits a finite string attractor $\Gamma$, then it contains at most $n + \text{span}(\Gamma)$ factors of length $n$.



# Sturmian words



- discrete line with irrational slope
- $n+1$ factors of length $n$
- **characteristic** word: intercept 0.

left-characteristic: $\ldots babaabaababaaba\ .ba\ abaababaabaabab \ldots$

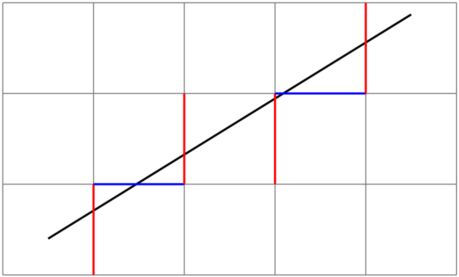
right-characteristic: $\ldots babaabaababaaba\ .ab\ abaababaabaabab \ldots$

# Words of span 1

**Theorem [Béaur, Gheeraert, H. 2024]**

A bi-infinite word admits a string attractor of span 1 if, and only if, it is a characteristic Sturmian word up to shift.

($\Leftarrow$) Take the factor $abaaba$:

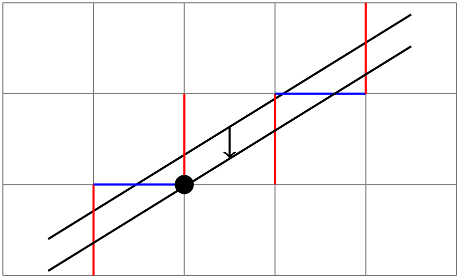


# Words of span 1

**Theorem [Béaur, Gheeraert, H. 2024]**

A bi-infinite word admits a string attractor of span 1 if, and only if, it is a characteristic Sturmian word up to shift.

($\Leftarrow$) Take the factor $abaaba$:

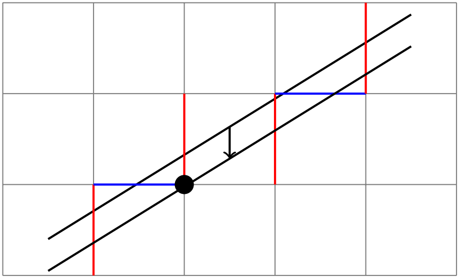


# Words of span 1

**Theorem [Béaur, Gheeraert, H. 2024]**

A bi-infinite word admits a string attractor of span 1 if, and only if, it is a characteristic Sturmian word up to shift.

($\Leftarrow$) Take the factor $abaaba$:



This direction was implicitely proved in [Barbieri, Labbé, Starosta 21].

($\Rightarrow$) Combinatorial proof with special words.

# Other Sturmian words

**Theorem [Béaur, Gheeraert, H. 2024]**

Sturmian words, except shifted characteristic words, have no finite string attractor.

A Sturmian word is the image of another Sturmian word under one of the substitutions:

$$\phi_a : \begin{cases} a \mapsto a \\ b \mapsto ba \end{cases} \quad \phi_b : \begin{cases} a \mapsto ab \\ b \mapsto b \end{cases}$$

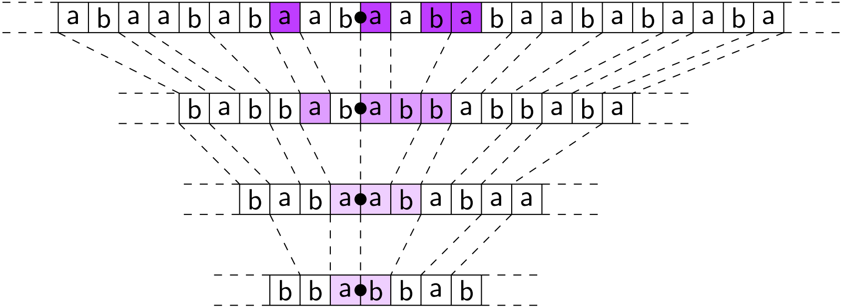
$$\dots |ba|ba|a|ba|a|ba|ba|a|ba|.ba|a|ba|a|ba|ba|a|ba|a|ba|b\dots$$

$$= \phi_a(\dots bbababbab.bababbababb\dots)$$

# Other Sturmian words

**Theorem [Béaur, Gheeraert, H. 2024]**

Sturmian words, except shifted characteristic words, have no finite string attractor.



# Words with finite string attractors

## Quasi-Sturmian words

A word $w$ is quasi-Sturmian if it has $n + k$ factors of length $n$, for $n$ large enough.

## Cassaigne 1997, Heinis 2001, Gheeraert 2023

$w$ is quasi-Sturmian if and only if $w = \phi(s)$ where $s$ is Sturmian, with some nice properties on $\phi$.

# Words with finite string attractors

**Quasi-Sturmian words**

A word $w$ is quasi-Sturmian if it has $n + k$ factors of length $n$, for $n$ large enough.

**Cassaigne 1997, Heinis 2001, Gheeraert 2023**

$w$ is quasi-Sturmian if and only if $w = \phi(s)$ where $s$ is Sturmian, with some nice properties on $\phi$.

**Theorem [Béaur, Gheeraert, H. 2024]**

A word $w$ has a finite string attractor $\Gamma$ if and only if $w$ is a characteristic quasi-Sturmian word.
In this case, $w$ has $n + \text{span}(\Gamma)$ factors of length $n$ for $n$ large enough.

Idea: String attractors behave well under (de)substitutions with nice properties.

# Side remarks

### Subshift

The set of infinite words where no forbidden factor from $\mathcal{F}$ occurs.

### Subshift attractors

A subshift where every word admits a finite string attractor, then it contains only periodic points.

# Side remarks

### Subshift

The set of infinite words where no forbidden factor from $\mathcal{F}$ occurs.

### Subshift attractors

A subshift where every word admits a finite string attractor, then it contains only periodic points.

What about "interesting" infinite string attractors?

### Arithmetic progressions as string attractors

A word admits every arithmetic progression as a string attractor if it is modulo recurrent.
A subshift admits every arithmetic progression as a string attractor if **and only if** it is totally minimal.

# Conclusion

- Complete characterisation of bi-infinite words that admit a finite string attractor.
- A new definition for characteristic words!

This is the same characterisation as for a related notion from symbolic dynamics, indistinguishable pairs [Barbieri, Labbé, Starosta 21].

# Conclusion

- Complete characterisation of bi-infinite words that admit a finite string attractor.
- A new definition for characteristic words!

This is the same characterisation as for a related notion from symbolic dynamics, indistinguishable pairs [Barbieri, Labbé, Starosta 21].

**Future work**

What are the infinite two-dimensional words that admit a finite string attractor?

[Barbieri and Labbé 24] decribe some indistinguishable pairs in higher dimension through cut & project schemes and multidimensional Sturmian words.