Developments in Language Theory

Palermo, Italy, July 8th, 2005

Collapsing Words: A Progress Report

D. S. Ananichev, I. V. Petrov and M. V. Volkov

Ural State University, Ekaterinburg, Russia

What Are Collapsing Words?

We work with DFA: $\mathscr{A} = \langle \boldsymbol{Q}, \boldsymbol{\Sigma}, \boldsymbol{\delta} \rangle$.

What Are Collapsing Words?

We work with DFA: $\mathscr{A} = \langle Q, \Sigma, \delta \rangle$.

The *deficiency* of a word $v \in \Sigma^*$ w.r.t. a DFA $\mathscr{A}$ is the difference $\mathrm{df}_{\mathscr{A}}(v) = |Q| - |Q\,.\,v|$. Here $Q\,.\,v$ stands for $\{\delta(q, v) \mid q \in Q\}$.

What Are Collapsing Words?

We work with DFA: $\mathscr{A} = \langle Q, \Sigma, \delta \rangle$.

The *deficiency* of a word $v \in \Sigma^*$ w.r.t. a DFA $\mathscr{A}$ is the difference $\mathrm{df}_{\mathscr{A}}(v) = |Q| - |Q \,.\, v|$.

A DFA $\mathscr{A}$ is *n-compressible* if $\mathrm{df}_{\mathscr{A}}(v) \geq n$ for some input word $v \in \Sigma^*$.

What Are Collapsing Words?

We work with DFA: $\mathscr{A} = \langle Q, \Sigma, \delta \rangle$.

The *deficiency* of a word $v \in \Sigma^*$ w.r.t. a DFA $\mathscr{A}$ is the difference $\mathrm{df}_{\mathscr{A}}(v) = |Q| - |Q \,.\, v|$.

A DFA $\mathscr{A}$ is *n-compressible* if $\mathrm{df}_{\mathscr{A}}(v) \geq n$ for some input word $v \in \Sigma^*$.

A word $w \in \Sigma^*$ is *n-collapsing* if $\mathrm{df}_{\mathscr{A}}(w) \geq n$ for **every** n-compressible DFA $\mathscr{A}$ whose input alphabet is Σ.

What Are Collapsing Words?

We work with DFA: $\mathscr{A} = \langle Q, \Sigma, \delta \rangle$.

The *deficiency* of a word $v \in \Sigma^*$ w.r.t. a DFA $\mathscr{A}$ is the difference $\mathrm{df}_{\mathscr{A}}(v) = |Q| - |Q \,.\, v|$.

A DFA $\mathscr{A}$ is *n-compressible* if $\mathrm{df}_{\mathscr{A}}(v) \geq n$ for some input word $v \in \Sigma^*$.

A word $w \in \Sigma^*$ is *n-collapsing* if $\mathrm{df}_{\mathscr{A}}(w) \geq n$ for **every** n-compressible DFA $\mathscr{A}$ whose input alphabet is Σ.

Thus, infinitely many n-compressible automata should all be ‘served’ by the same word!

What Are Collapsing Words?

We work with DFA: $\mathscr{A} = \langle Q, \Sigma, \delta \rangle$.

The *deficiency* of a word $v \in \Sigma^*$ w.r.t. a DFA $\mathscr{A}$ is the difference $\mathrm{df}_{\mathscr{A}}(v) = |Q| - |Q\,.\,v|$.

A DFA $\mathscr{A}$ is n*-compressible* if $\mathrm{df}_{\mathscr{A}}(v) \geq n$ for some input word $v \in \Sigma^*$.

A word $w \in \Sigma^*$ is n*-collapsing* if $\mathrm{df}_{\mathscr{A}}(w) \geq n$ for **every** n-compressible DFA $\mathscr{A}$ whose input alphabet is Σ.

Thus, infinitely many n-compressible automata should all be ‘served’ by the same word! Say, a 3-collapsing word:

$4 \longrightarrow 1$;

What Are Collapsing Words?

We work with DFA: $\mathscr{A} = \langle \boldsymbol{Q}, \boldsymbol{\Sigma}, \boldsymbol{\delta} \rangle$.

The *deficiency* of a word $\boldsymbol{v} \in \boldsymbol{\Sigma}^*$ w.r.t. a DFA $\mathscr{A}$ is the difference $\mathbf{df}_{\mathscr{A}}(\boldsymbol{v}) = |\boldsymbol{Q}| - |\boldsymbol{Q} \,.\, \boldsymbol{v}|$.

A DFA $\mathscr{A}$ is $\boldsymbol{n}$*-compressible* if $\mathbf{df}_{\mathscr{A}}(\boldsymbol{v}) \geq \boldsymbol{n}$ for some input word $\boldsymbol{v} \in \boldsymbol{\Sigma}^*$.

A word $\boldsymbol{w} \in \boldsymbol{\Sigma}^*$ is $\boldsymbol{n}$*-collapsing* if $\mathbf{df}_{\mathscr{A}}(\boldsymbol{w}) \geq \boldsymbol{n}$ for **every** $\boldsymbol{n}$-compressible DFA $\mathscr{A}$ whose input alphabet is $\boldsymbol{\Sigma}$.

Thus, infinitely many $\boldsymbol{n}$-compressible automata should all be 'served' by the same word! Say, a 3-collapsing word: $\mathbf{4} \longrightarrow \mathbf{1}$; at the same time, $\mathbf{1000000} \longrightarrow \mathbf{999997}$, etc.

What Are Collapsing Words?

A natural interpretation as a solitaire-like game:

What Are Collapsing Words?

A natural interpretation as a solitaire-like game:

- The underlying digraph $\Gamma(\mathscr{A})$ — the game-board.

What Are Collapsing Words?

A natural interpretation as a solitaire-like game:

- The underlying digraph $\Gamma(\mathscr{A})$ — the game-board.
- The initial position — each state holds a token.

What Are Collapsing Words?

A natural interpretation as a solitaire-like game:

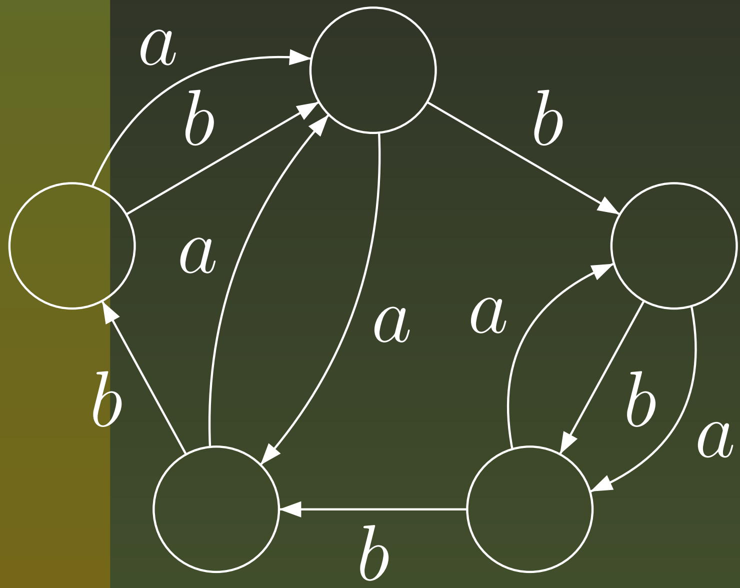
- The underlying digraph $\Gamma(\mathscr{A})$ — the game-board.
- The initial position — each state holds a token.
- Each letter $a \in \Sigma$ defines a move — tokens slide along the arrows labelled a and, whenever several tokens arrive at the same state, all but one of them are removed.

What Are Collapsing Words?

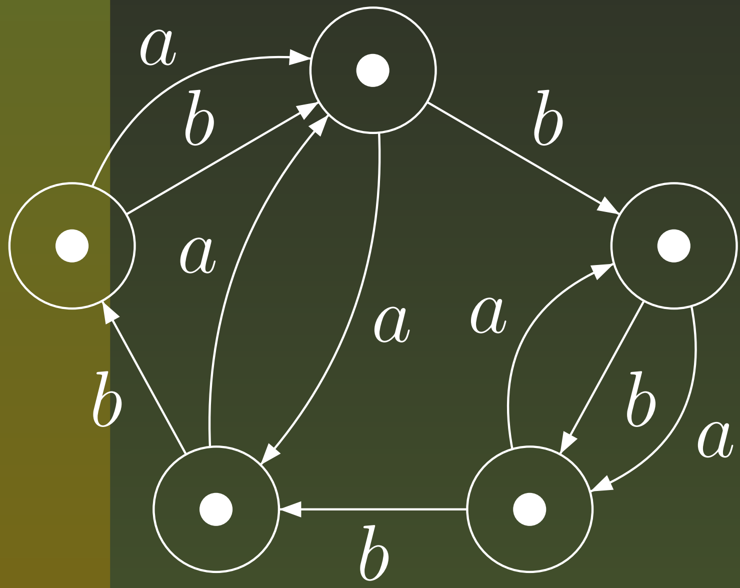
A natural interpretation as a solitaire-like game:

- The underlying digraph $\Gamma(\mathscr{A})$ — the game-board.
- The initial position — each state holds a token.
- Each letter $a \in \Sigma$ defines a move — tokens slide along the arrows labelled a and, whenever several tokens arrive at the same state, all but one of them are removed.
- The goal — to free at least n states (n is a given number).

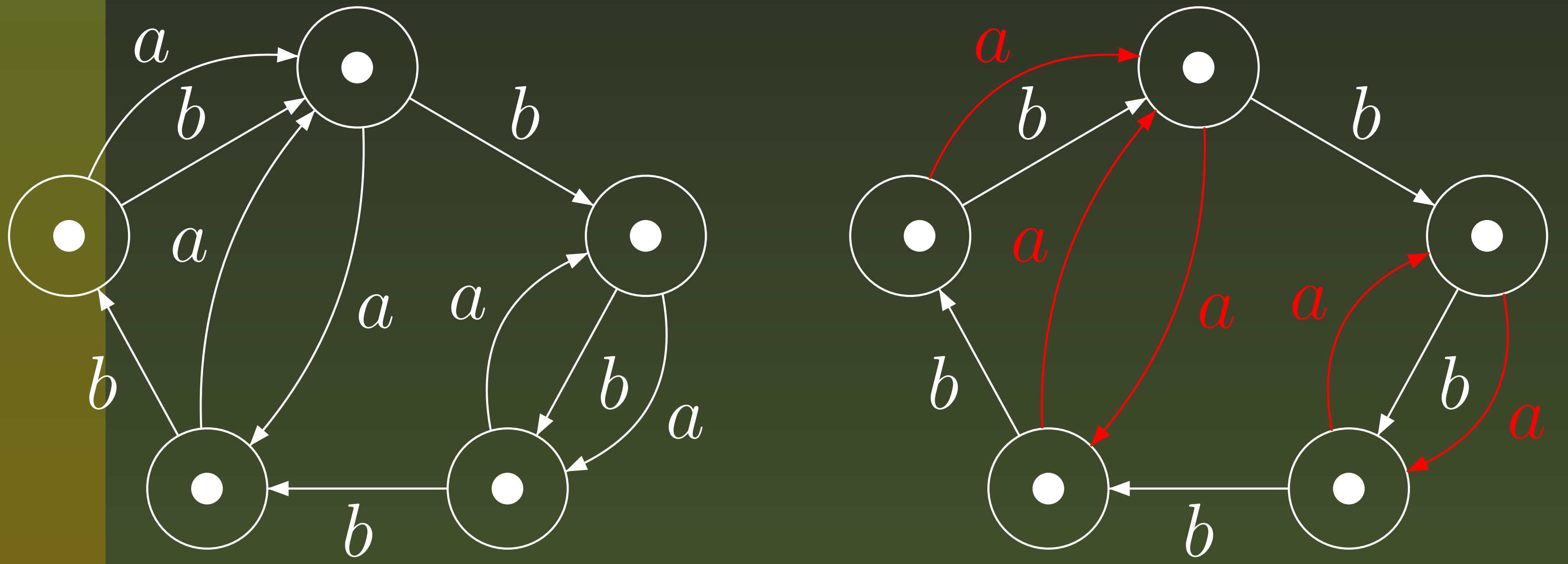
What Are Collapsing Words?



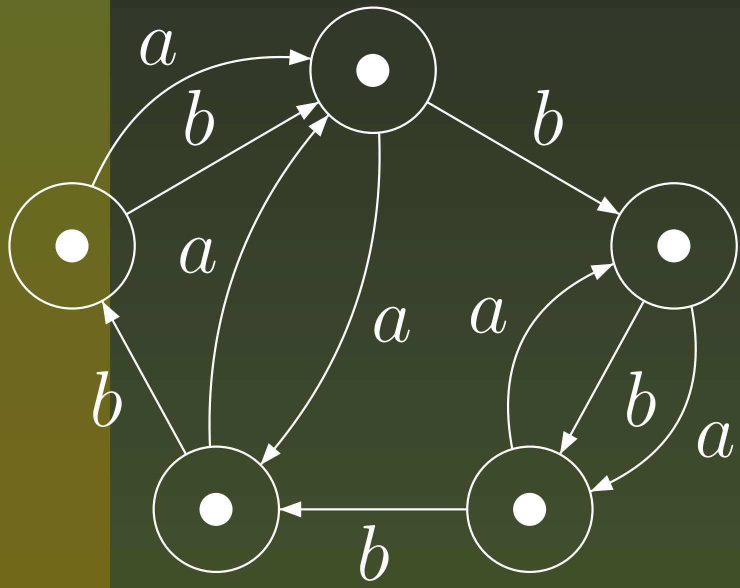
What Are Collapsing Words?

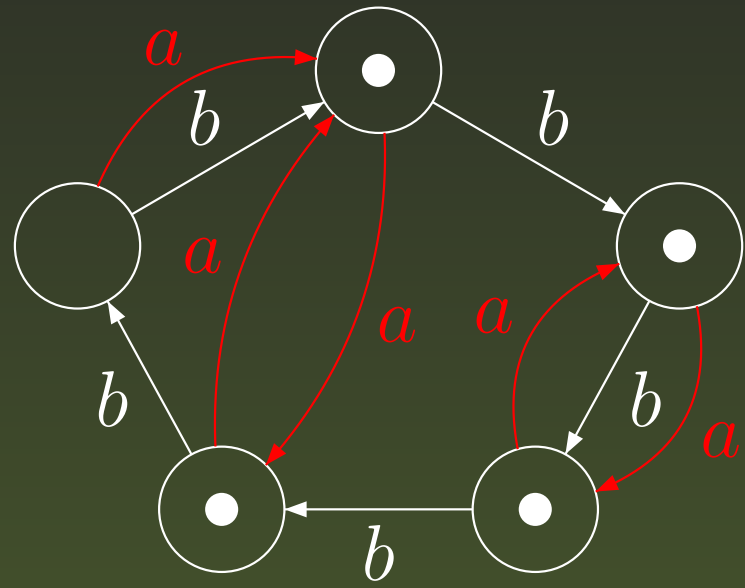


What Are Collapsing Words?

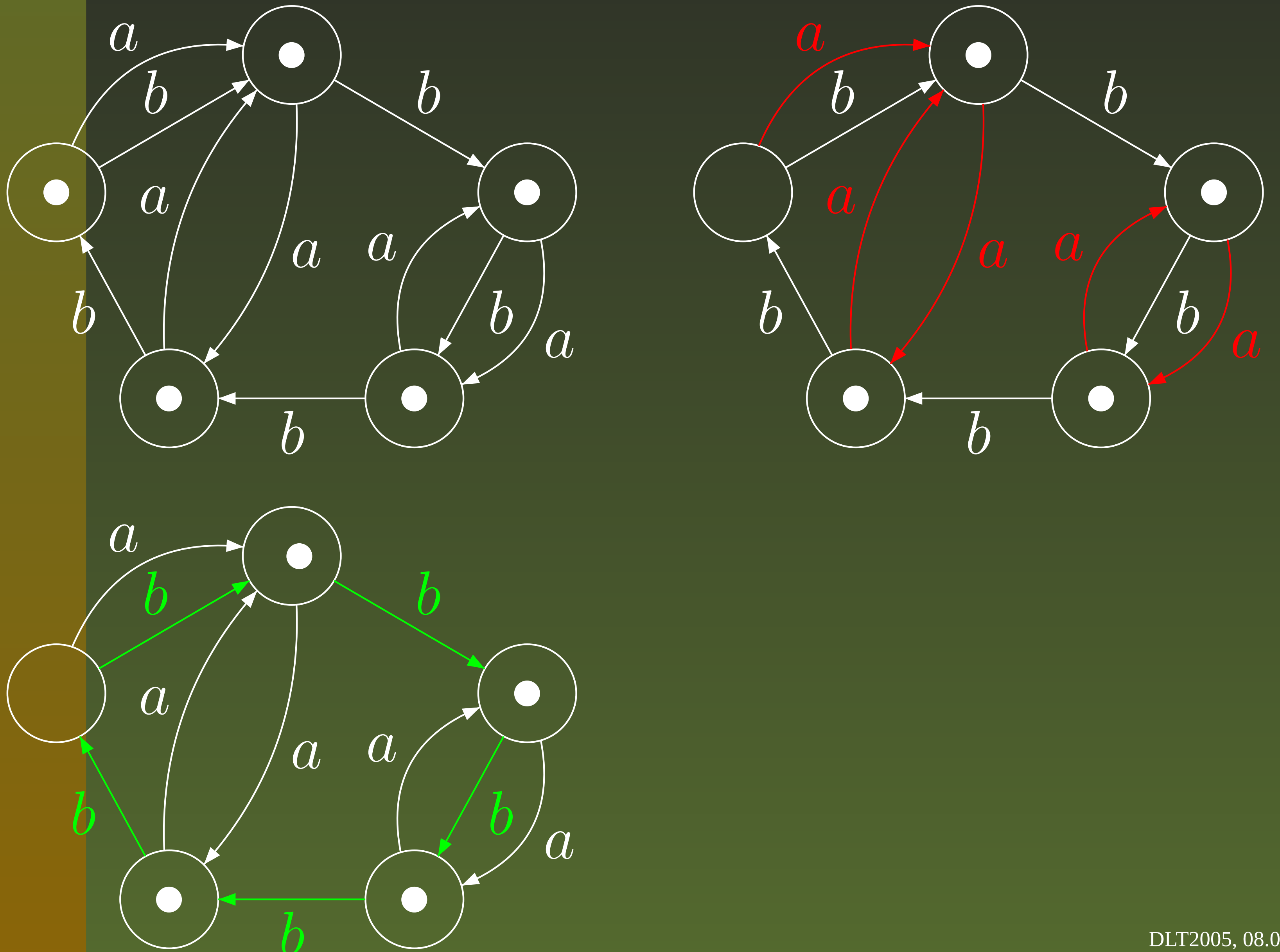


What Are Collapsing Words?

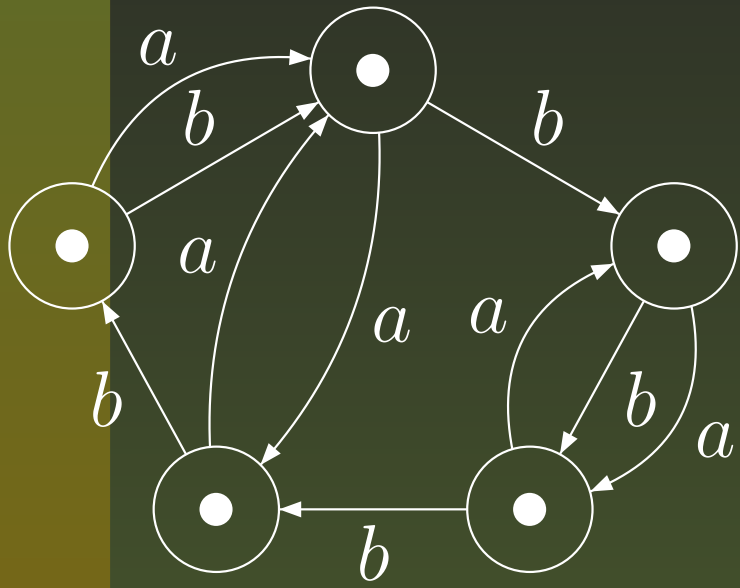


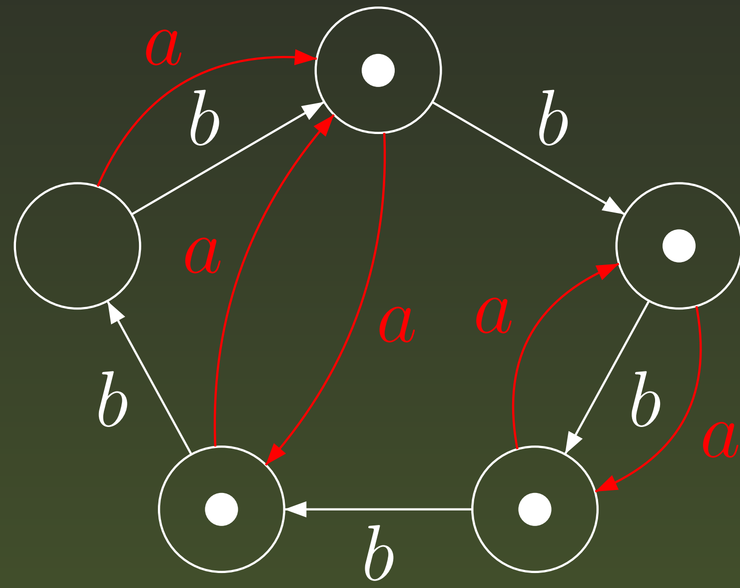


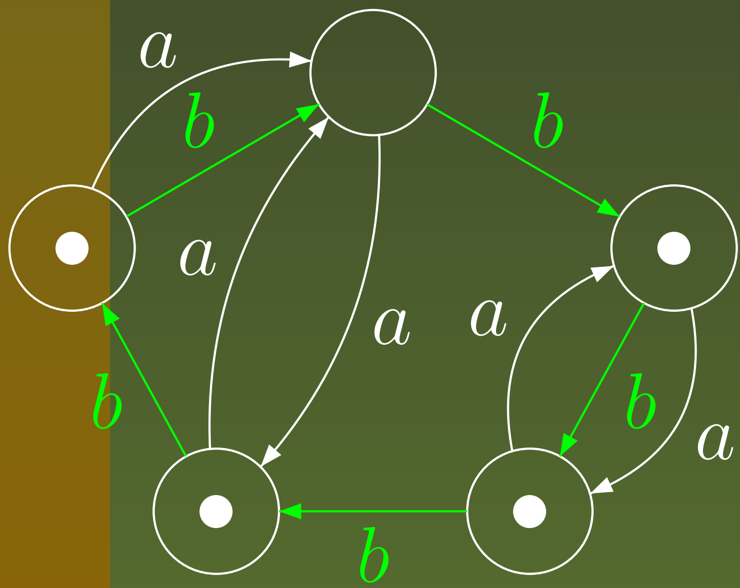
What Are Collapsing Words?



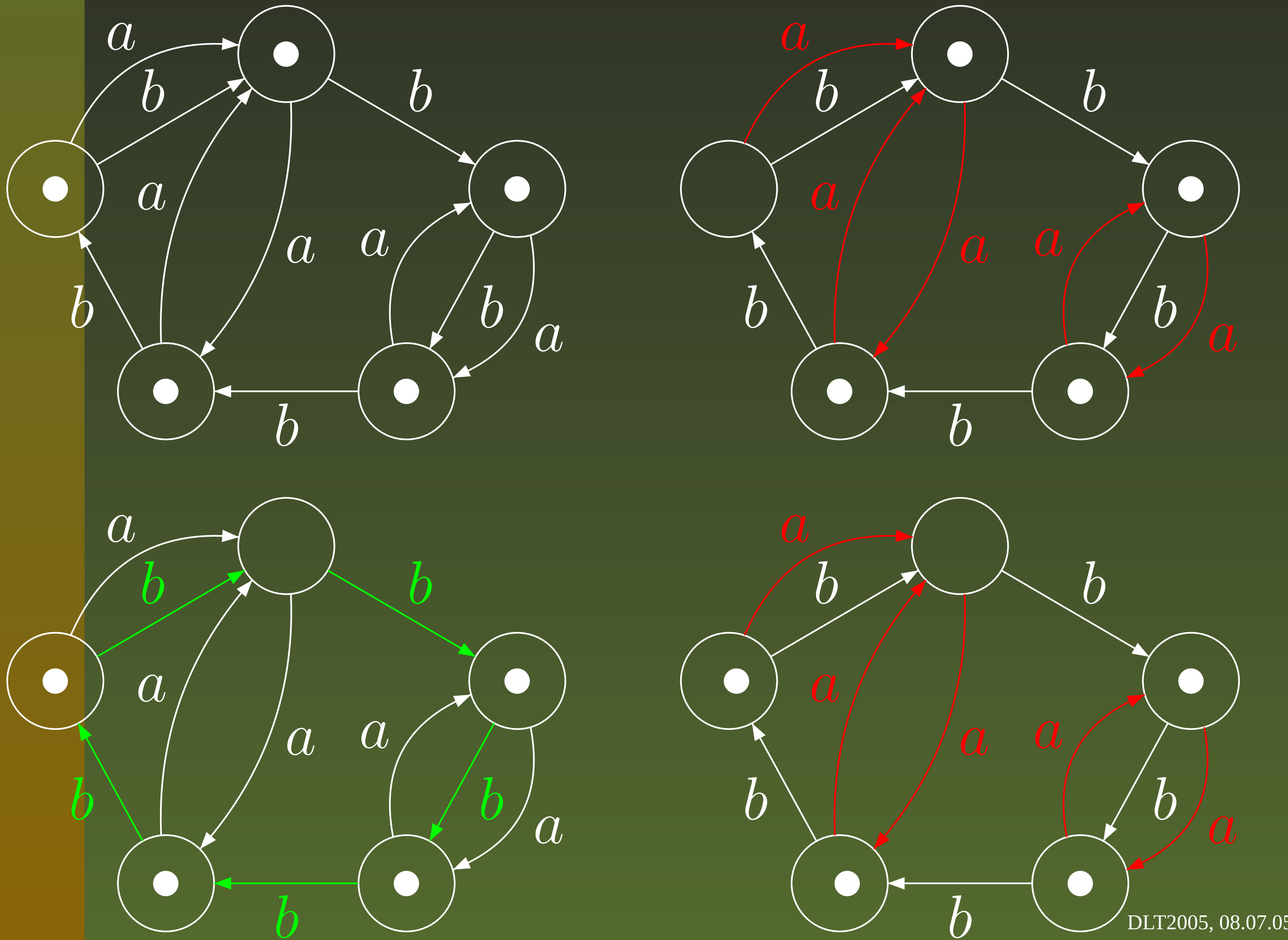
What Are Collapsing Words?



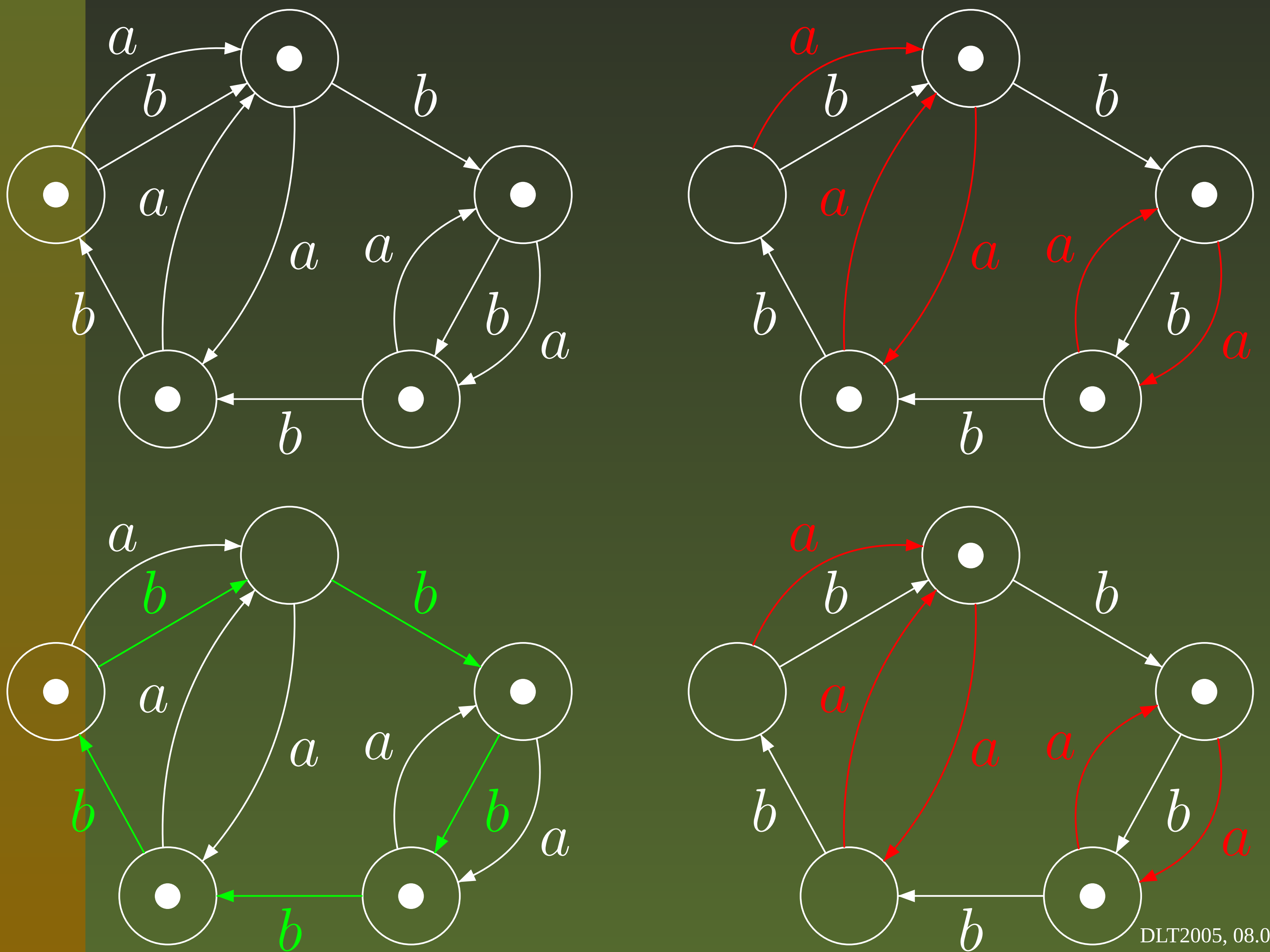




What Are Collapsing Words?



What Are Collapsing Words?



What Are Collapsing Words?

It is easy to realize that, for any word $v \in \Sigma^*$, the corresponding sequence of moves leaves tokens precisely on the states in $Q\,.\,v$.

What Are Collapsing Words?

It is easy to realize that, for any word $v \in \Sigma^*$, the corresponding sequence of moves leaves tokens precisely on the states in $Q \,.\, v$.

Therefore:

What Are Collapsing Words?

It is easy to realize that, for any word $v \in \Sigma^*$, the corresponding sequence of moves leaves tokens precisely on the states in $Q \,.\, v$.

Therefore:

- the deficiency $\mathrm{df}_{\mathscr{A}}(v)$ of v is exactly the number of freed states;

What Are Collapsing Words?

It is easy to realize that, for any word $v \in \Sigma^*$, the corresponding sequence of moves leaves tokens precisely on the states in $Q \, . \, v$.

Therefore:

- the deficiency $\mathrm{df}_{\mathscr{A}}(v)$ of v is exactly the number of freed states;
- $\mathscr{A}$ is n-compressible iff there is a winning sequence of moves for the game on the board $\Gamma(\mathscr{A})$ with the score n to be achieved;

What Are Collapsing Words?

It is easy to realize that, for any word $v \in \Sigma^*$, the corresponding sequence of moves leaves tokens precisely on the states in $Q \,.\, v$.

Therefore:

- the deficiency $\mathrm{df}_{\mathscr{A}}(v)$ of v is exactly the number of freed states;
- $\mathscr{A}$ is n-compressible iff there is a winning sequence of moves for the game on the board $\Gamma(\mathscr{A})$ with the score n to be achieved;
- a word w is n-collapsing iff it represents a universal winning strategy for all games (with the same names of moves and the score n) that can be won.

What Are They For?

Collapsing word admit several interesting applications in:

What Are They For?

Collapsing word admit several interesting applications in:

- universal algebra (Pöschel, Sapir, Sauer, Stone, and $\sim$, 1994);

What Are They For?

Collapsing word admit several interesting applications in:

- universal algebra (Pöschel, Sapir, Sauer, Stone, and $\sim$, 1994);
- theory of finite and profinite semigroups (Almeida and $\sim$, 2003, 2005; Reilly and Zhang, 2000);

What Are They For?

Collapsing word admit several interesting applications in:

- universal algebra (Pöschel, Sapir, Sauer, Stone, and $\sim$, 1994);
- theory of finite and profinite semigroups (Almeida and $\sim$, 2003, 2005; Reilly and Zhang, 2000);
- computational complexity ($\sim$, see Proceedings).

What Are They For?

Collapsing word admit several interesting applications in:

- universal algebra (Pöschel, Sapir, Sauer, Stone, and $\sim$, 1994);
- theory of finite and profinite semigroups (Almeida and $\sim$, 2003, 2005; Reilly and Zhang, 2000);
- computational complexity ($\sim$, see Proceedings).

Roughly speaking, collapsing words allow one to reduce in a uniform way studying arbitrary transformations (semigroup theory) to studying reversible transformations (group theory).

What May They Be For?

- *DNA-computing*, cf. reports of Shapiro’s group on ‘soup of automata’:

What May They Be For?

- *DNA-computing*, cf. reports of Shapiro's group on 'soup of automata':
  - Y. Benenson et al, Programmable and autonomous computing machine made of biomolecules, Nature Vol.414, no.1 (November 22, 2001) 430–434.

What May They Be For?

- *DNA-computing*, cf. reports of Shapiro's group on 'soup of automata':
  - Y. Benenson et al, Programmable and autonomous computing machine made of biomolecules, Nature Vol.414, no.1 (November 22, 2001) 430–434.
  - Y. Benenson et al, DNA molecule provides a computing machine with both data and fuel, Proc. National Acad. Sci. USA Vol.100 (2003) 2191–2196.

What May They Be For?

- *DNA-computing*, cf. reports of Shapiro's group on 'soup of automata':
  - Y. Benenson et al, Programmable and autonomous computing machine made of biomolecules, Nature Vol.414, no.1 (November 22, 2001) 430–434.
  - Y. Benenson et al, DNA molecule provides a computing machine with both data and fuel, Proc. National Acad. Sci. USA Vol.100 (2003) 2191–2196.

Here collapsing words (over the DNA alphabet) might play a role of a 'RESET' button.

What May They Be For?

- *robotics*, more precisely, part handling problems in industrial automation such as part feeding, fixturing, loading, assembly and packing, cf., e.g.:

What May They Be For?

- *robotics*, more precisely, part handling problems in industrial automation such as part feeding, fixturing, loading, assembly and packing, cf., e.g.:
  - B. K. Natarajan, An algorithmic approach to the automated design of parts orienters, Proc. 27th Annual Symp. Foundations Comput. Sci., IEEE, 1986, 132–142.

What May They Be For?

- *robotics*, more precisely, part handling problems in industrial automation such as part feeding, fixturing, loading, assembly and packing, cf., e.g.:
  - B. K. Natarajan, An algorithmic approach to the automated design of parts orienters, Proc. 27th Annual Symp. Foundations Comput. Sci., IEEE, 1986, 132–142.
  - B. K. Natarajan, Some paradigms for the automated design of parts feeders, Internat. J. Robotics Research, Vol.8, no.6 (1989) 89–109.

What May They Be For?

- *robotics*, more precisely, part handling problems in industrial automation such as part feeding, fixturing, loading, assembly and packing, cf., e.g.:
  - B. K. Natarajan, An algorithmic approach to the automated design of parts orienters, Proc. 27th Annual Symp. Foundations Comput. Sci., IEEE, 1986, 132–142.
  - B. K. Natarajan, Some paradigms for the automated design of parts feeders, Internat. J. Robotics Research, Vol.8, no.6 (1989) 89–109.

Here collapsing words precisely correspond to engineers' dreams about UFO (Universal Feeder-Orienter).

How Can They Be Constructed?

Fact: n-collapsing words exist for each n and over each finite alphabet Σ (Sauer and Stone, 1991).

How Can They Be Constructed?

Fact: n-collapsing words exist for each n and over each finite alphabet Σ (Sauer and Stone, 1991).

At present, there are two general constructions.

How Can They Be Constructed?

Fact: n-collapsing words exist for each n and over each finite alphabet Σ (Sauer and Stone, 1991).

At present, there are two general constructions.

- An inductive construction invented by Sauer and Stone and improved by Margolis, Pin and $\sim$, 2004.

How Can They Be Constructed?

Fact: n-collapsing words exist for each n and over each finite alphabet Σ (Sauer and Stone, 1991).

At present, there are two general constructions.

- An inductive construction invented by Sauer and Stone and improved by Margolis, Pin and $\sim$, 2004.

  It leads to a word of length $|\Sigma|^{\frac{n^2-n}{2}} + \dots$.

How Can They Be Constructed?

Fact: n-collapsing words exist for each n and over each finite alphabet Σ (Sauer and Stone, 1991).

At present, there are two general constructions.

- An inductive construction invented by Sauer and Stone and improved by Margolis, Pin and $\sim$, 2004.

  It leads to a word of length $|\Sigma|^{\frac{n^2-n}{2}} + \dots$.
- A construction related to the so-called rank conjecture, a generalization of the well known Černý's conjecture.

How Can They Be Constructed?

Fact: n-collapsing words exist for each n and over each finite alphabet Σ (Sauer and Stone, 1991).

At present, there are two general constructions.

- An inductive construction invented by Sauer and Stone and improved by Margolis, Pin and $\sim$, 2004.

  It leads to a word of length $|\Sigma|^{\frac{n^2-n}{2}} + \ldots$.
- A construction related to the so-called rank conjecture, a generalization of the well known Černý's conjecture. It leads to a word of length $|\Sigma|^{n^2} + \ldots$ modulo the conjecture and of length $|\Sigma|^{\frac{n(n+1)(n+2)}{6}} + \ldots$ unconditionally.

The Inductive Construction

Observe that if a DFA $\mathscr{A}$ is 1-compressible then there exists a letter $\boldsymbol{a} \in \boldsymbol{\Sigma}$ with $\mathrm{df}_{\mathscr{A}}(\boldsymbol{a}) \geq \mathbf{1}$.

The Inductive Construction

Observe that if a DFA $\mathscr{A}$ is 1-compressible then there exists a letter $a \in \Sigma$ with $\mathrm{df}_{\mathscr{A}}(a) \geq 1$. Therefore the product w_1 of all letters in Σ (in any order) is a 1-collapsing word over Σ.

The Inductive Construction

Observe that if a DFA $\mathscr{A}$ is 1-compressible then there exists a letter $a \in \Sigma$ with $\mathrm{df}_{\mathscr{A}}(a) \geq 1$. Therefore the product w_1 of all letters in Σ (in any order) is a 1-collapsing word over Σ.

Now for $n > 1$ define

$$w_n = w_{n-1} \prod_{0 \leq |v| \leq n} (v w_{n-1}).$$

Again, the order of the v's plays no role.

The Inductive Construction

Observe that if a DFA $\mathscr{A}$ is 1-compressible then there exists a letter $a \in \Sigma$ with $\mathrm{df}_{\mathscr{A}}(a) \geq 1$. Therefore the product w_1 of all letters in Σ (in any order) is a 1-collapsing word over Σ.

Now for $n > 1$ define

$$w_n = w_{n-1} \prod_{0 \leq |v| \leq n} (v w_{n-1}).$$

Again, the order of the v's plays no role.
Then w_n turns out to be an n-collapsing word over Σ.

The Inductive Construction

Observe that if a DFA $\mathscr{A}$ is 1-compressible then there exists a letter $a \in \Sigma$ with $\mathrm{df}_{\mathscr{A}}(a) \geq 1$. Therefore the product w_1 of all letters in Σ (in any order) is a 1-collapsing word over Σ.

Now for $n > 1$ define

$$w_n = w_{n-1} \prod_{0 \leq |v| \leq n} (v w_{n-1}).$$

Again, the order of the v's plays no role.
Then w_n turns out to be an n-collapsing word over Σ.

In Sauer-Stone's version, the product was over all $v \in \Sigma^*$ with $|v| \leq 2^{n-3} + 1$.

The Rank Conjecture Construction

Define the *deficiency* of a DFA $\mathscr{A}$ as

$$\mathrm{df}(\mathscr{A}) = \max_{v \in \Sigma^*}\{\mathrm{df}_{\mathscr{A}}(v)\}.$$

The Rank Conjecture Construction

Define the *deficiency* of a DFA $\mathscr{A}$ as

$$\mathrm{df}(\mathscr{A}) = \max_{v \in \Sigma^*} \{\mathrm{df}_{\mathscr{A}}(v)\}.$$

This is the maximum score that can be achieved in the solitaire game on the board $\Gamma(\mathscr{A})$.

The Rank Conjecture Construction

Define the *deficiency* of a DFA $\mathscr{A}$ as

$$\mathbf{df}(\mathscr{A}) = \max_{v \in \Sigma^*} \{\mathbf{df}_{\mathscr{A}}(v)\}.$$

This is the maximum score that can be achieved in the solitaire game on the board $\Gamma(\mathscr{A})$.

The rank conjecture claims that if $\mathbf{df}(\mathscr{A}) = k$ then there exists a word $w \in \Sigma^*$ such that $|w| = k^2$ and $\mathbf{df}_{\mathscr{A}}(w) = k$.

The Rank Conjecture Construction

Define the *deficiency* of a DFA $\mathscr{A}$ as

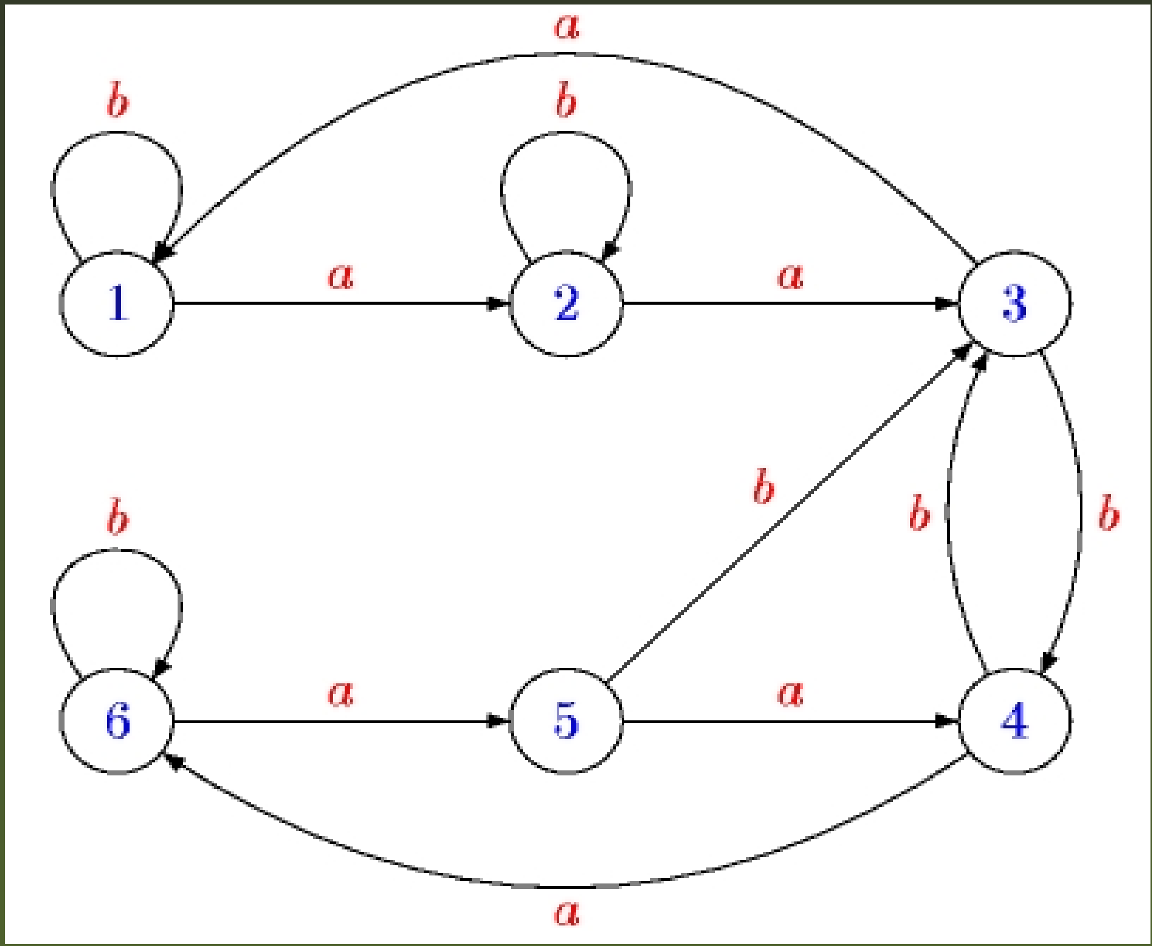
$$\mathrm{df}(\mathscr{A}) = \max_{v \in \Sigma^*}\{\mathrm{df}_{\mathscr{A}}(v)\}.$$

This is the maximum score that can be achieved in the solitaire game on the board $\Gamma(\mathscr{A})$.

The rank conjecture claims that if $\mathrm{df}(\mathscr{A}) = k$ then there exists a word $w \in \Sigma^*$ such that $|w| = k^2$ and $\mathrm{df}_{\mathscr{A}}(w) = k$. Thus, the highest possible score can be always achieved via a suitable sequence of k^2 moves.

The Rank Conjecture Construction

Kari's automaton does **not** refute the rank conjecture!



The Rank Conjecture Construction

In the solitaire game on Kari’s digraph, no sequence of 16 moves removes 4 tokens.

The Rank Conjecture Construction

In the solitaire game on Kari's digraph, no sequence of 16 moves removes 4 tokens. However, 4 is **not** the maximum number of tokens that can be removed!

The Rank Conjecture Construction

In the solitaire game on Kari’s digraph, no sequence of 16 moves removes 4 tokens. However, 4 is **not** the maximum number of tokens that can be removed! One can show that 5 states can be freed by a sequence of 25 moves — in full accordance with the rank conjecture.

The Rank Conjecture Construction

In the solitaire game on Kari's digraph, no sequence of 16 moves removes 4 tokens. However, 4 is **not** the maximum number of tokens that can be removed! One can show that 5 states can be freed by a sequence of 25 moves — in full accordance with the rank conjecture.

Thus, the conjecture remains open.

The Rank Conjecture Construction

In the solitaire game on Kari's digraph, no sequence of 16 moves removes 4 tokens. However, 4 is **not** the maximum number of tokens that can be removed! One can show that 5 states can be freed by a sequence of 25 moves — in full accordance with the rank conjecture.

Thus, the conjecture remains open. It can be shown that if it holds true then any word containing all $v \in \Sigma^*$ with $|v| = n^2$ as factors is an n-collapsing word over Σ.

The Rank Conjecture Construction

In the solitaire game on Kari's digraph, no sequence of 16 moves removes 4 tokens. However, 4 is **not** the maximum number of tokens that can be removed! One can show that 5 states can be freed by a sequence of 25 moves — in full accordance with the rank conjecture.

Thus, the conjecture remains open. It can be shown that if it holds true then any word containing all $v \in \Sigma^*$ with $|v| = n^2$ as factors is an n-collapsing word over Σ.

The best estimate on the rank conjecture known so far is due to Pin, 1983.

The Rank Conjecture Construction

In the solitaire game on Kari's digraph, no sequence of 16 moves removes 4 tokens. However, 4 is **not** the maximum number of tokens that can be removed! One can show that 5 states can be freed by a sequence of 25 moves — in full accordance with the rank conjecture.

Thus, the conjecture remains open. It can be shown that if it holds true then any word containing all $v \in \Sigma^*$ with $|v| = n^2$ as factors is an n-collapsing word over Σ.

The best estimate on the rank conjecture known so far is due to Pin, 1983. Using it, one can show any word containing all $v \in \Sigma^*$ with $|v| = n(n+1)(n+2)/6$ as factors is n-collapsing.

Discussion

- Both the constructions depend on non-trivial results from combinatorics of finite sets.

Discussion

- Both the constructions depend on non-trivial results from combinatorics of finite sets.
- Both the constructions are unsatisfactory from the viewpoint of possible real world applications.

Discussion

- Both the constructions depend on non-trivial results from combinatorics of finite sets.
- Both the constructions are unsatisfactory from the viewpoint of possible real world applications.
- The upper bounds provided by the constructions are very far from being precise.

Discussion

- Both the constructions depend on non-trivial results from combinatorics of finite sets.
- Both the constructions are unsatisfactory from the viewpoint of possible real world applications.
- The upper bounds provided by the constructions are very far from being precise.
- The inductive construction is better even if the rank conjecture is true. The optimal universal winning strategy should not consist of optimal local winning strategies glued together!

Discussion

- Both the constructions depend on non-trivial results from combinatorics of finite sets.
- Both the constructions are unsatisfactory from the viewpoint of possible real world applications.
- The upper bounds provided by the constructions are very far from being precise.
- The inductive construction is better even if the rank conjecture is true. The optimal universal winning strategy should not consist of optimal local winning strategies glued together!
- The rank conjecture construction may be better for some restricted classes (for aperiodic automata, say).

Discussion

The rank conjecture construction is based on the worst case approach. However, experiments show that ‘complicated’ game-boards (where the length of the shortest winning sequence is close to the square of score) are extremely rare.

Discussion

The rank conjecture construction is based on the worst case approach. However, experiments show that ‘complicated’ game-boards (where the length of the shortest winning sequence is close to the square of score) are extremely rare. We present sample results of such experiments on the next two slides.

Discussion

The rank conjecture construction is based on the worst case approach. However, experiments show that 'complicated' game-boards (where the length of the shortest winning sequence is close to the square of score) are extremely rare. We present sample results of such experiments on the next two slides.

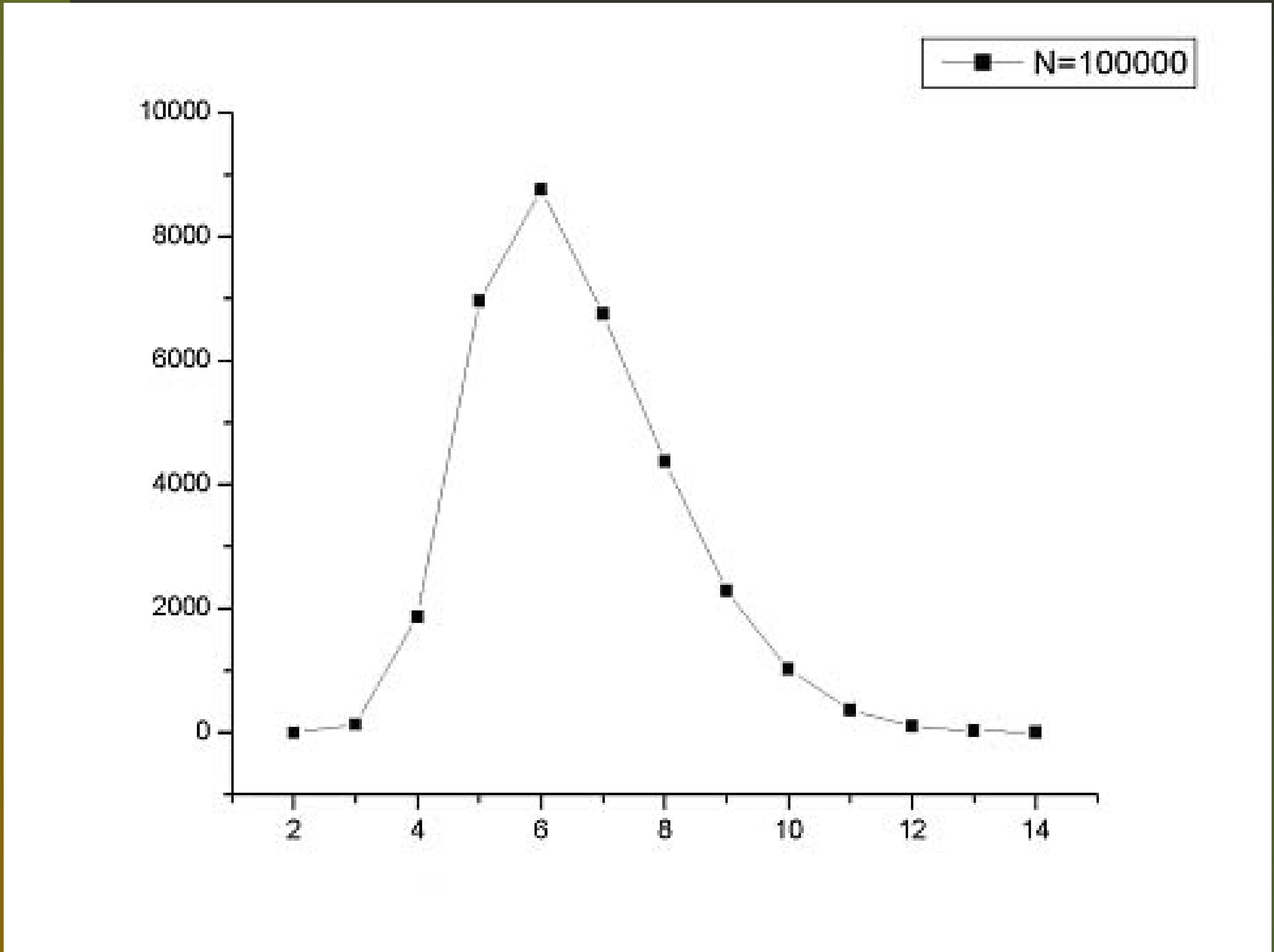
They give some indication that a more economic construction for n-collapsing words may based on the average case approach:

Discussion

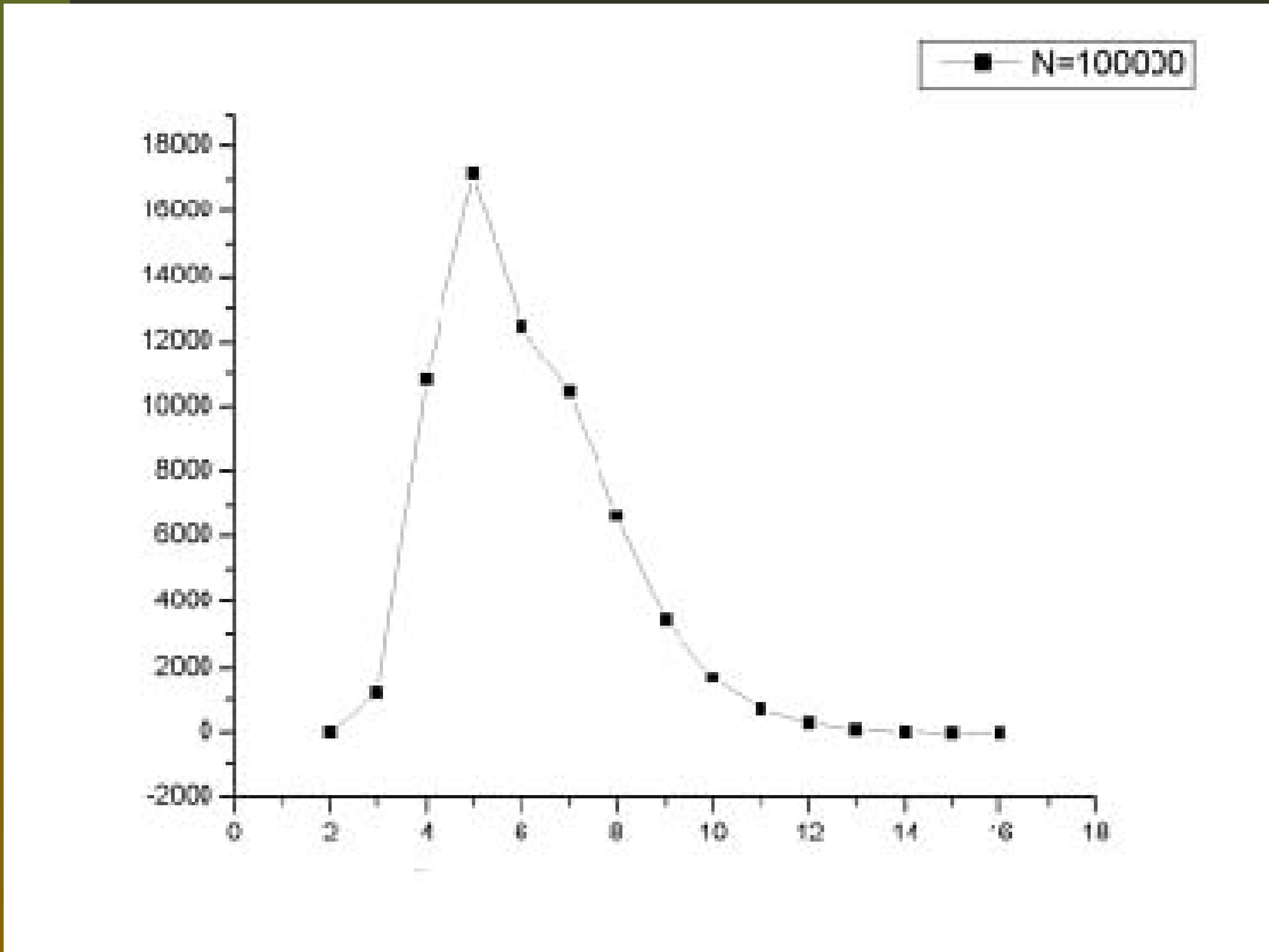
The rank conjecture construction is based on the worst case approach. However, experiments show that ‘complicated’ game-boards (where the length of the shortest winning sequence is close to the square of score) are extremely rare. We present sample results of such experiments on the next two slides.

They give some indication that a more economic construction for n-collapsing words may based on the average case approach: include all $v \in \Sigma^*$ with $|v| = n$ as factors (this is certainly necessary) and add just a few longer factors to serve all possible ‘complicated’ game-boards.

20-State Game-Boards



30-State Game-Boards



Recognizing Collapsing Words

Given a word w and a positive integer n, can we recognize whether or not w is n-collapsing?

Recognizing Collapsing Words

Given a word w and a positive integer n, can we recognize whether or not w is n-collapsing?

In other terms: suppose that w is **not** n-collapsing.

Recognizing Collapsing Words

Given a word w and a positive integer n, can we recognize whether or not w is n-collapsing?

In other terms: suppose that w is **not** n-collapsing. Then there exists an n-compressible DFA $\mathscr{A}$ such that $\mathrm{df}_{\mathscr{A}}(w) < n$.

Recognizing Collapsing Words

Given a word w and a positive integer n, can we recognize whether or not w is n-collapsing?

In other terms: suppose that w is **not** n-collapsing. Then there exists an n-compressible DFA $\mathscr{A}$ such that $\mathrm{df}_{\mathscr{A}}(w) < n$. Observe that $\mathscr{A}$ should be determined entirely by w. Can we efficiently extract $\mathscr{A}$ from w?

Recognizing Collapsing Words

Given a word w and a positive integer n, can we recognize whether or not w is n-collapsing?

In other terms: suppose that w is **not** n-collapsing. Then there exists an n-compressible DFA $\mathscr{A}$ such that $\mathrm{df}_{\mathscr{A}}(w) < n$. Observe that $\mathscr{A}$ should be determined entirely by w. Can we efficiently extract $\mathscr{A}$ from w?

The answer is 'yes' for $n = 2$ (Ananichev, Cherubini and $\sim$, 2003).

Recognizing Collapsing Words

Given a word w and a positive integer n, can we recognize whether or not w is n-collapsing?

In other terms: suppose that w is **not** n-collapsing. Then there exists an n-compressible DFA $\mathscr{A}$ such that $\mathrm{df}_{\mathscr{A}}(w) < n$. Observe that $\mathscr{A}$ should be determined entirely by w. Can we efficiently extract $\mathscr{A}$ from w?

The answer is 'yes' for $n = 2$ (Ananichev, Cherubini and $\sim$, 2003). The algorithm that relies on combinatorial group theory has been implemented by Petrov, and experiments with the implementation have led to many interesting examples and several useful observations.

Recognizing Collapsing Words

So far we have not succeeded in finding an algorithm of the same flavor for recognizing n-collapsing words with $n > 2$. Therefore we have tried another approach.

Recognizing Collapsing Words

So far we have not succeeded in finding an algorithm of the same flavor for recognizing n-collapsing words with $n > 2$. Therefore we have tried another approach.

A byproduct of our algorithm for $n = 2$ is the following fact: if w is not 2-collapsing, then $\mathbf{df}_{\mathscr{A}}(w) < 2$ for some 2-compressible DFA $\mathscr{A}$ with $\leq |w|$ states.

Recognizing Collapsing Words

So far we have not succeeded in finding an algorithm of the same flavor for recognizing n-collapsing words with $n > 2$. Therefore we have tried another approach.

A byproduct of our algorithm for $n = 2$ is the following fact: if w is not 2-collapsing, then $\mathbf{df}_{\mathscr{A}}(w) < 2$ for some 2-compressible DFA $\mathscr{A}$ with $\leq |w|$ states.

Since the number of DFA with $\leq |w|$ states and a fixed input alphabet Σ is finite, this corollary implies that the set $\mathcal{C}_2$ of all 2-collapsing words over Σ is recursive.

Recognizing Collapsing Words

So far we have not succeeded in finding an algorithm of the same flavor for recognizing n-collapsing words with $n > 2$. Therefore we have tried another approach.

A byproduct of our algorithm for $n = 2$ is the following fact: if w is not 2-collapsing, then $\mathbf{df}_{\mathscr{A}}(w) < 2$ for some 2-compressible DFA $\mathscr{A}$ with $\leq |w|$ states.

Since the number of DFA with $\leq |w|$ states and a fixed input alphabet Σ is finite, this corollary implies that the set $\mathcal{C}_2$ of all 2-collapsing words over Σ is recursive.

Moreover, it leads to a non-deterministic linear space and polynomial time algorithm recognizing the complement of $\mathcal{C}_2$ whence $\mathcal{C}_2$ is a context-sensitive language.

Recognizing Collapsing Words

Recently Petrov has managed to find a similar result for an arbitrary n.

Recognizing Collapsing Words

Recently Petrov has managed to find a similar result for an arbitrary n.

Namely, for every word $\boldsymbol{w}$ which is not $\boldsymbol{n}$-collapsing, there exists an $\boldsymbol{n}$-compressible DFA $\mathscr{A}$ with $\leq 3|\boldsymbol{w}|(\boldsymbol{n}-1)+\boldsymbol{n}+1$ states and such that $\mathrm{df}_{\mathscr{A}}(\boldsymbol{w})<\boldsymbol{n}$.

Recognizing Collapsing Words

Recently Petrov has managed to find a similar result for an arbitrary n.

Namely, for every word $\boldsymbol{w}$ which is not $\boldsymbol{n}$-collapsing, there exists an $\boldsymbol{n}$-compressible DFA $\mathscr{A}$ with $\leq 3|\boldsymbol{w}|(\boldsymbol{n}-1)+\boldsymbol{n}+1$ states and such that $\mathbf{df}_{\mathscr{A}}(\boldsymbol{w})<\boldsymbol{n}$.

This implies that the set of all 2-collapsing words over each finite alphabet is a recursive and even context-sensitive language.

Recognizing Collapsing Words

The main idea is simple. Since w is not n-collapsing, there is an n-compressible DFA $\mathscr{B}$ and such that $\mathrm{df}_{\mathscr{B}}(w) < n$.

Recognizing Collapsing Words

The main idea is simple. Since w is not n-collapsing, there is an n-compressible DFA $\mathscr{B}$ and such that $\mathrm{df}_{\mathscr{B}}(w) < n$. If we treat w as a sequence of moves in the solitaire game on $\Gamma(\mathscr{B})$, then after each move at most $n - 1$ states are free of tokens.

Recognizing Collapsing Words

The main idea is simple. Since w is not n-collapsing, there is an n-compressible DFA $\mathscr{B}$ and such that $\mathrm{df}_{\mathscr{B}}(w) < n$. If we treat w as a sequence of moves in the solitaire game on $\Gamma(\mathscr{B})$, then after each move at most $n-1$ states are free of tokens. It turns out that all essential information about the game can be retrieved if at each move we control the following state sets:

Recognizing Collapsing Words

The main idea is simple. Since w is not n-collapsing, there is an n-compressible DFA $\mathscr{B}$ and such that $\mathrm{df}_{\mathscr{B}}(w) < n$. If we treat w as a sequence of moves in the solitaire game on $\Gamma(\mathscr{B})$, then after each move at most $n - 1$ states are free of tokens. It turns out that all essential information about the game can be retrieved if at each move we control the following state sets:

- the set of all current free states;

Recognizing Collapsing Words

The main idea is simple. Since w is not n-collapsing, there is an n-compressible DFA $\mathscr{B}$ and such that $\mathrm{df}_{\mathscr{B}}(w) < n$. If we treat w as a sequence of moves in the solitaire game on $\Gamma(\mathscr{B})$, then after each move at most $n - 1$ states are free of tokens. It turns out that all essential information about the game can be retrieved if at each move we control the following state sets:

- the set of all current free states;
- the set of all 'post-free' states to which the current letter brings states that were empty before the move;

Recognizing Collapsing Words

The main idea is simple. Since w is not n-collapsing, there is an n-compressible DFA $\mathscr{B}$ and such that $\mathrm{df}_{\mathscr{B}}(w) < n$. If we treat w as a sequence of moves in the solitaire game on $\Gamma(\mathscr{B})$, then after each move at most $n-1$ states are free of tokens. It turns out that all essential information about the game can be retrieved if at each move we control the following state sets:

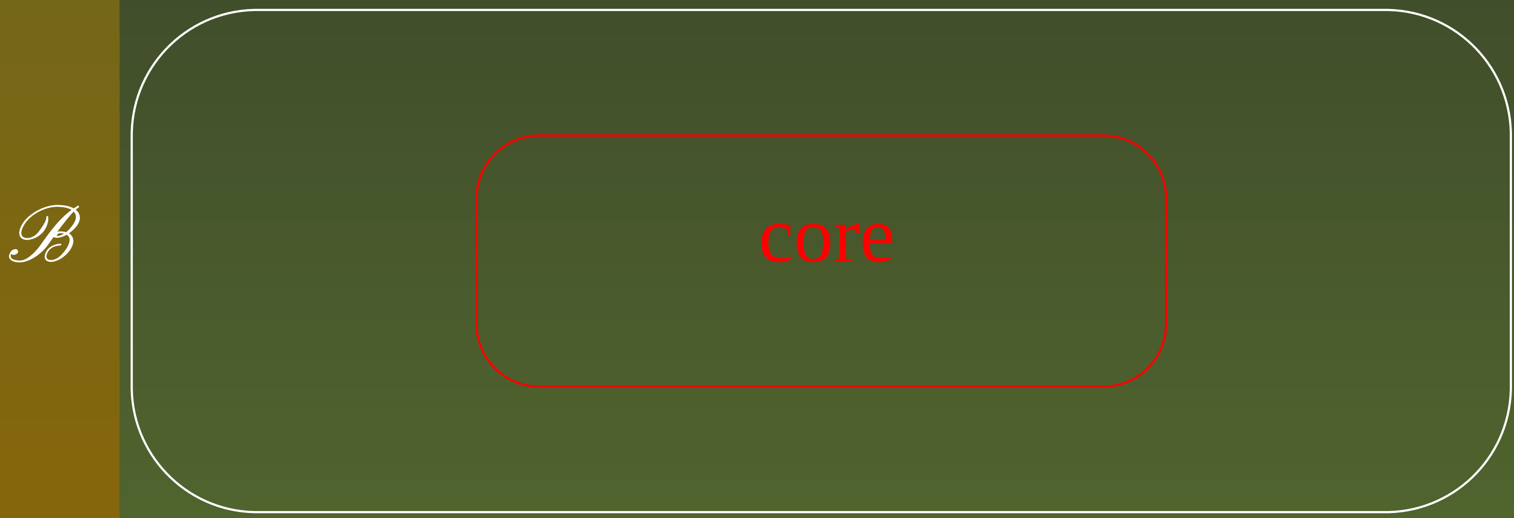
- the set of all current free states;
- the set of all 'post-free' states to which the current letter brings states that were empty before the move;
- the set of all 'next-to-free' states that could be achieved from current free states if we would repeat the same move.

Recognizing Collapsing Words

The total number of states subject to control does not exceed $3|w|(n-1)$.

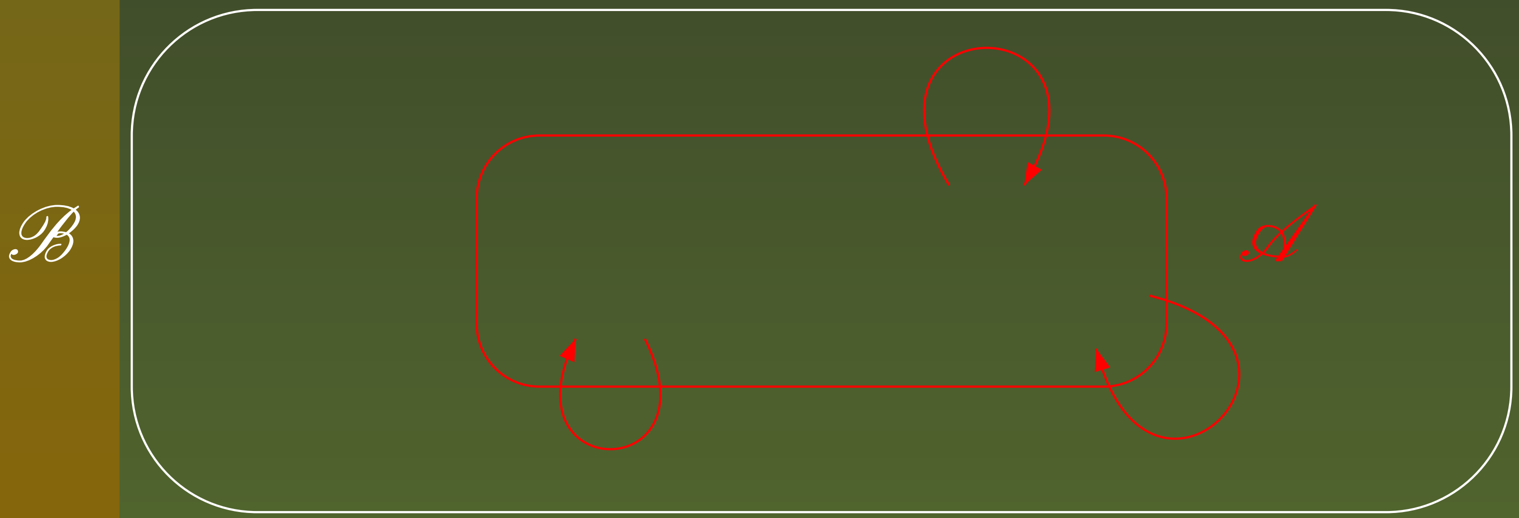
Recognizing Collapsing Words

The total number of states subject to control does not exceed $3|w|(n-1)$. They form a core of the 'small' n-compressible DFA $\mathscr{A}$ which witness that w is not n-collapsing



Recognizing Collapsing Words

The total number of states subject to control does not exceed $3|w|(n-1)$. They form a core of the ‘small’ n-compressible DFA $\mathscr{A}$ which witness that w is not n-collapsing while ‘irrelevant’ parts of $\mathscr{B}$ are substituted by some tiny buffer automata adding at most $n+1$ extra states.



Conclusion

- Collapsing words constitute a natural object for combinatorics of words and admit numerous applications, especially in algebra.

Conclusion

- Collapsing words constitute a natural object for combinatorics of words and admit numerous applications, especially in algebra.
- A significant progress in studying collapsing words has been achieved over the last 5 years. Still many intriguing questions remain widely open.

Conclusion

- Collapsing words constitute a natural object for combinatorics of words and admit numerous applications, especially in algebra.
- A significant progress in studying collapsing words has been achieved over the last 5 years. Still many intriguing questions remain widely open.
- The game interpretation underlying in this talk is not only entertaining but also provides some important insights. In particular, it was essential for finding a decision procedure for the property of being n-collapsing.

Conclusion

- Collapsing words constitute a natural object for combinatorics of words and admit numerous applications, especially in algebra.
- A significant progress in studying collapsing words has been achieved over the last 5 years. Still many intriguing questions remain widely open.
- The game interpretation underlying in this talk is not only entertaining but also provides some important insights. In particular, it was essential for finding a decision procedure for the property of being n-collapsing.
- I hope this talk did not make you collapse…