

# Automata, words and groups

Dominique Perrin

19 janvier 2011

- 1 Historical background
  - Fundamental groups
  - One-relator groups
  - Lyndon words
  - Bifix codes
  - Sturmian words
- 2 Bifix codes in Sturmian sets
  - Episturmian words
  - Cardinality
  - Periodicity
  - Sturmian bases
  - Syntactic groups

## Fundamental groups of surfaces



Max Dehn.

Fundamental group of a surface : Equivalence classes of closed paths around the origin modulo homotopy (Poincaré, 1892).

Homeomorphism problem for surfaces  $\rightarrow$  Isomorphism problem for groups.

## Theorem (Dehn, Heegard, 1907)

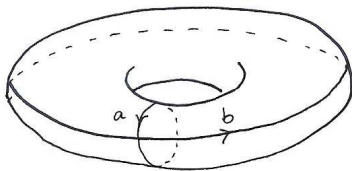
*The fundamental group of a finite closed surface is a one-relator group.*

Orientable case :

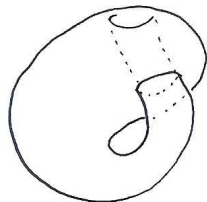
$$\langle a_1, b_1, \dots, a_n, b_n \mid [a_1 b_1] \cdots [a_n b_n] \rangle$$

Nonorientable case

$$\langle a_1, \dots, a_n, \mid a_1^2 \cdots a_n^2 \rangle$$

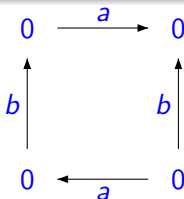
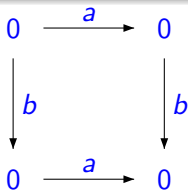


Torus



Klein's bottle

## Examples



Combinatorial representations of the torus (left) and Klein's bottle (right).

Fundamental groups :

$$\langle a, b \mid ab\bar{a}\bar{b} \rangle,$$

$$\langle a, b \mid a\bar{b}ab \rangle,$$

The last one is isomorphic (via  $c = \bar{a}\bar{b}$ ) to  $\langle b, c \mid c^2b^2 \rangle$ .

# The Freiheitsatz

## Theorem (Magnus, 1932)

Let  $G = \langle A \mid r \rangle$  with  $r$  cyclically reduced containing  $a \in A$ . Then  $A \setminus a$  is a basis of a free subgroup of  $G$ .

The word problem in a one-relator group is solvable. The answer is not known for one-relator semigroups.

# Lyndon words



A **lyndon word** is a primitive word which is minimal in its conjugacy class.

*a, b*

*ab,*

*aab, abb,*

*aaab, aabb, abbb,*

*aaaab, aaabb, aabab, aabbb, abbbb*

Any Lyndon word which is not a letter has a **standard factorization** obtained as follows : let  $w$  be a Lyndon word and let  $u$  be the longest proper prefix of  $w$  which is a Lyndon word. Set  $w = uv$ .

Then  $(u, v)$  is the standard factorization of  $w$ .

Iterating the standard factorization defines a standard **bracketting** of any Lyndon word.

For example  $aabab \rightarrow [[a[ab]][ab]]$  where  $[u, v]$  denotes the **commutator** of  $u, v$  defined by  $[u, v] = uvu^{-1}v^{-1}$ .

Let  $G$  be a group. Let  $\Gamma_1(G) = G$  and for  $k \geq 1$ , let  $\Gamma_{k+1}(G)$  be the group generated by the commutators  $[u, v]$  for  $u \in G$  and  $v \in \Gamma_k(G)$ .

The sequence  $(G, \Gamma_1(G), \Gamma_2(G), \dots)$  is called the *lower central series* of  $G$ .

**Theorem (Chen, Fox, Lyndon, 1958)**

*Bracketted Lyndon words give a basis of the quotients of the lower central series of the free group.*

Commutators of weight 5 :

$$[a[a[a[ab]]]], [a[a[ab]b]], [[a[ab]][ab]], [[[a[ab]]b]b], [[[[ab]b]b]b]$$

# Bifix codes

A **prefix code** is a set of words which does not contain any proper prefix of its elements

A **bifix code** is a set of words which does not contain any proper prefix or suffix of its elements (Schützenberger, 1956, Gilbert and Moore, 1959).

A set  $X$  of words over  $A \cup A^{-1}$  is **Nielsen reduced** if no element of  $X$  can disappear in a product of elements of  $X \cup X^{-1}$ .

Any set  $X$  of words in the free group which is Nielsen reduced is a bifix code (the converse is not true).

## Example

The set  $X = \{a, bab\}$  is a bifix code.

A prefix code is **maximal** if it is not properly contained in another prefix code on the same alphabet.

A set  $X$  of words is **limited** if there are no words of  $X$  which are not factors of any other word of  $X$  (example : a finite set).

A Bernoulli distribution is a morphism  $\pi : A^* \rightarrow [0, 1]$  such that  $\sum_{a \in A} \pi(a) = 1$ . For  $X \subset A^*$ , denote

$$\pi(X) = \sum_{x \in X} \pi(x), \quad \lambda(X) = \sum_{x \in X} |x| \pi(x).$$

### Theorem

*Let  $X$  be a limited maximal prefix code and let  $P$  be the set of proper prefixes of  $X$ . Let  $\pi$  a positive Bernoulli distribution on the alphabet. Then  $\pi(X) = 1$  and  $\lambda(X) = \pi(P) < \infty$ .*

## Group codes

Let  $\varphi : A^* \rightarrow G$  be a morphism from  $A^*$  onto a group  $G$ . If  $H$  is a subgroup of  $G$ , then  $\varphi^{-1}(H)$  is generated by a maximal bifix code called a **group code**.

The group code is limited if and only if the subgroup  $H$  is of finite index.

### Example

The set  $X = a \cup ba^*b$  is a group code. It is limited since  $b^3$  is not a factor of  $X$ . One has  $X^* = \varphi^{-1}(0)$  for the morphism  $\varphi : \{a, b\}^* \rightarrow \mathbb{Z}/2\mathbb{Z}$  defined by  $\varphi(w) = 0$  if  $w$  has an even number of  $b$  and 1 otherwise.

## Theorem (Schützenberger, 1961)

*The average length of a limited maximal bifix code with respect to a positive Bernoulli distribution on the alphabet is an integer.*

This integer is independent of the distribution and called the **degree** of the bifix code.

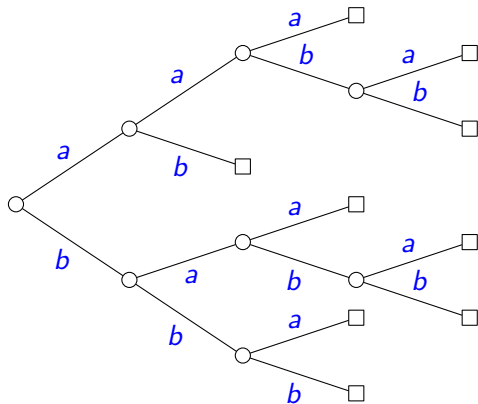
The degree of a group code  $X$  defined by a morphism  $\varphi : A^* \rightarrow G$  is equal to the index of the subgroup  $\varphi(X^*)$ .

### Example

The group code  $X = a \cup ba^*b$  has degree 2. With  $p = \pi(a)$  and  $q = \pi(b)$ , we have

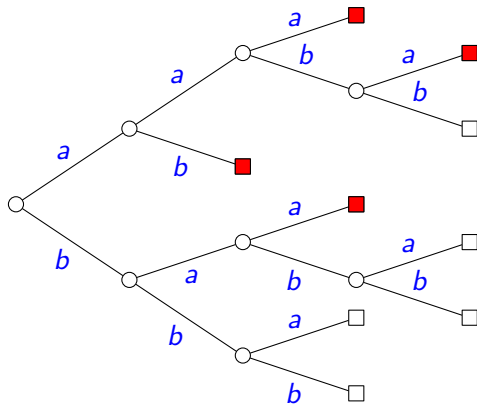
$$\lambda(X) = \pi(1 \cup ba^*) = 1 + q/(1 - p) = 1 + 1 = 2.$$

A finite maximal bifix code of degree 3. Many of the 4 subsets of  $X$  are basis of subgroups of index 3. For example :



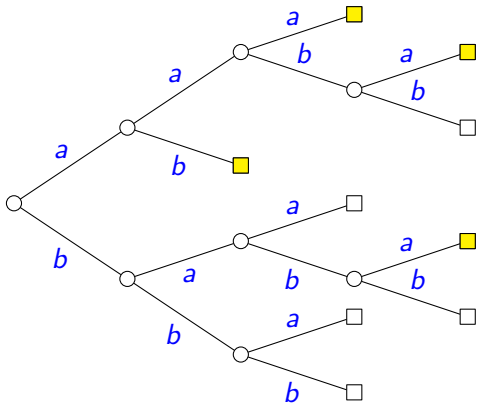
More about that in a moment.

A finite maximal bifix code of degree 3. Many of the 4 subsets of  $X$  are basis of subgroups of index 3. For example :



More about that in a moment.

A finite maximal bifix code of degree 3. Many of the 4 subsets of  $X$  are basis of subgroups of index 3. For example :



More about that in a moment.





# Sturmian words

An infinite word over a binary alphabet is called **Sturmian** if for all  $n \geq 0$ , the number of its factors of length  $n$  is  $n + 1$ .



As an example, the **Fibonacci word**  $f = abaababaabaab \dots$  which is defined as the unique fixed point of the substitution  $(a \mapsto ab, b \mapsto a)$  is Sturmian.



Leonardo di Pisa (Fibonacci)



Turku Energia chimney.

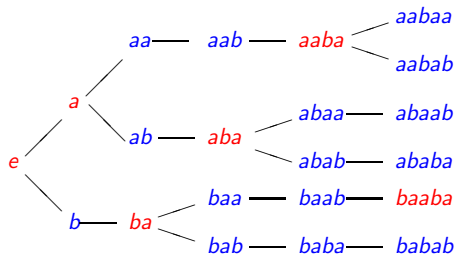


FIG.: The factors of the Fibonacci word.

## Mechanical words

There is also an alternative definition of the Fibonacci word using approximations of irrationals by rationals. Let  $\alpha$  be some irrational with  $0 < \alpha < 1$ , and let  $s(\alpha) = (s_n)$  be the sequence

$$s_n = \begin{cases} a & \text{if } \lfloor (n+1)\alpha \rfloor = \lfloor n\alpha \rfloor, \\ b & \text{otherwise} \end{cases}$$

For  $\alpha = 2/(3 + \sqrt{5})$ , one has  $s(\alpha) = af$ . This formula shows that the symbols  $s_n$  can be interpreted as the approximation of a line of slope  $\alpha$  by points with integer coordinates. It is a theorem due to Morse and Hedlund that Sturmian words can be defined equivalently by a formula as above with  $\alpha$  an irrational number.

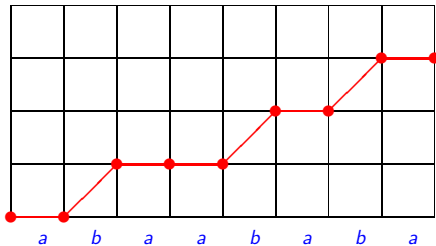


FIG.: The graphical representation of the Fibonacci word

Mechanical words (also called Christoffel words) have historical roots in the work of the astronomer Jean **Bernoulli** III who studied these words in connection with continued fractions.

We will establish new links between Sturmian words and free groups. A previous result in this direction is the following. Set  $A = \{a, b\}$ . A **Sturmian morphism** is a morphism  $f : A^* \rightarrow A^*$  such that  $f(x)$  is Sturmian for any Sturmian word  $x$ .

Theorem (Wen, Wen, 1994)

*The Sturmian morphisms are the positive automorphisms of the free group on  $A$ .*

The positive automorphisms are those which preserve the set of positive words. They are generated by the three elementary morphisms defined as

$$\psi_a : \begin{array}{l} a \mapsto a \\ b \mapsto ab \end{array}, \quad E : \begin{array}{l} a \mapsto b \\ b \mapsto a \end{array}, \quad \psi_b : \begin{array}{l} a \mapsto ba \\ b \mapsto b \end{array},$$

## Episturmian words

**Episturmian words** are a generalization of Sturmian words to arbitrary finite alphabets based on ideas of Rauzy and De Luca. Given a set  $F$  of words over an alphabet  $A$ , the right order of a word  $u$  in  $F$  is the number of letters  $a$  such that  $ua \in F$ . A word  $u$  is **right-special** if its right order is at least 2. A right-special word is **strict** if its right order is equal to  $\text{Card}(A)$ . In the case of a 2-letter alphabet, all special words are strict.

By definition, an infinite word  $x$  is **episturmian** if  $F(x)$  is closed under reversal and if  $F(x)$  contains, for each  $n \geq 1$ , at most one word  $u$  of length  $n$  which is right-special.

As a particular case, a **strict** episturmian word is an episturmian word  $x$  such that it has exactly one right-special factor of each length and moreover each right-special factor  $u$  is strict, that is satisfies the inclusion  $uA \subset F(x)$ .

An episturmian word is called **standard** if all its prefixes are left-special. For any episturmian word  $s$ , there is a standard one  $t$  such that  $F(s) = F(t)$ . This is the word that has as prefixes the left-special factors of  $s$ .

For a strict episturmian word  $x$  on an alphabet  $A$  with  $k$  letters, the set  $F(x) \cap A^n$  has  $(k-1)n + 1$  elements for each  $n$ . Thus, for a binary alphabet, the strict episturmian words are just the Sturmian words.

## Example

Consider the following generalization of the Fibonacci word to the ternary alphabet  $A = \{a, b, c\}$ . Let  $f : A^* \rightarrow A^*$  be the morphism defined by  $f(a) = ab$ ,  $f(b) = ac$  and  $f(c) = a$ . The fixpoint

$$f^\omega(a) = abacabaabacababacabaabacabacabaabacab \dots$$

is the **Tribonacci word**. It is a strict standard episturmian word.

Let  $F$  be a Sturmian set. A bifix code  $X \subset F$  is  $F$ -maximal in  $F$  if it is not properly contained in any other bifix code  $Y \subset F$ .

### Example

Let  $F$  be the Fibonacci set. The set of factors of length  $d$  of the Fibonacci word is an  $F$ -maximal bifix code with  $d + 1$  elements.

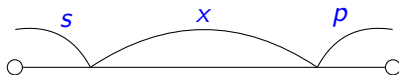
### Example

Let  $F$  be the Fibonacci set. The set  $X = \{a, bab, baab\}$  is an  $F$ -maximal bifix code.

Let  $X$  be a bifix code. A **parse** of a word  $w$  with respect to  $X$  is a triple  $(s, x, p)$  such that

- $s$  has no suffix in  $X$ ,
- $x \in X^*$
- $p$  has no prefix in  $X$
- $w = sxp$

The **degree** of  $X$ , denoted  $d(X)$  is the maximal number of parses of a word in  $X$ .



### Example

Let  $X = \{a, bab\}$ . The word  $w = abab$  has two parses, namely  $(1, abab, 1)$  and  $(ab, a, b)$ .

An **internal factor** of a word  $x$  is a word  $v$  such that  $x = uvw$  with  $u, w$  nonempty.

### Theorem

*Let  $F$  be a Sturmian set and let  $X \subset F$  be an  $F$ -maximal bifix code. The number of parses of a word  $w \in F$  is maximal if and only if  $w$  is not an internal factor of a word in  $X$ .*

### Example

Let  $F$  be the Fibonacci set. The bifix code  $X = \{a, bab, baab\}$  is an  $F$ -maximal bifix code of  $F$ -degree 2. Indeed, the word  $ba$  is not an internal factor and has two parses  $(b, a, 1)$  and  $(1, 1, ba)$ .

# The Cardinality Theorem

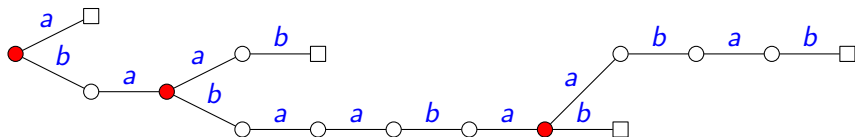
The following result generalizes the fact that a Sturmian word has  $d + 1$  factors of length  $d$ .

Theorem (Berstel, De Felice, Reutenauer, Rindone, P. 2010)

*Let  $F$  be a Sturmian set on an alphabet with  $k$  letters. For any finite  $F$ -maximal bifix code  $X \subset F$ , one has*  
$$\text{Card}(X) = (k - 1)d(X) + 1.$$

## Example

Consider the following bifix code  $X$ . It is  $F$ -maximal because any word in  $F$  is prefix comparable with a word of  $X$ . It has degree 3 because the word  $bab$  is not an internal factor and has 3 parses.



There are exactly three right-special words which are proper prefixes of words of  $X$ .

Let  $x = a_0a_1 \cdots$ , with  $a_i \in A$ , be an infinite word. It is **periodic** if there is an integer  $n \geq 1$  such that  $a_{i+n} = a_i$  for all  $i \geq 0$ . It is **ultimately periodic** if the equalities hold for all  $i$  large enough. Thus,  $x$  is ultimately periodic if there is a word  $u$  and a periodic infinite word  $y$  such that  $x = uy$ . The following result, due to Coven and Hedlund, is well-known.

### Theorem (Coven, Hedlund, 1973)

*Let  $x \in A^{\mathbb{N}}$  be an infinite word. If there exists an integer  $d \geq 1$  such that  $x$  has at most  $d$  factors of length  $d$  then  $x$  is ultimately periodic.*

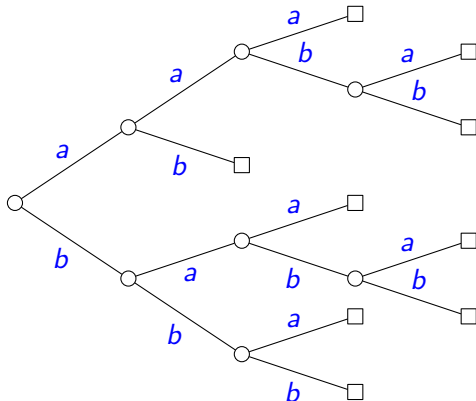
# The Periodicity Theorem

The following statement implies the Coven-Hedlund Theorem since  $A^d$  is a maximal bifix code of degree  $d$ .

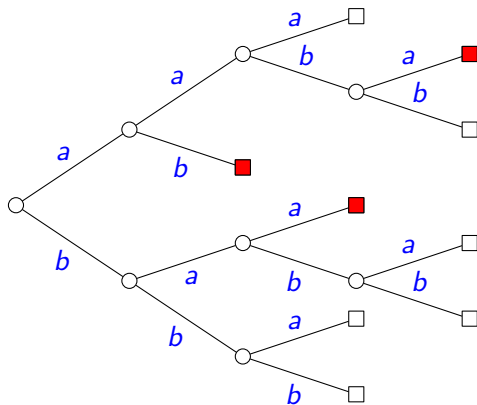
## Theorem (BDRRP, 2010)

*Let  $x \in A^{\mathbb{N}}$  be an infinite word. If there exists a finite maximal bifix code  $X$  of degree  $d$  such that  $\text{Card}(X \cap F(x)) \leq d$ , then  $x$  is ultimately periodic.*

Consider the maximal bifix code of degree 3 below.



Consider the maximal bifix code of degree 3 below.



Assume that  $X \cap F(x)$  is the set of red nodes. Then a factor  $aab$  can only be followed by a second  $aab$ . Thus  $x = u(aab)^\omega$ .

The proof uses the Critical Factorization Theorem that we recall below. For a pair of words  $(p, s) \neq (1, 1)$ , consider the set of nonempty words  $r$  such that

$$A^*p \cap A^*r \neq \emptyset, \quad sA^* \cap rA^* \neq \emptyset.$$

This is the set of nonempty words  $r$  which are prefix-comparable with  $s$  and suffix-comparable with  $p$ . This set is nonempty since it contains  $r = sp$ . The **repetition**  $\text{rep}(p, s)$  is the minimal length of such a nonempty word  $r$ .

Let  $w = a_1a_2 \cdots a_m$  be a word with  $a_i \in A$ . An integer  $n \geq 1$  is a **period** of  $w$  if for  $1 \leq i \leq j \leq m$ ,  $j - i = n$  implies  $a_i = a_j$ . A **factorization** of a word  $w \in A^*$  is a pair  $(p, s)$  of words such that  $w = ps$ .

# The Critical Factorization Theorem

Theorem (Césari, 1978, Duval, 1981)

*For any word  $w \in A^+$ , the maximal value of  $\text{rep}(p, s)$  for all factorizations  $(p, s)$  of  $w$  is the least period of  $w$ .*

# The Sturmian Basis Theorem

## Theorem (BDRRP, 2010)

Let  $F$  be a Sturmian set and let  $d \geq 1$  be an integer. A bifix code  $X \subset F$  is a basis of a subgroup of index  $d$  of the free group on  $A$  if and only if it is a finite  $F$ -maximal bifix code of degree  $d$ .

Note that this theorem contains the Cardinality Theorem. Indeed, by **Schreier's formula**, if  $H$  is a subgroup of rank  $n$  and index  $d$  of a free group of rank  $k$ , then

$$n - 1 = d(k - 1)$$

Let  $X$  be a  $F$ -maximal bifix code of  $F$ -degree  $d$ . By the above theorem, it is a basis of a subgroup of index  $d$  of the free group  $A^\circ$  which has rank  $k$ . Thus  $\text{Card}(X) = (k - 1)d + 1$  by Schreier's Formula.

## Corollary

Let  $F$  be a Sturmian set. For any  $n \geq 1$ , the set  $F \cap A^n$  is a basis of the subgroup of  $A^\circ$  generated by  $A^n$ .

Direct proof : set  $X = F \cap A^d$ . Show by descending induction on  $i = d, \dots, 0$  that for any  $u \in F \cap A^i$ , one has  $uA^{d-i} \subset \langle X \rangle$ . It is true for  $i = d$ . Next consider a right-special word  $u \in F \cap A^i$ . By induction hypothesis, we have  $uaA^{d-i-1} \subset \langle X \rangle$  for any  $a \in A$ . Thus  $uA^{d-i} \subset \langle X \rangle$ . For another  $v \in A^i$ , let  $w$  be such that  $vw \in X$ . Then  $vt = vw(uw)^{-1}ut$  for any  $t \in A^{d-i}$ .

## Example

Let  $F$  be the Fibonacci set. We have  $F \cap A^2 = \{aa, ab, ba\}$  and  $bb = ba(aa)^{-1}ab$ .

# Holonomy groups

Let  $M$  be a monoid of transformations on a set  $Q$ . For  $P \subset Q$ , let

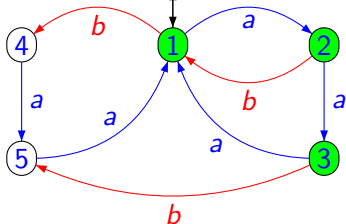
$$\text{Stab}(P) = \{m \in M \mid Pm = P\}$$

be the **stabilizer** of  $P$ . Let  $P$  be of the form  $Qe$  for some idempotent  $e \in M$ . The restriction of  $\text{Stab}(P)$  to  $P$  is a permutation group called the **holonomy group** of  $M$  relative to  $P$  (Eilenberg).

The holonomy groups of an automaton  $\mathcal{A}$  are those of its monoid of transitions  $M(\mathcal{A})$ . A **syntactic group** of a prefix code  $X$  is a holonomy group in the monoid of transitions of  $\mathcal{A}(X^*)$ .

# Example

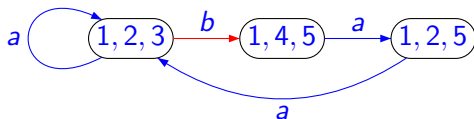
Let  $\mathcal{A}$  be the automaton represented below.



The holonomy group relative to  $\{1, 2, 3\}$  is the symmetric group  $S_3$  generated by  $a = (123)$  and  $baa = (23)$ .

# Computation of the holonomy group

Action on the sets with 3 elements.



# The Rank Conjecture

Let us call **minimal rank** of a group  $G$  the minimal cardinality of a generating set for  $G$ . We will use Sturmian words to discuss the following conjecture (P. 1981).

## Conjecture

*Let  $X$  be a finite bifix code and let  $G$  be a transitive permutation group of degree  $d$  and minimal rank  $k$ . If  $G$  is a syntactic group of  $X$ , then  $\text{Card}(X) \geq (k - 1)d + 1$ .*

The inequality is related to Schreier's Formula.

# The Syntactic Group Theorem

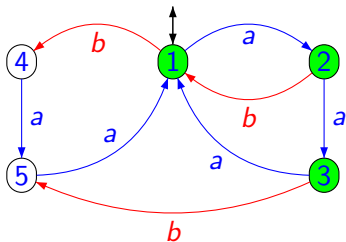
The following result shows that the bound can be reached for any transitive permutation group.

## Theorem (BDPRR, 2010)

*Any transitive permutation group of degree  $d$  which can be generated by  $k$  elements is a syntactic group of a bifix code with  $(k - 1)d + 1$  elements.*

Example 1 : The symmetric group  $S_3$ 

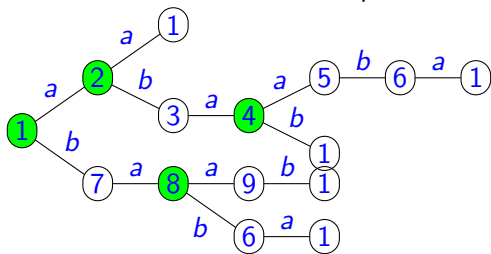
Let  $X = \{aaa, aaba, ab, baa\}$ . The minimal automaton of  $X^*$  is represented below.



The holonomy group relative to  $\{1, 2, 3\}$  is the symmetric group  $S_3$  generated by  $a = (123)$  and  $baa = (23)$ . Such a construction can be used to realize any group generated by a  $d$ -cycle  $\alpha$  and another permutation  $\beta$  using  $X = a^d \cup \{a^i ba^{d-(i+1)\beta} \mid 0 \leq i \leq d-1\}$ .

## Example 2 : The abelian group $(\mathbb{Z}/2\mathbb{Z})^2$

Let  $X$  be the bifix code with 5 elements represented below.



The action on the sets of states with four elements is shown on Figure 3.

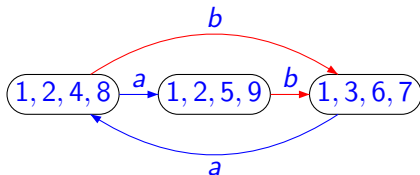


FIG.: The action on the sets of states

The word  $ba$  defines the permutation  $(18)(24)$  and the word  $aba$  the permutation  $(14)(28)$ . Thus  $(\mathbb{Z}/2\mathbb{Z})^2$  is a syntactic group of this code.