A robust extension of ω -regular word languages.

Mikołaj Bojańczyk Warsaw University

- regular expressions
- automata
- monadic second-order logic
- closure properties
- semigroups
- Myhill-Nerode equivalence

- regular expressions
- automata
- monadic second-order logic
- closure properties
- semigroups
- Myhill-Nerode equivalence

What about infinite words?

- regular expressions
- automata
- monadic second-order logic
- closure properties
- semigroups
- Myhill-Nerode equivalence

What about infinite words? ω-regular languages

- regular expressions
- automata
- monadic second-order logic
- closure properties
- semigroups
- Myhill-Nerode equivalence

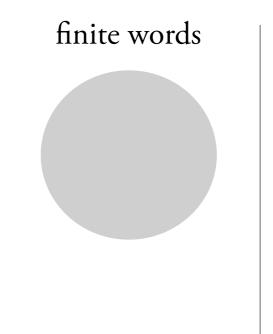
What about infinite words? ω-regular languages

Claim: there are robust extensions of ω -regular languages

- regular expressions
- automata
- monadic second-order logic
- closure properties
- semigroups
- Myhill-Nerode equivalence

What about infinite words? ω-regular languages

Claim: there are robust extensions of ω -regular languages

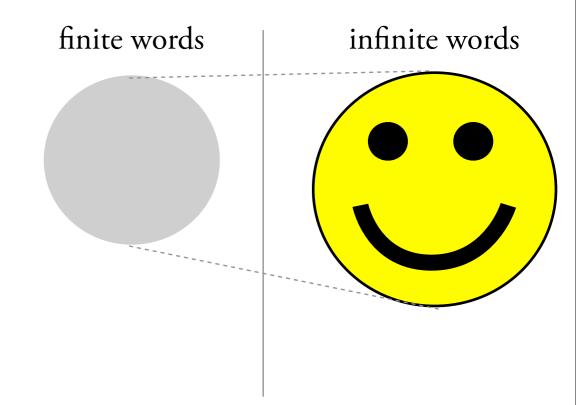


infinite words

- regular expressions
- automata
- monadic second-order logic
- closure properties
- semigroups
- Myhill-Nerode equivalence

What about infinite words? ω-regular languages

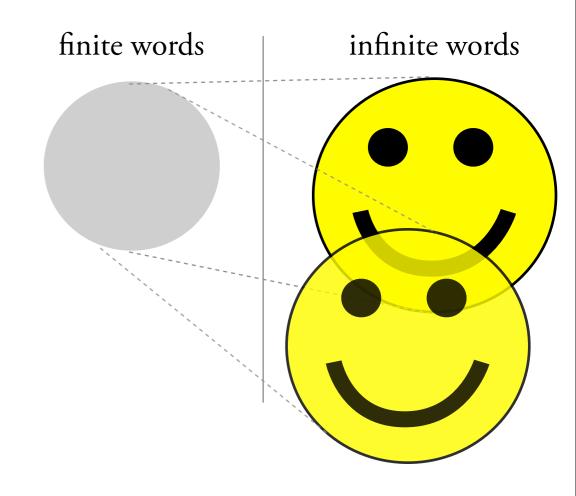
Claim: there are robust extensions of ω -regular languages



- regular expressions
- automata
- monadic second-order logic
- closure properties
- semigroups
- Myhill-Nerode equivalence

What about infinite words? ω-regular languages

Claim: there are robust extensions of ω -regular languages



Which properties of number sequences are regular?

- odd numbers on even positions

Which properties of number sequences are regular?

– odd numbers on even positions– infinitely many odd numbers

- odd numbers on even positions
- infinitely many odd numbers
- the first number n_1 is prime

- odd numbers on even positions
- infinitely many odd numbers
- the first number n_1 is prime
- infinitely many prime numbers

- odd numbers on even positions
- infinitely many odd numbers
- the first number n_1 is prime
- infinitely many prime numbers
- sequence is ultimately constant

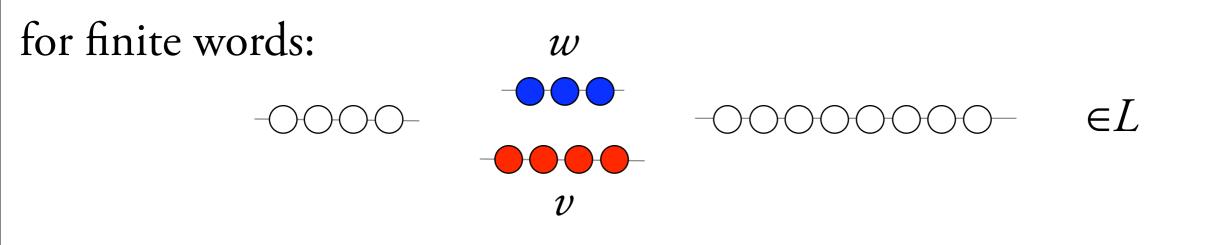
- odd numbers on even positions
- infinitely many odd numbers
- the first number n_1 is prime
- infinitely many prime numbers
- sequence is ultimately constant
- sequence is bounded

- odd numbers on even positions
- infinitely many odd numbers
- the first number n_1 is prime
- infinitely many prime numbers
- sequence is ultimately constant
- sequence is bounded
- sequence tends to ∞

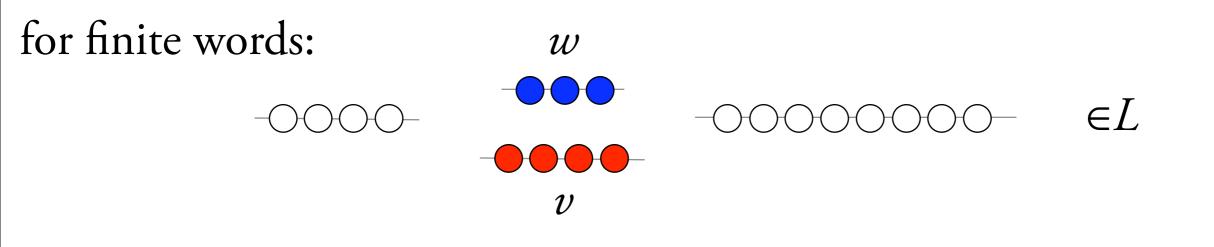
- odd numbers on even positions
- infinitely many odd numbers
- the first number n_1 is prime
- infinitely many prime numbers
- sequence is ultimately constant
- sequence is bounded
- sequence tends to ∞
- exists a bounded subsequence

Two finite words v, w are equivalent for a language L if they can be swapped in any environment, and L will not notice.

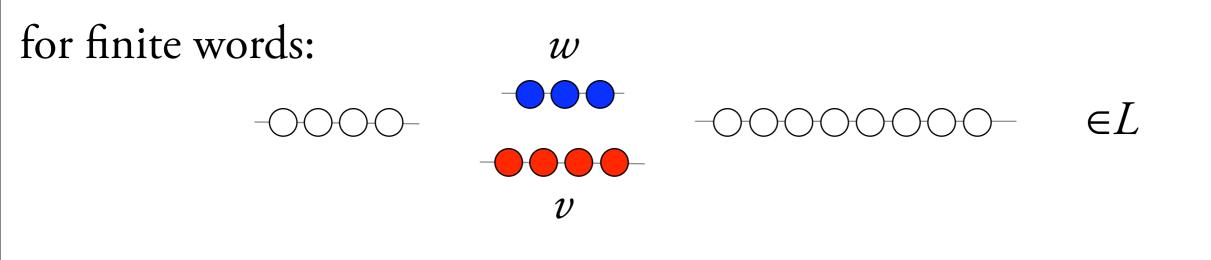
Two finite words v, w are equivalent for a language L if they can be swapped in any environment, and L will not notice.

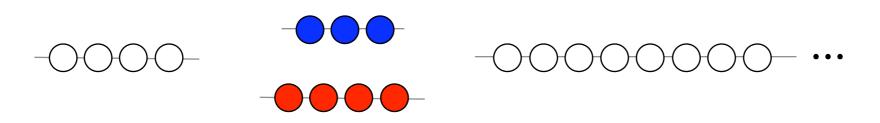


Two finite words v, w are equivalent for a language L if they can be swapped in any environment, and L will not notice.

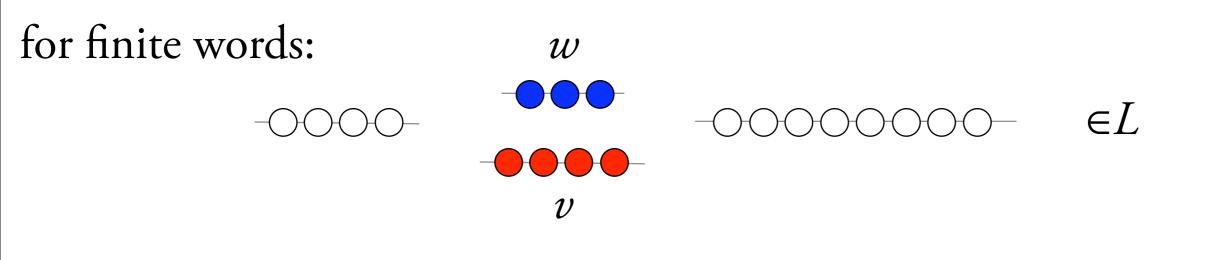


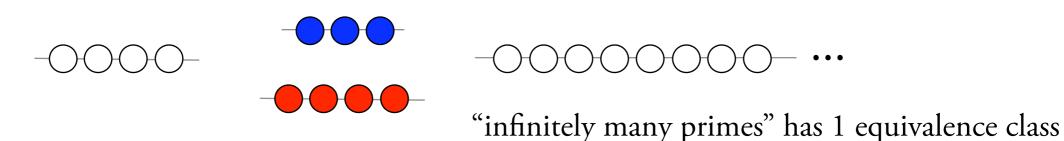
Two finite words v, w are equivalent for a language L if they can be swapped in any environment, and L will not notice.



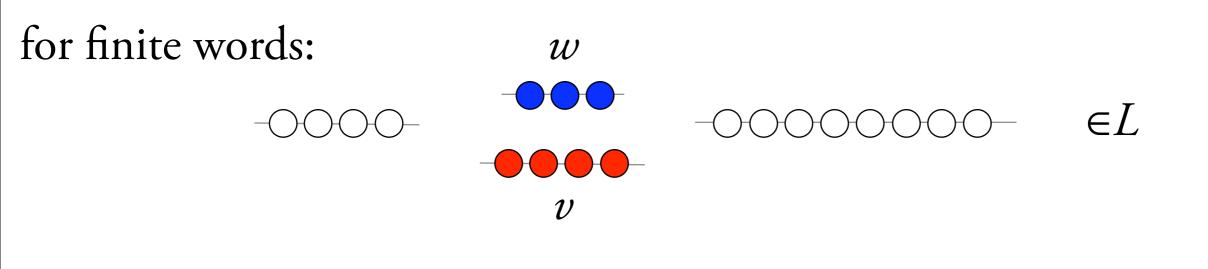


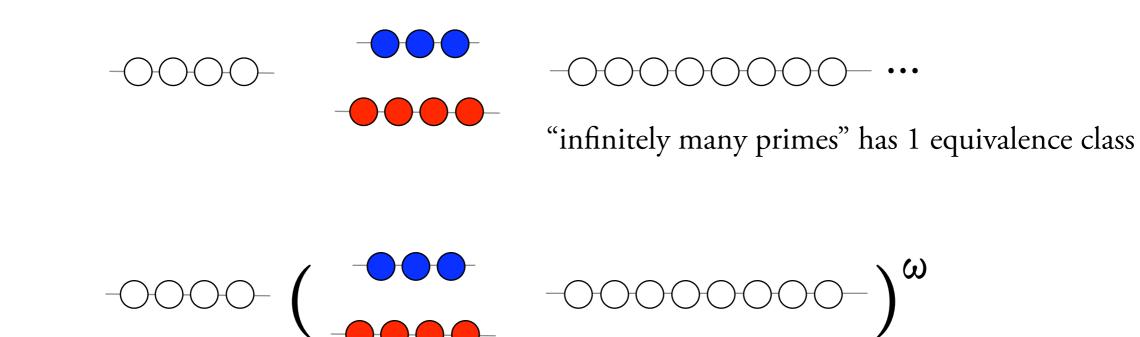
Two finite words v, w are equivalent for a language L if they can be swapped in any environment, and L will not notice.



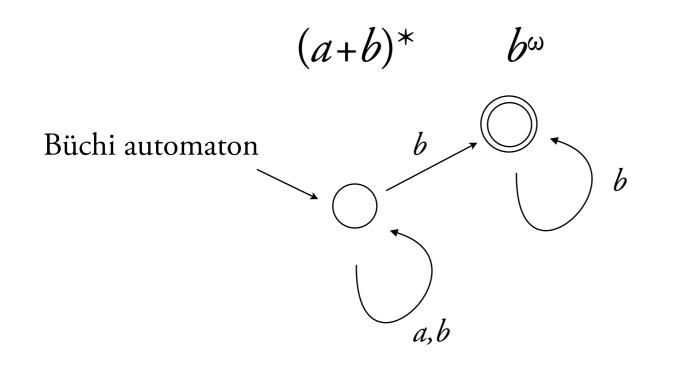


Two finite words v, w are equivalent for a language L if they can be swapped in any environment, and L will not notice.

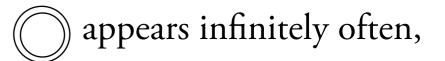


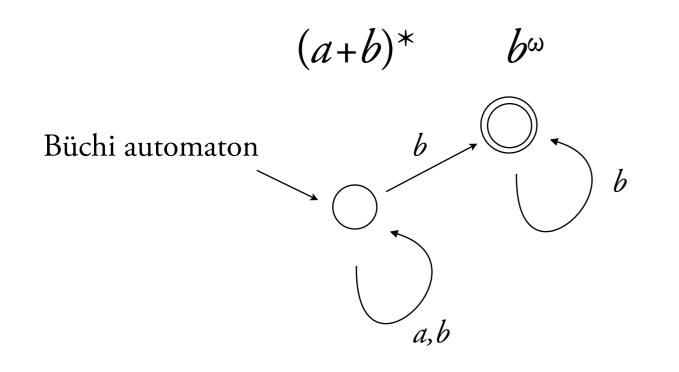


 $(a+b)^*$ b^{ω}

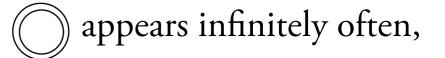


acceptance condition:



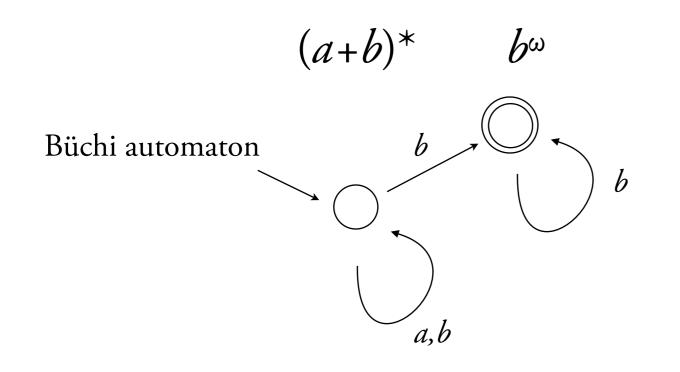


acceptance condition:

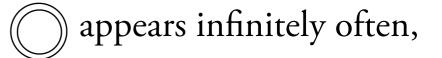


Thm. (McNaughton `66)

For every nondeterministic Büchi automaton, there is an equivalent deterministic Muller automaton.

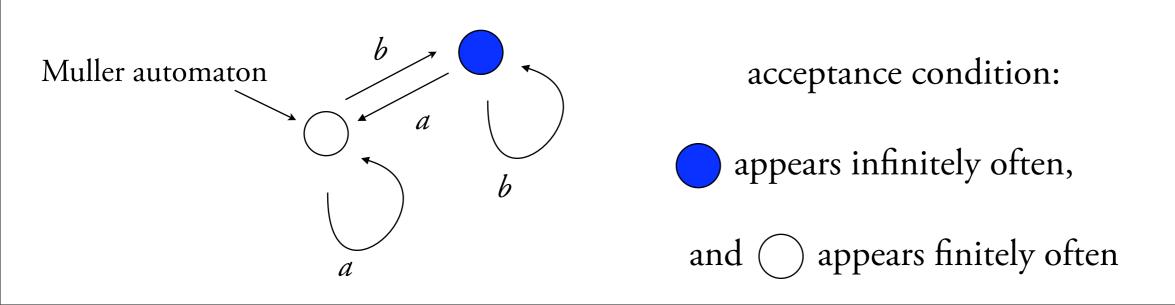


acceptance condition:



Thm. (McNaughton `66)

For every nondeterministic Büchi automaton, there is an equivalent deterministic Muller automaton.



"infinitely many *a*'s on even positions"

"infinitely many *a*'s on even positions"

$$\exists X \exists y \le x \quad y \in X$$
$$\exists X \left\{ \begin{array}{l} \forall x \ \forall y \quad suc(x,y) \Rightarrow (x \in X \Leftrightarrow y \notin X) \\ \forall x \ \forall y \ge x \quad a(y) \land y \in X \end{array} \right.$$

"infinitely many *a*'s on even positions"

contains the first position

There is a set X of positions

contains every second position

 $\forall x \exists y \le x \quad y \in X$ $\exists X \left\{ \begin{array}{c} \forall x \forall y \quad \operatorname{suc}(x, y) \Rightarrow (x \in X \Leftrightarrow y \notin X) \end{array} \right\}$

 $\forall x \; \forall y \ge x \quad a(y) \land y \in X$

contains infinitely many a's

"infinitely many a's on even positions"

contains the first position

There is a set X of positions

contains every second position

$$\exists X \left\{ \begin{array}{ll} \forall x \forall y & \operatorname{suc}(x, y) \Rightarrow (x \in X \Leftrightarrow y \notin X) \\ \forall x \forall y \ge x & a(y) \land y \in X \end{array} \right.$$

contains infinitely many *a*'s

Thm. (Büchi `60) Büchi automata and MSO have the same expressive power.

 $\forall x \exists y \leq x \quad y \in X$

"infinitely many a's on even positions"

contains the first position

There is a set X of positions

contains every second position

 $\exists X \{ \forall x \exists y \le x \quad y \in X \}$ $\exists X \{ \forall x \forall y \quad suc(x,y) \Rightarrow (x \in X \Leftrightarrow y \notin X) \}$ $\forall x \forall y \ge x \quad a(y) \land y \in X$

contains infinitely many a's

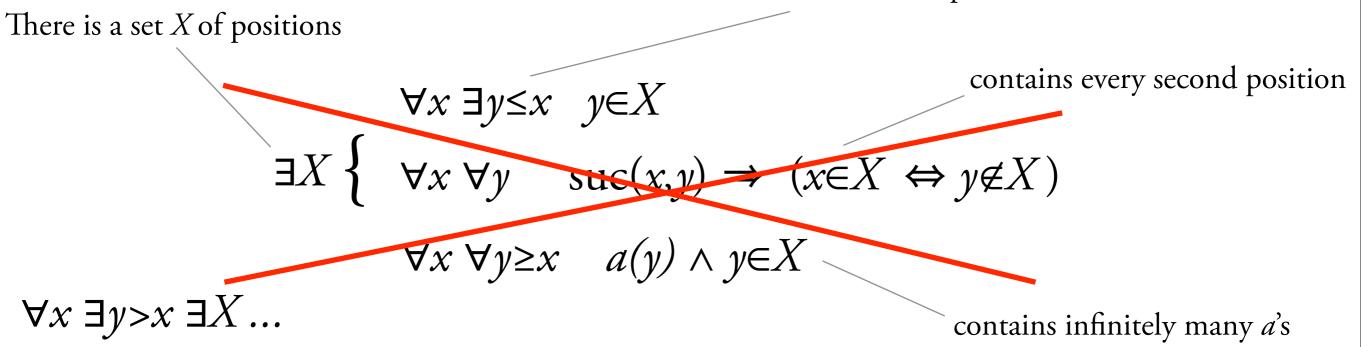
Thm. (Büchi `60)

Büchi automata and MSO have the same expressive power.

Corollary of determinization. For infinite words, MSO = Weak MSO (Weak MSO: set quantification only over finite sets.)

"infinitely many a's on even positions"

contains the first position

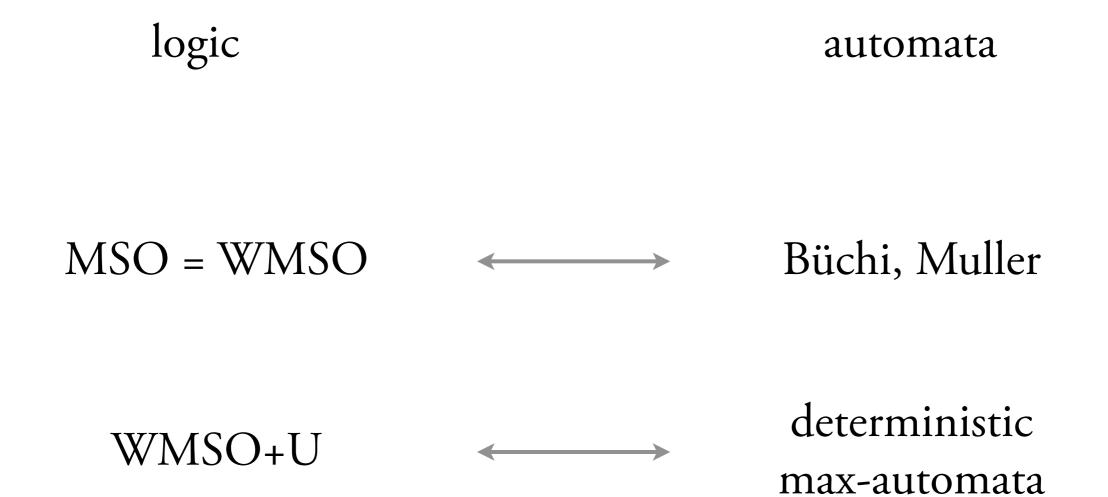


Thm. (Büchi `60)

Büchi automata and MSO have the same expressive power.

Corollary of determinization. For infinite words, MSO = Weak MSO (Weak MSO: set quantification only over finite sets.)





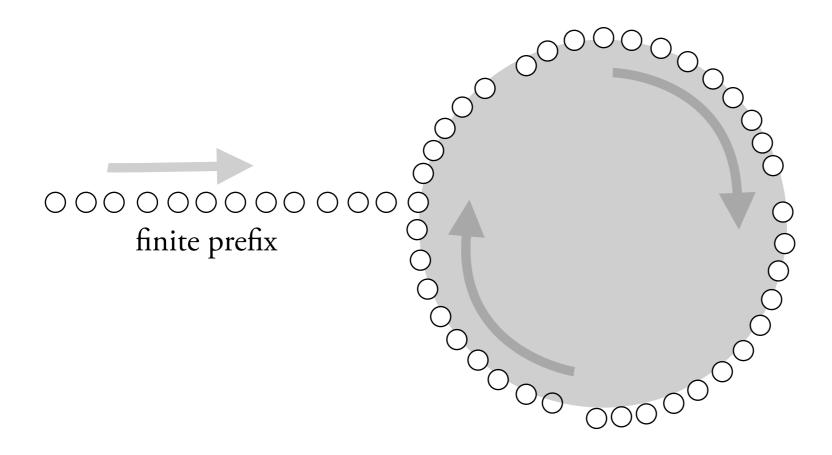
Has finite state space Q and a finite set of counters C. The counters are only read by the acceptance condition, and not during the run.

Has finite state space Q and a finite set of counters C. The counters are only read by the acceptance condition, and not during the run.

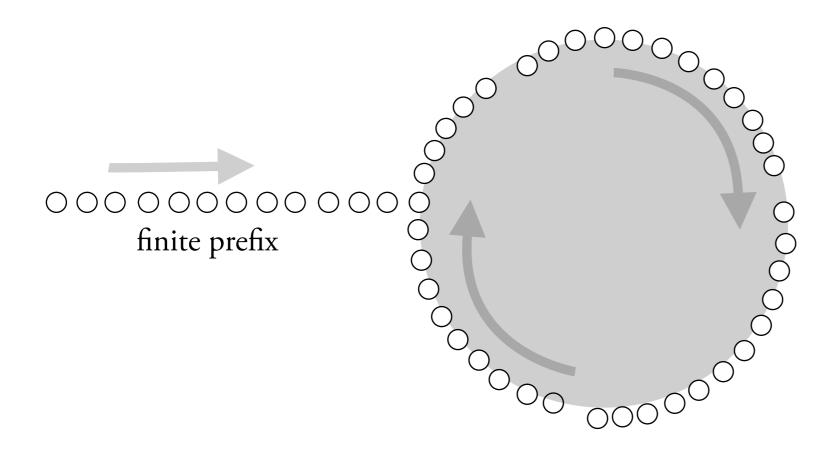
Transitions can do counter operations: c:=c+1 c:=0 c:=max(c,d)

Has finite state space Q and a finite set of counters C. The counters are only read by the acceptance condition, and not during the run.

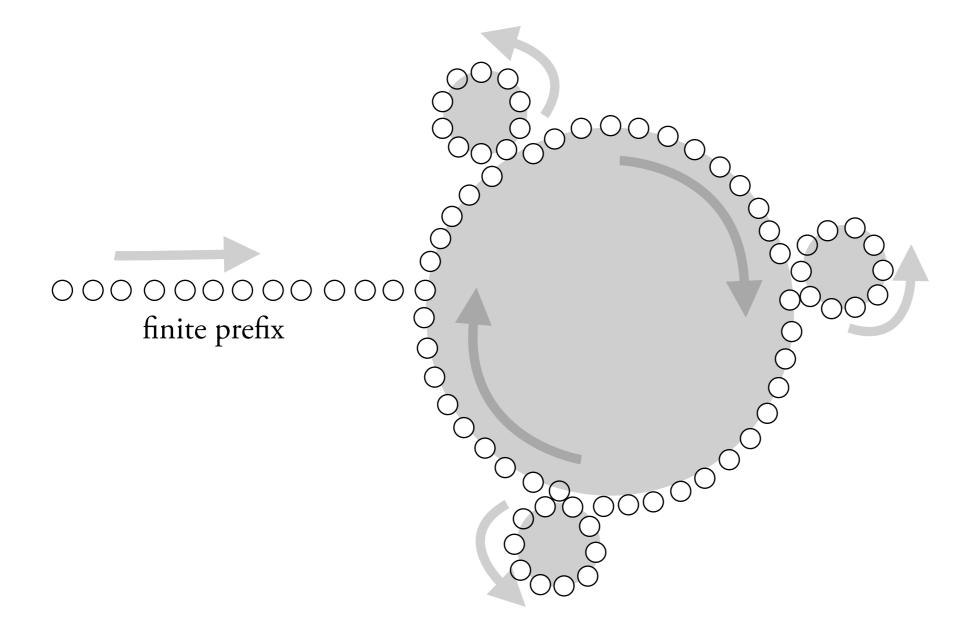
Acceptance condition: boolean combination of clauses "counter *c* is bounded"



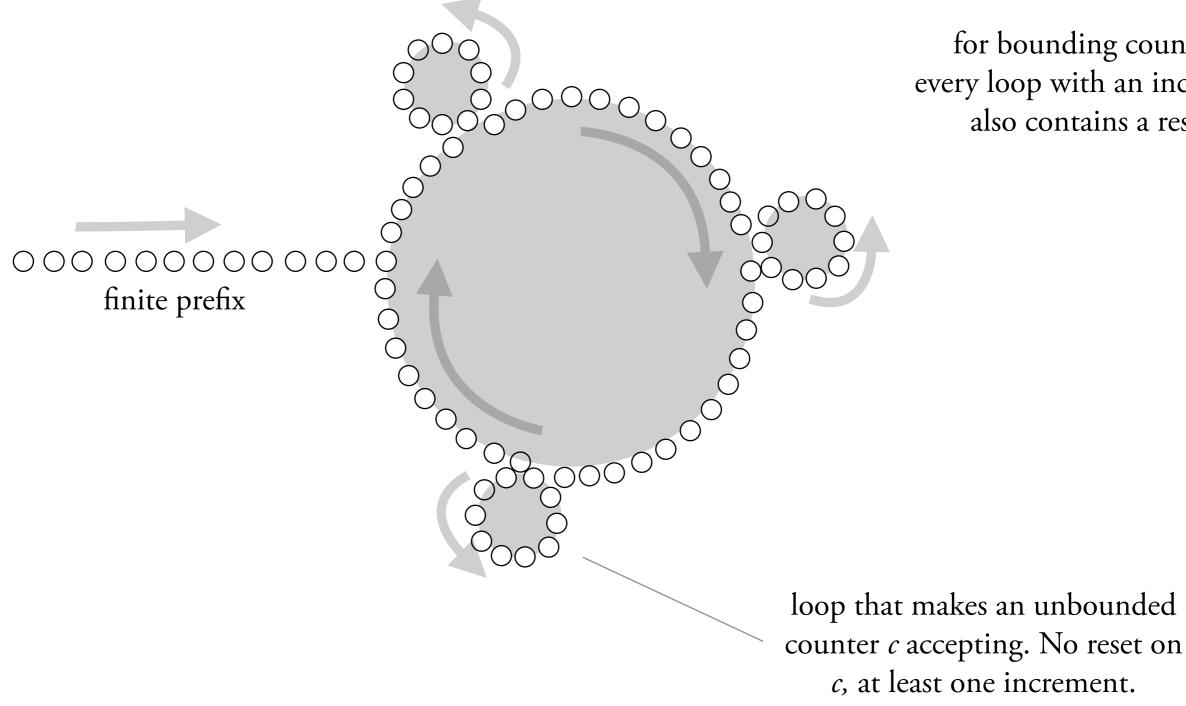
For a max-automaton, the accepting condition says some counters are bounded, and some are not.



For a max-automaton, the accepting condition says some counters are bounded, and some are not.



For a max-automaton, the accepting condition says some counters are bounded, and some are not.



for bounding counters: every loop with an increment also contains a reset.

What is the logic for max-automata?

UX $\varphi(X)$

which is the same as

" $\varphi(X)$ holds for finite sets X of arbitrarily large size"

which is the same as

 $\bigwedge_{n} \varphi(X) \wedge n < |X| < \infty$

UX $\varphi(X)$

which is the same as

" $\varphi(X)$ holds for finite sets X of arbitrarily large size"

which is the same as

 $\bigwedge_{n} \varphi(X) \wedge n < |X| < \infty$

Example: { $a^{n_1}b a^{n_2}b a^{n_3}b... : n_1 n_2 n_3...$ is not bounded}

UX $\varphi(X)$

which is the same as

" $\varphi(X)$ holds for finite sets X of arbitrarily large size"

which is the same as

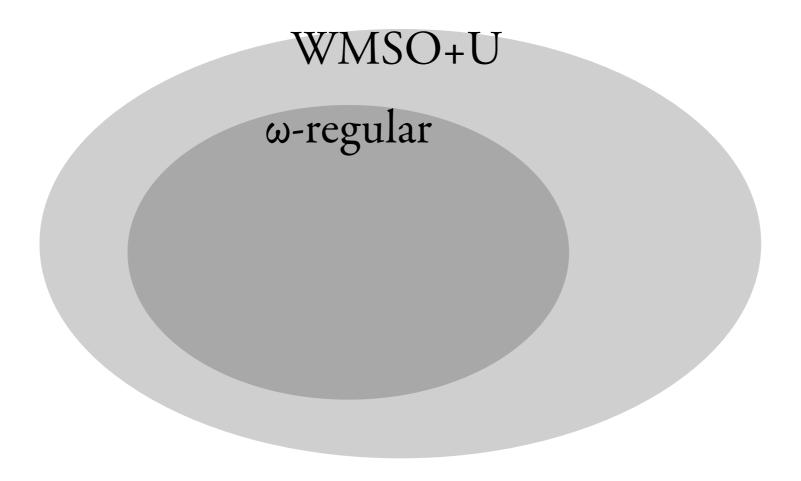
 $\bigwedge_{n} \varphi(X) \wedge n < |X| < \infty$

Example: { $a^{n_1}b a^{n_2}b a^{n_3}b... : n_1 n_2 n_3...$ is not bounded} UX "X is a set of consecutive *a*'s" **Thm.** Deterministic max-automata recognize the same langauges as weak MSO with the unbounding quantifier.

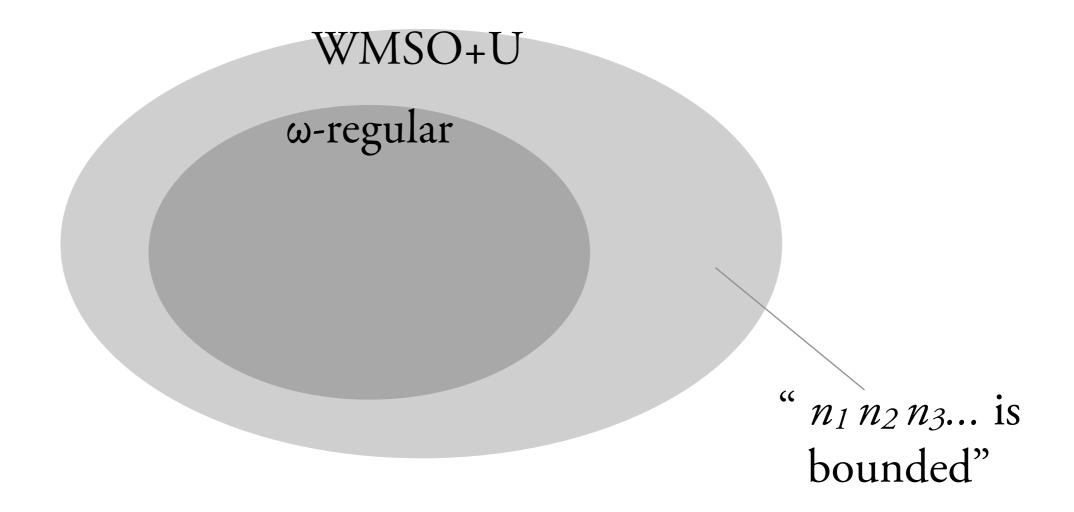
Thm. Deterministic max-automata recognize the same langauges as weak MSO with the unbounding quantifier.

Proof. Effective translations both ways.

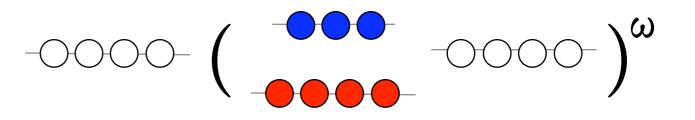
Thm. Deterministic max-automata recognize the same langauges as weak MSO with the unbounding quantifier. **Proof.** Effective translations both ways.

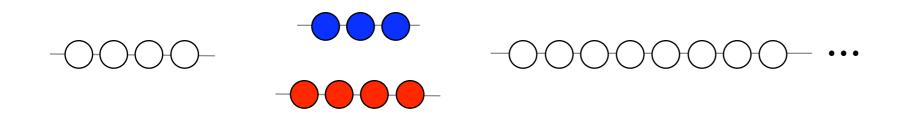


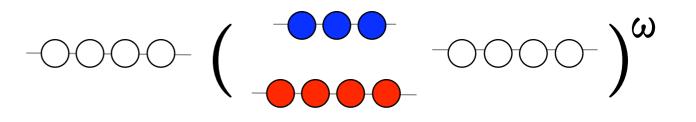
Thm. Deterministic max-automata recognize the same langauges as weak MSO with the unbounding quantifier. **Proof.** Effective translations both ways.

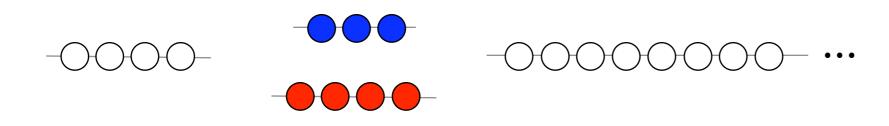


-logic
-automata
-decidability
_?

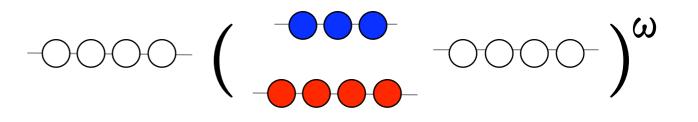


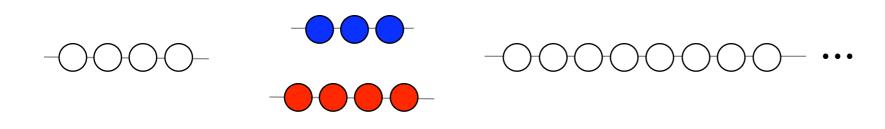






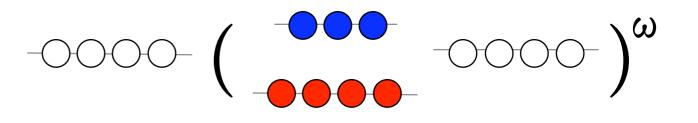
Prop. Languages recognized by max-automata have finitely many equivalence classes. Each class is a regular language of finite words.

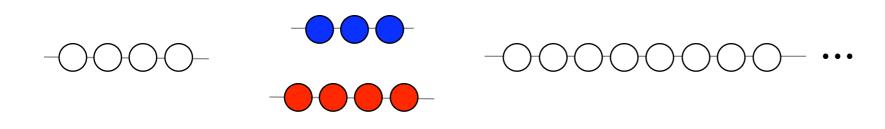




Prop. Languages recognized by max-automata have finitely many equivalence classes. Each class is a regular language of finite words.

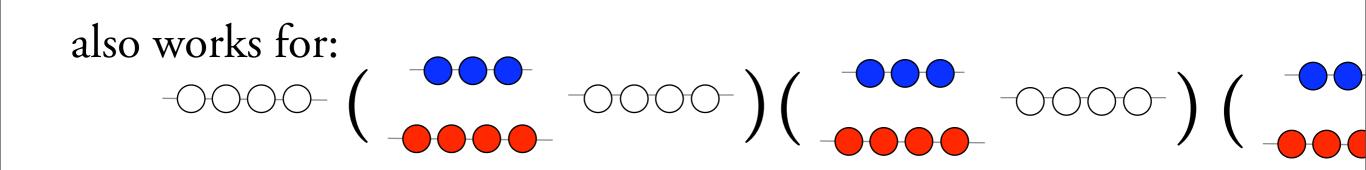
Proof sketch. Equivalence class of — depends on state transformations, which counters are incremented (but not how much), and which counters are reset.





Prop. Languages recognized by max-automata have finitely many equivalence classes. Each class is a regular language of finite words.

Proof sketch. Equivalence class of — depends on state transformations, which counters are incremented (but not how much), and which counters are reset.



Thm. MSO+U is strictly more expressive than WMSO+U

Thm. MSO+U is strictly more expressive than WMSO+U separating language $L = \{a^{n_1}b \ a^{n_2}b \ a^{n_3}b... : n_1 n_2 n_3... \text{ tends to } \infty\}$

Thm. MSO+U is strictly more expressive than WMSO+U separating language $L = \{a^{n_1}b \ a^{n_2}b \ a^{n_3}b... : n_1 n_2 n_3... \text{ tends to } \infty\}$

 $L \in \mathrm{MSO}{+}\mathrm{U}$

Thm. MSO+U is strictly more expressive than WMSO+U separating language $L = \{a^{n_1}b \ a^{n_2}b \ a^{n_3}b... : n_1 n_2 n_3... \text{ tends to } \infty\}$

 $L \in MSO+U$ complement of L: exists a bounded subsequence.

Thm. MSO+U is strictly more expressive than WMSO+U separating language $L = \{a^{n_1}b \ a^{n_2}b \ a^{n_3}b... : n_1 n_2 n_3... \text{ tends to } \infty\}$

 $L \in MSO+U$ complement of L: exists a bounded subsequence.

 $L \not\in \text{WMSO+U}$

Thm. MSO+U is strictly more expressive than WMSO+U separating language $L = \{a^{n_1}b \ a^{n_2}b \ a^{n_3}b... : n_1 n_2 n_3... \text{ tends to } \infty\}$

 $L \in MSO+U$ complement of L: exists a bounded subsequence.

 $L \notin WMSO+U$ topological argument.

Thm. MSO+U is strictly more expressive than WMSO+U separating language $L = \{a^{n_1}b \ a^{n_2}b \ a^{n_3}b... : n_1 n_2 n_3... \text{ tends to } \infty\}$

 $L \in MSO+U$ complement of L: exists a bounded subsequence.

 $L \notin WMSO+U$ topological argument. acceptance condition " $n_1 n_2 n_3$... is bounded"

Thm. MSO+U is strictly more expressive than WMSO+U separating language $L = \{a^{n_1}b \ a^{n_2}b \ a^{n_3}b... : n_1 n_2 n_3... \text{ tends to } \infty\}$

 $L \in MSO+U$ complement of L: exists a bounded subsequence.

 $L \notin WMSO+U$ topological argument. acceptance condition " $n_1 n_2 n_3$... is bounded" is a countable union of closed sets (Σ_2)

Thm. MSO+U is strictly more expressive than WMSO+U separating language $L = \{a^{n_1}b \ a^{n_2}b \ a^{n_3}b... : n_1 n_2 n_3... \text{ tends to } \infty\}$

 $L \in MSO+U$ complement of L: exists a bounded subsequence.

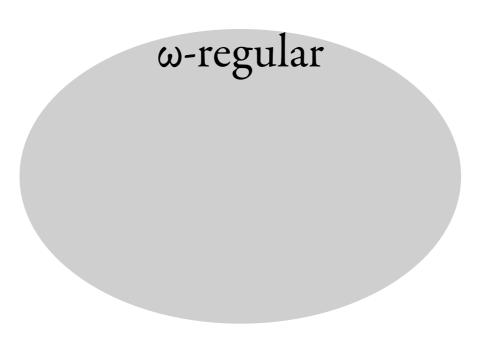
 $L \notin WMSO+U$ topological argument. acceptance condition " $n_1 n_2 n_3$... is bounded" is a countable union of closed sets (Σ_2) "sequence bounded by N" is a closed set

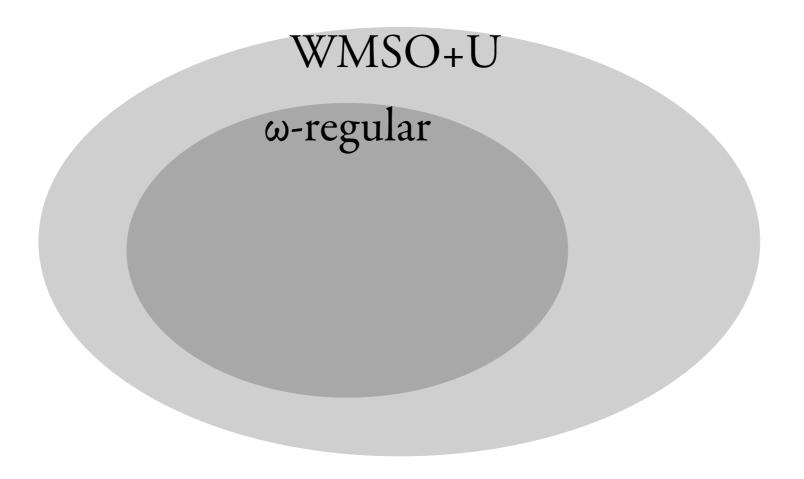
Thm. MSO+U is strictly more expressive than WMSO+U separating language $L = \{a^{n_1}b \ a^{n_2}b \ a^{n_3}b... : n_1 n_2 n_3... \text{ tends to } \infty\}$

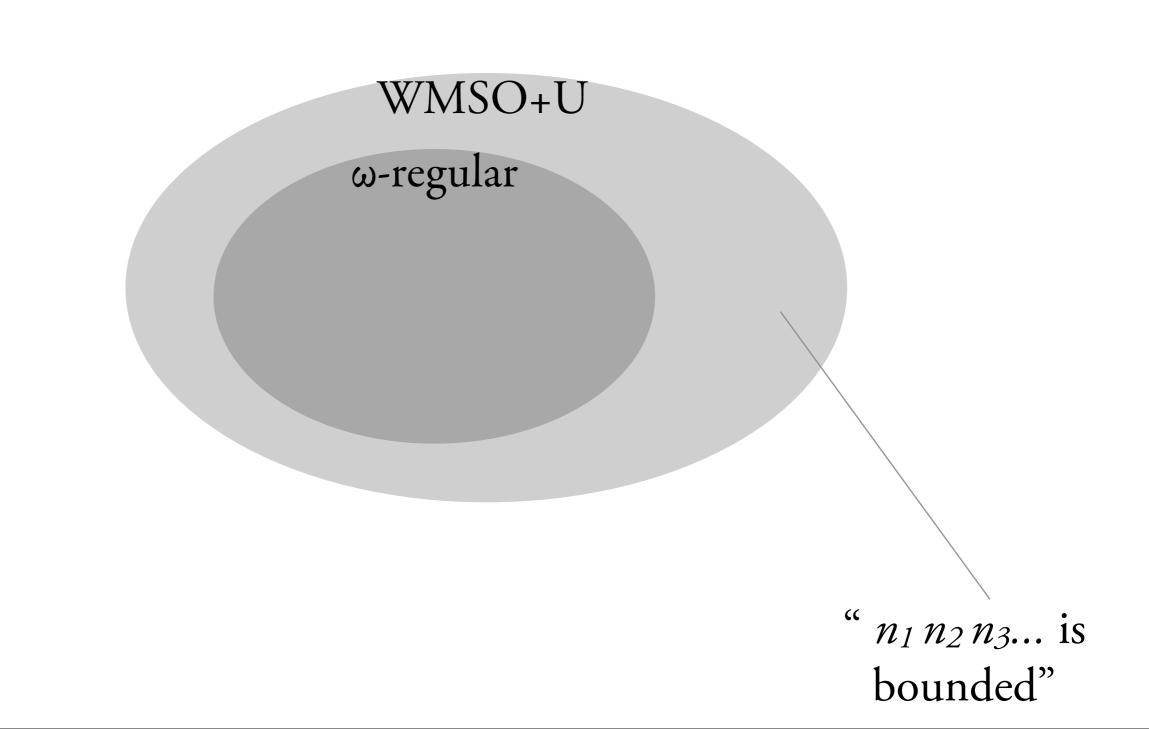
 $L \in MSO+U$ complement of L: exists a bounded subsequence.

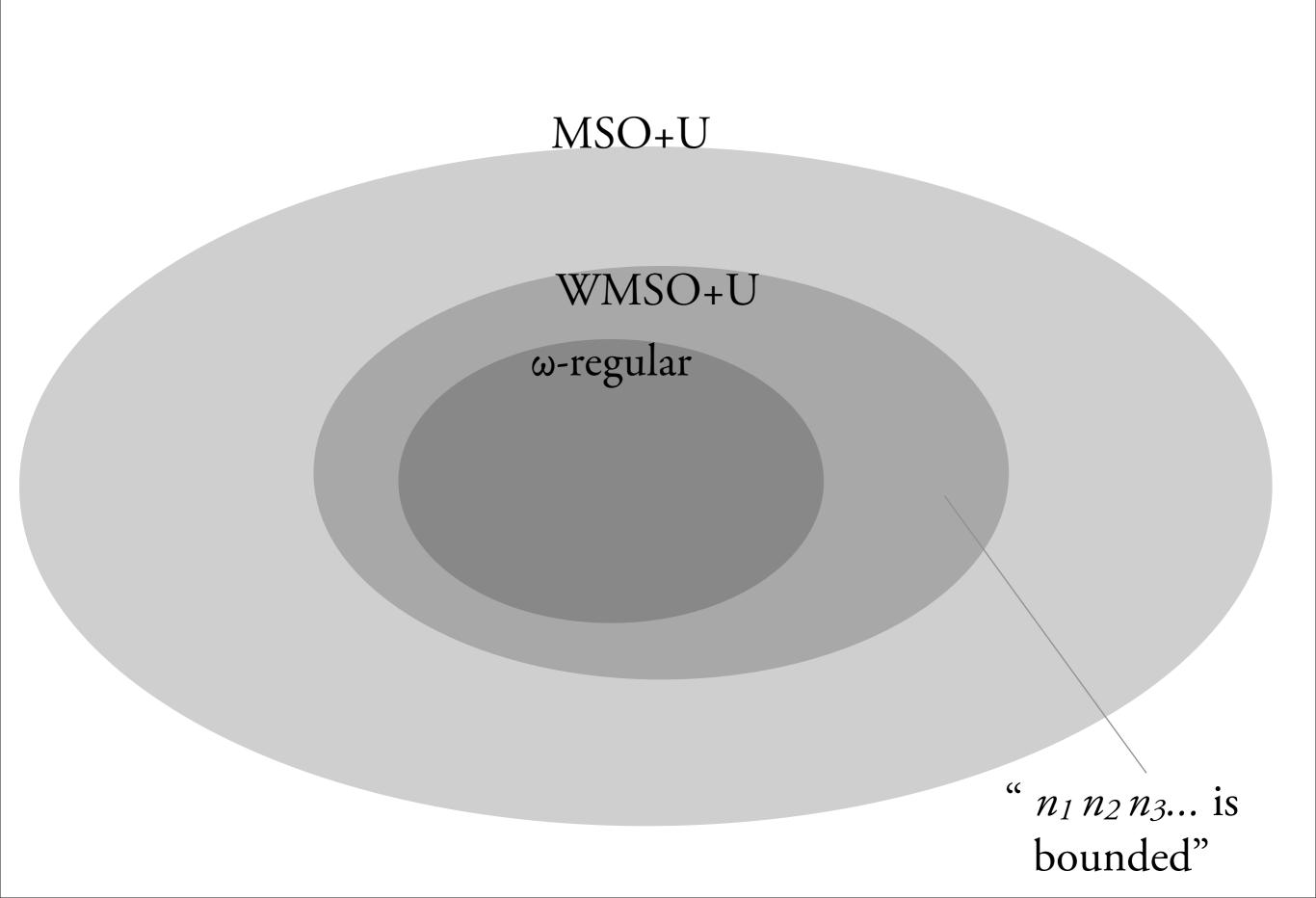
 $L \notin WMSO+U$ topological argument. acceptance condition " $n_1 n_2 n_3$... is bounded" is a countable union of closed sets (Σ_2) "sequence bounded by N" is a closed set

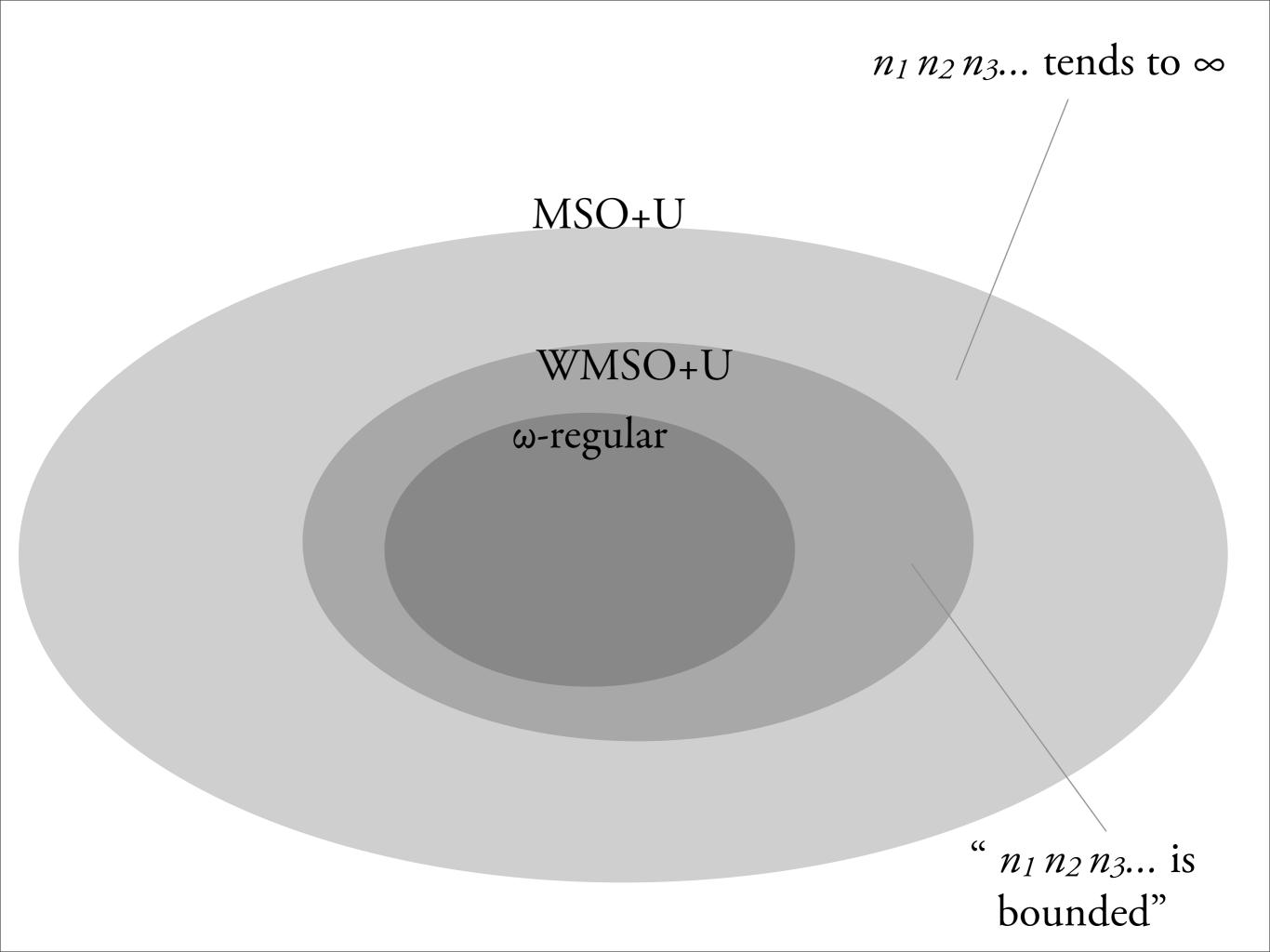
Prop. A language recognized by a max automaton is a boolean combination of Σ_2 sets, while *L* is not.

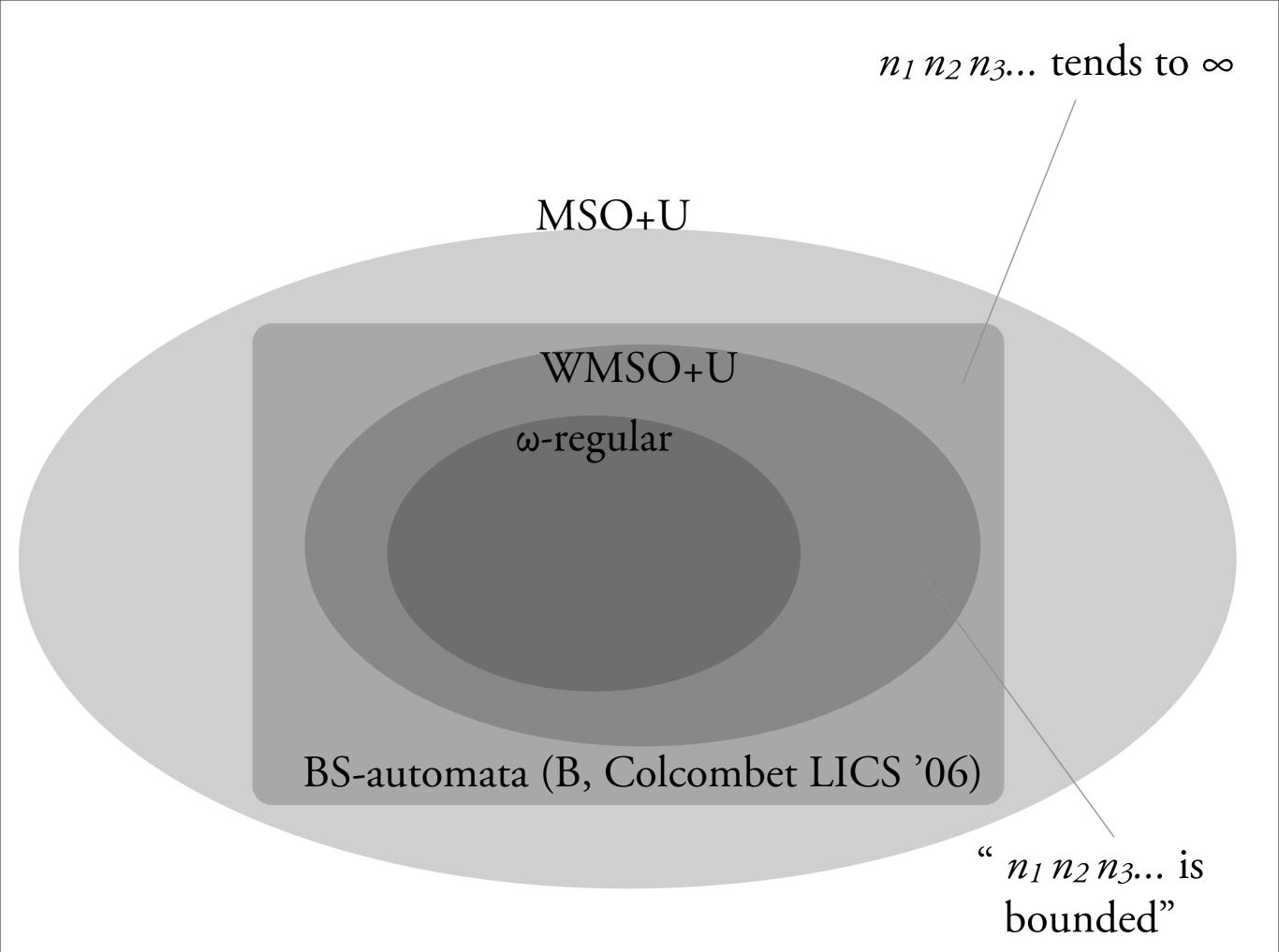


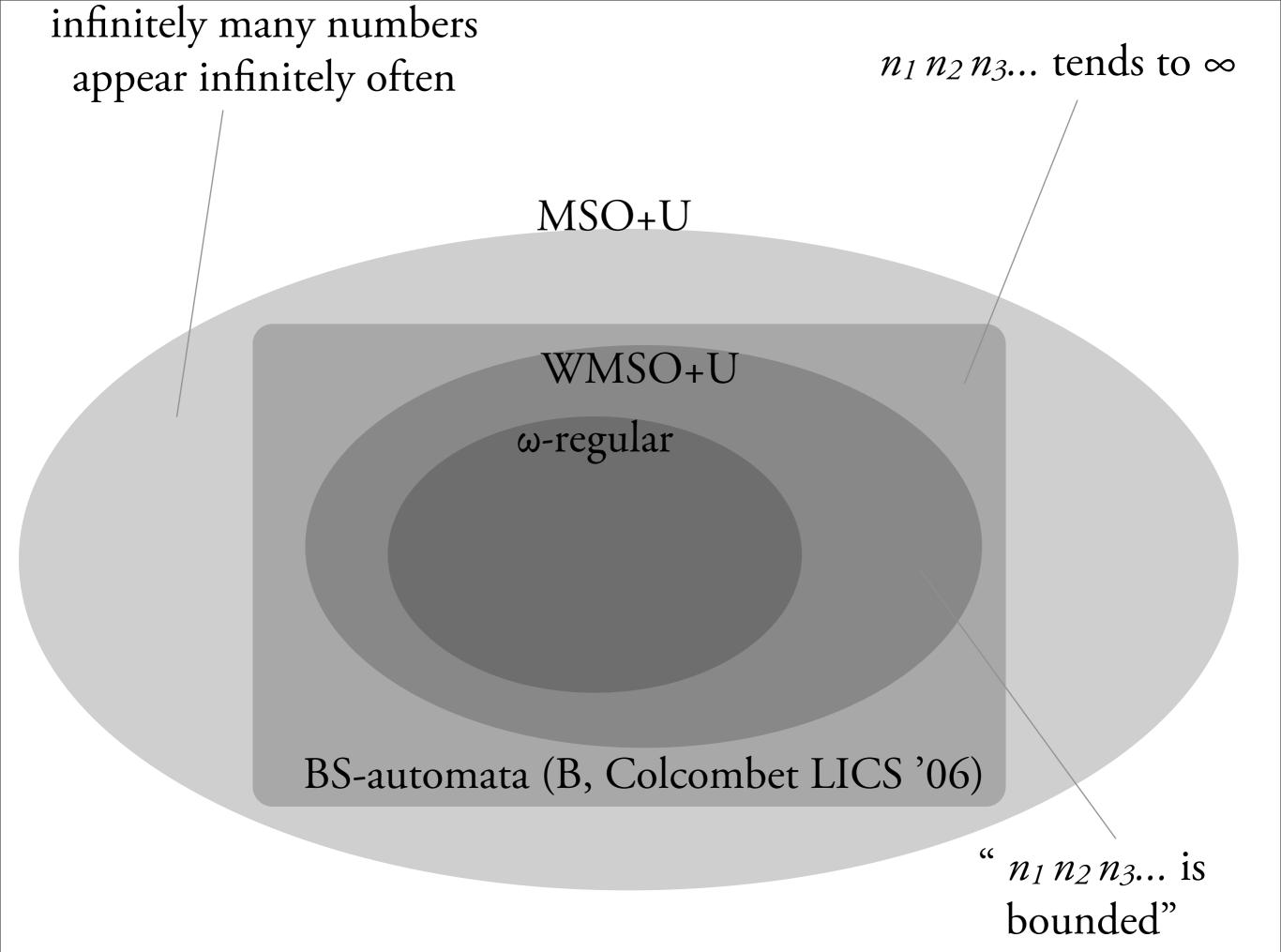












New robust class of languages extending ω -regular languages. (automata, logic, decidability)

New robust class of languages extending ω -regular languages. (automata, logic, decidability)

New robust class of languages extending ω -regular languages. (automata, logic, decidability)

Future work

– Full MSO+U

New robust class of languages extending ω -regular languages. (automata, logic, decidability)

- Full MSO+U
- Tree extensions

New robust class of languages extending ω -regular languages. (automata, logic, decidability)

- Full MSO+U
- Tree extensions
- Algebra

New robust class of languages extending ω -regular languages. (automata, logic, decidability)

- Full MSO+U
- Tree extensions
- Algebra
- Regular expressions

a bit about the proofs





Proof strategy: Automata are closed under all operations in the logic.



Proof strategy: Automata are closed under all operations in the logic.

Boolean operations: free for a deterministic automaton.



Proof strategy: Automata are closed under all operations in the logic.

Boolean operations: free for a deterministic automaton.

Weak existential quantification

Unbounding quantification



Proof strategy: Automata are closed under all operations in the logic.

Boolean operations: free for a deterministic automaton.

Let w be a word over alphabet Σ , and X a set of positions. w[X]: word over alphabet $\Sigma \times \{0,1\}$

Weak existential quantification

Unbounding quantification

Proof strategy: Automata are closed under all operations in the logic. Boolean operations: free for a deterministic automaton.

Let w be a word over alphabet Σ , and X a set of positions. w[X]: word over alphabet $\Sigma \times \{0,1\}$

Weak existential quantification **Prop.** If $L \subseteq (\Sigma \times \{0,1\})^{\omega}$ is recognized by a deterministic maxautomaton, then so is $\{w : w[X] \in L$ for some finite set $X\} \subseteq \Sigma^{\omega}$

Unbounding quantification

Proof strategy: Automata are closed under all operations in the logic. Boolean operations: free for a deterministic automaton.

Let w be a word over alphabet Σ , and X a set of positions. w[X]: word over alphabet $\Sigma \times \{0,1\}$

Weak existential quantification **Prop.** If $L \subseteq (\Sigma \times \{0,1\})^{\omega}$ is recognized by a deterministic maxautomaton, then so is $\{w : w[X] \in L$ for some finite set $X\} \subseteq \Sigma^{\omega}$

Unbounding quantification **Prop.** If $L \subseteq (\Sigma \times \{0,1\})^{\omega}$ is recognized by a deterministic maxautomaton, then so is $\{w : w[X] \in L$ for arbitrarily large $X\} \subseteq \Sigma^{\omega}$ **Proof strategy:** Automata are closed under all operations in the logic. Boolean operations: free for a deterministic automaton.

Let w be a word over alphabet Σ , and X a set of positions. w[X]: word over alphabet $\Sigma \times \{0,1\}$

Weak existential quantification **Prop.** If $L \subseteq (\Sigma \times \{0,1\})^{\omega}$ is recognized by a deterministic maxautomaton, then so is $\{w : w[X] \in L$ for some finite set $X\} \subseteq \Sigma^{\omega}$

Unbounding quantification

Prop. If $L \subseteq (\Sigma \times \{0,1\})^{\omega}$ is recognized by a deterministic maxautomaton, then so is $\{w : w[X] \in L$ for arbitrarily large $X\} \subseteq \Sigma^{\omega}$ The proof uses a combinatoric theorem of I. Simon.

Rule 1. split into two parts

abaabbbababbabba bbabbbabbabbabbababa

Rule 1. split into two parts

abaabbbababbabba bbabbbabbabbabbababa

Rule 2. split into many parts, each with the same transformation $\delta: Q \to Q$

abaab	bbababb	babba	bba	bbbabb	babba	ba	
δ							

Rule 1. split into two parts

abaabbbababbabba bbabbbabbabbabbababa

Rule 2. split into many parts, each with the same transformation $\delta: Q \to Q$

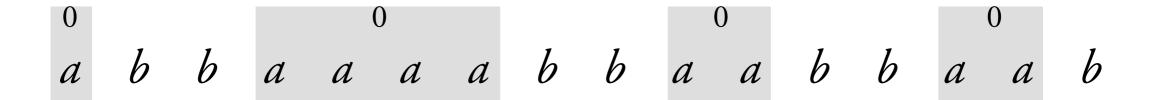
aba	ab	bbababb	babba	bba	bbbabb	babba	ba				
δ		δ	δ	δ	δ	δ	δ				
$\delta\circ\delta=\delta$											

two transition functions: even (0) and odd (1)

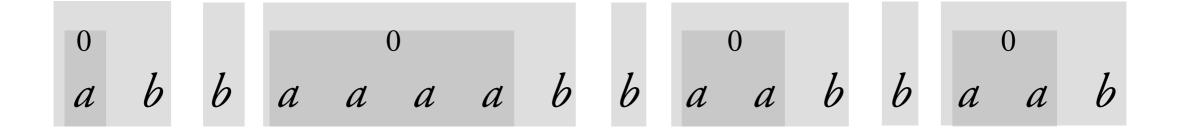
two transition functions: even (0) and odd (1)

a b b a a a a b b a a b b a a b

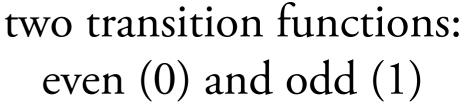
two transition functions: even (0) and odd (1)

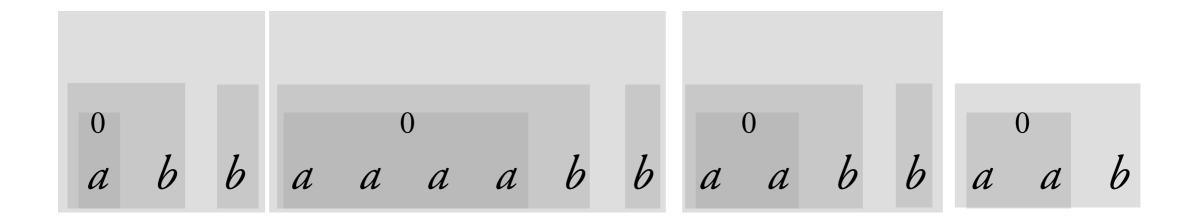


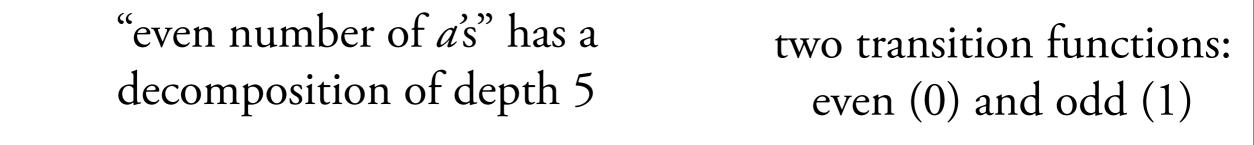
two transition functions: even (0) and odd (1)

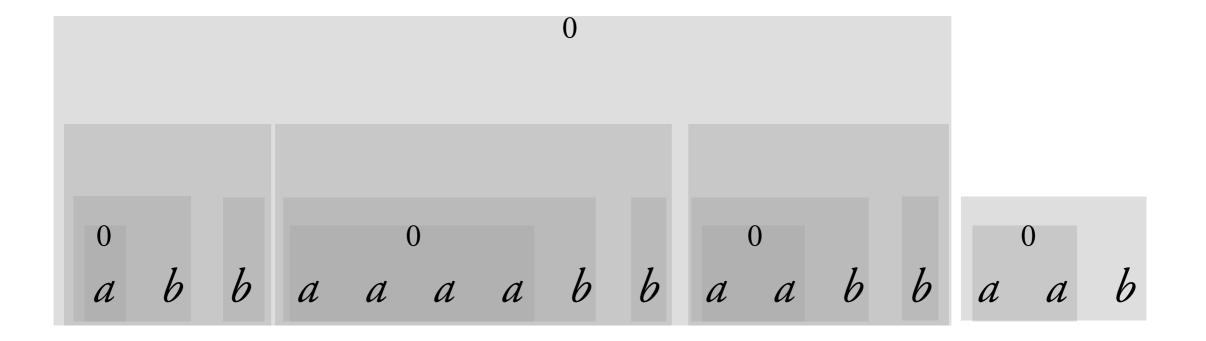


"even number of a's" has a two transf decomposition of depth 5 even (0)

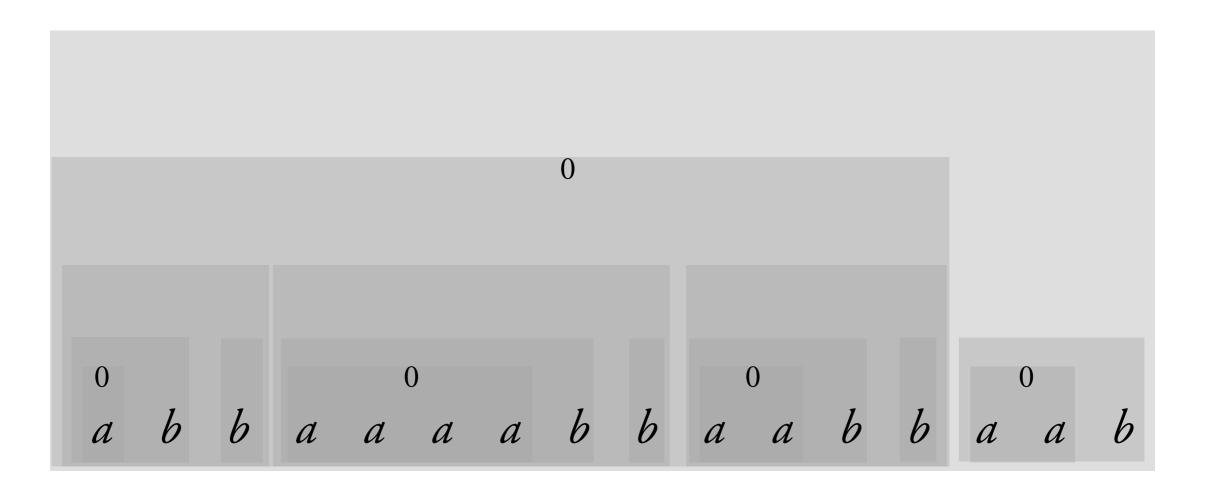




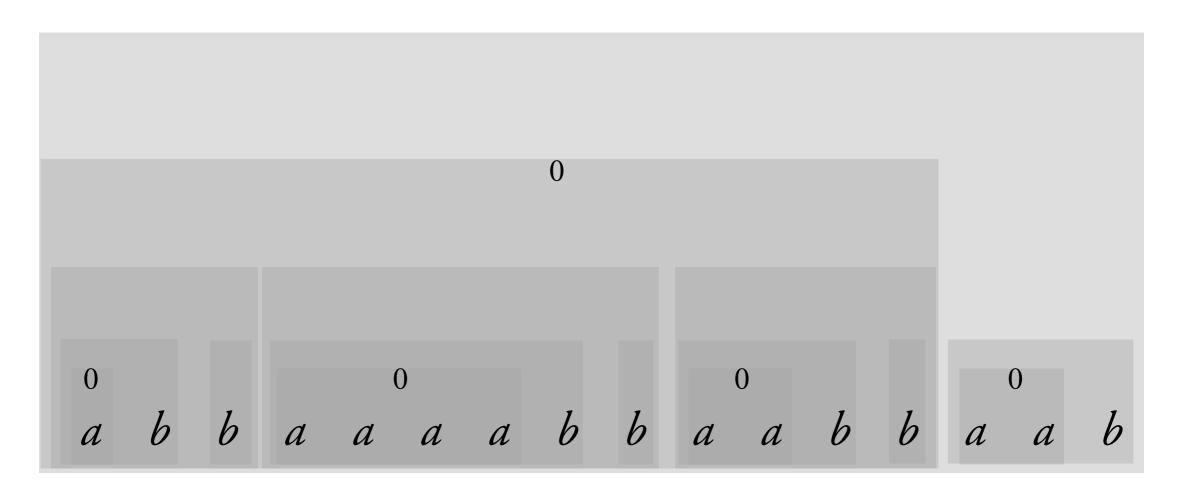




"even number of a's" has a two transition functions: decomposition of depth 5 even (0) and odd (1)







Thm. (Colcombet '07)

The decomposition can be output by a deterministic finite state transducer.

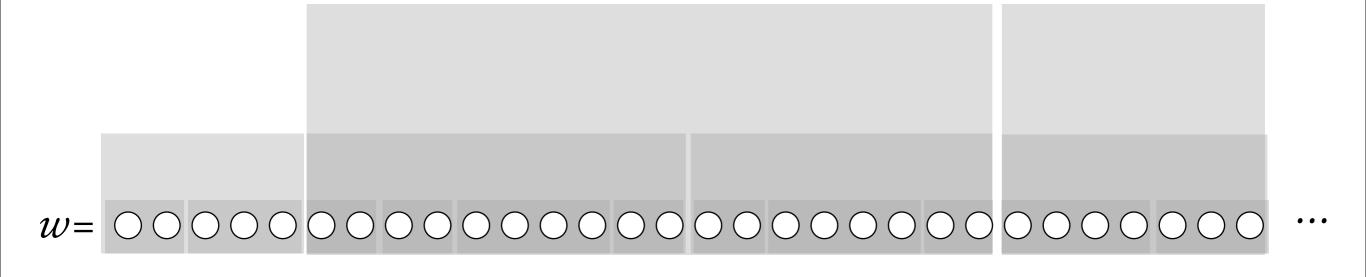
given a word $w \in \Sigma^{\omega}$, how can a deterministic automaton tell if $w[X] \in L$ holds for arbitrarily large X?

given a word $w \in \Sigma^{\omega}$, how can a deterministic automaton tell if $w[X] \in L$ holds for arbitrarily large X?

given a word $w \in \Sigma^{\omega}$, how can a deterministic automaton tell if $w[X] \in L$ holds for arbitrarily large X?

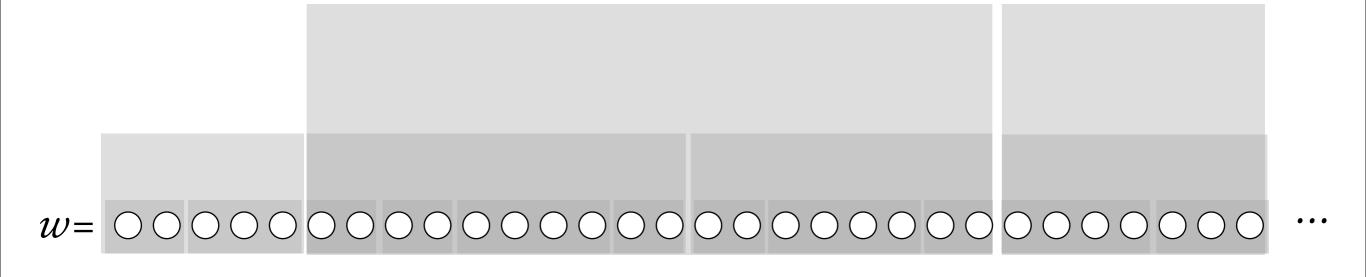
1. compute Simon decomposition for well chosen automaton (a modification of the automaton for *L*)

given a word $w \in \Sigma^{\omega}$, how can a deterministic automaton tell if $w[X] \in L$ holds for arbitrarily large X?



1. compute Simon decomposition for well chosen automaton (a modification of the automaton for *L*)

given a word $w \in \Sigma^{\omega}$, how can a deterministic automaton tell if $w[X] \in L$ holds for arbitrarily large X?



1. compute Simon decomposition for well chosen automaton (a modification of the automaton for L)

2. find large boxes