

## A QUADRATIC UPPER BOUND ON THE SIZE OF A SYNCHRONIZING WORD IN ONE-CLUSTER AUTOMATA

MARIE-PIERRE BÉAL

*Université Paris-Est  
Laboratoire d'informatique Gaspard-Monge, CNRS  
77454 Marne-la-Vallée Cedex 2, France  
beal@univ-mlv.fr*

MIKHAIL V. BERLINKOV

*Department of Algebra and Discrete Mathematics  
Ural State University  
620083 Ekaterinburg, Russia  
berlm@mail.ru*

DOMINIQUE PERRIN

*Université Paris-Est  
Laboratoire d'informatique Gaspard-Monge, CNRS  
77454 Marne-la-Vallée Cedex 2, France  
perrin@univ-mlv.fr*

Černý's conjecture asserts the existence of a synchronizing word of length at most  $(n - 1)^2$  for any synchronized  $n$ -state deterministic automaton. We prove a quadratic upper bound on the length of a synchronizing word for any synchronized  $n$ -state deterministic automaton satisfying the following additional property: there is a letter  $a$  such that for any pair of states  $p, q$ , one has  $p \cdot a^r = q \cdot a^s$  for some integers  $r, s$  (for a state  $p$  and a word  $w$ , we denote by  $p \cdot w$  the state reached from  $p$  by the path labeled  $w$ ). As a consequence, we show that for any finite synchronized prefix code with an  $n$ -state decoder, there is a synchronizing word of length  $O(n^2)$ . This applies in particular to Huffman codes.

*Keywords:* Černý's Conjecture; Road Coloring Problem; synchronized automata

2010 Mathematics Subject Classification: 68Q45, 68S10, 68R15

### 1. Introduction

Synchronized automata are deterministic and complete finite-state automata admitting a synchronizing word, that is a word which takes each state of the automaton to a single special state. Černý conjecture claims that each  $n$ -state synchronized automaton has a synchronizing word of length at most  $(n - 1)^2$  [9]. An extension of this conjecture due to Pin [16, 18] was shown to be false by Kari [13]. The conjecture has been shown to be true for particular classes of automata like the class of circular automata by Dubuc [10] (see also [17]). The linearisation of automata is used in [19] where a  $2(n - 1)^2$  bound is obtained for regular automata. A  $n(n - 1)/2$

upper bound has been obtained by Trahtman [24, 25] for aperiodic automata. This upper bound was improved to  $n(n+1)/6$  by Volkov [27]. We refer to [28] for a recent beautiful thorough survey on synchronizing automata and Černý's conjecture.

In a previous note [1], the first author gave a proof of a quadratic bound for circular automata which is simpler than the one given in [10]. Nevertheless, it does not allow one to get the tight  $(n-1)^2$  bound. The proof uses rational series.

The formulation of the problem in terms of rational series is also used in [1] to provide a simple proof of a result from Kari [14] which proves Černý's conjecture for automata with an underlying Eulerian graph.

Later, the result on circular automata was generalized by Carpi and d'Alessandro to a larger class called strongly transitive automata [6] and [7], see also [4]. Their proof uses rational series as in [1]. They use the same methods to generalize the result of Kari to unambiguous automata. Carpi and D'Alessandro investigated the synchronization problem for the class locally strongly transitive deterministic automata in [8].

In this paper, we prove the existence of a quadratic upper bound for the length of a synchronizing word for a class of finite automata called one-cluster. This means that, for some letter  $a$ , there is only one cycle with all edges labeled by  $a$ . The proof is an extension of the argument of [1] and uses again rational series. We here slightly improve the upper bound on the size of a shortest synchronizing word obtained in [3] from  $2n^2 - 6n + 5$  to  $2n^2 - 7n + 7$ .

The class of one-cluster automata contains in particular the automata associated with finite prefix codes. We thus obtain the existence of a quadratic bound on the length of a synchronizing word for a finite maximal synchronized prefix code. This applies in particular to Huffman codes.

Let us mention two recent results connected to our work (we thank Hugo Vaccaro for pointing out these references to us). First, it is proved in [12] that almost all finite maximal prefix codes are synchronizing. Next, in [5], it is proved that a finite maximal synchronized prefix code with  $n$  codewords of maximal length  $h$  has a synchronizing word of length  $O(nh \log n)$ . This bound is not comparable with ours. Indeed, since  $\log n \leq h \leq n-1$ , one has  $n(\log n)^2 \leq nh \log n \leq n^2 \log n$ .

A preliminary version of this paper appeared in [3].

## 2. Automata and series

Let  $A$  be a finite alphabet and  $A^*$  be the set of finite words drawn from the alphabet  $A$ , the empty word  $\epsilon$  included. A (finite) *automaton*  $\mathcal{A}$  over some (finite) alphabet  $A$  is composed of a finite set  $Q$  of states and a finite set  $E$  of edges which are triples  $(p, a, q)$  where  $p, q$  are states and  $a$  is a symbol from  $A$  called the *label* of the edge. Note that we do not specify a set of terminal states and that, for this reason, our automata are sometimes called semi-automata.

An automaton is *deterministic* if, for each state  $p$  and each letter  $a$ , there is at most one edge starting in  $p$  and labeled with  $a$ . It is *complete deterministic* if, for

each state  $p$  and each letter  $a$ , there is exactly one edge starting in  $p$  and labeled with  $a$ . This implies that, for each state  $p$  and each word  $w$ , there is exactly one path starting in  $p$  and labeled with  $w$ . We denote by  $p \cdot w$  the state which is the end of this unique path.

An automaton is *irreducible* if its underlying graph is strongly connected.

A *synchronizing word* for a deterministic complete automaton is a word  $w$  such that for any states  $p, q$ , one has  $p \cdot w = q \cdot w$ . A synchronizing word is also called a *reset sequence* or a *magic sequence*, or also a *homing word*. An automaton which has a synchronizing word is called *synchronized* (see an example on the right part of Figure 1).



Fig. 1. Two complete deterministic automata labeled on  $A = \{a, b\}$ . A thick plain edge is an edge labeled by  $a$  while a thin dashed edge is an edge labeled by  $b$ . The automaton on the left is not synchronized. The one on the right is synchronized; for instance, the word  $aaa$  is a synchronizing word.

Let  $\mathcal{A} = (Q, E)$  be a complete deterministic automaton. For any word  $u \in A^*$ , we denote by  $M_u$  the transition matrix of the action of  $u$  on the states  $Q$ . It is defined by:

$$(M_u)_{pq} = \begin{cases} 1 & \text{if } p \cdot u = q, \\ 0 & \text{otherwise.} \end{cases}$$

Note that if  $u, v$  are two words, we have

$$M_{uv} = M_u M_v.$$

We define the *rank* of a word  $u$  as the cardinality of  $Q \cdot u$ . Note that since the automaton is complete deterministic, this rank is non null, and that a word is synchronizing if and only if his rank is 1.

A *circular* automaton is a deterministic complete automaton on the alphabet  $A$  such that there is a letter  $a$  of  $A$  which induces a circular permutation of the states, *i.e.* such that  $M_a$  is a circular permutation matrix.

We shall consider the set of non commutative formal series with coefficients in a ring  $K$  (with  $K = \mathbb{Z}$  or  $K = \mathbb{Q}$ ), which are applications from  $A^*$  to  $K$ . If  $S$  is such a series, the image of a word  $u$  of  $A^*$  by  $S$  is denoted by  $\langle S, u \rangle$  and called the coefficient of  $u$  in  $S$ .

As an example, the series  $S$  on  $\{a, b\}^*$  with coefficients in  $\mathbb{Z}$  defined by  $\langle S, u \rangle = |u|_a - |u|_b$  maps a word  $u \in \{a, b\}^*$  to the difference between the number of occurrences of  $a$  and  $b$  in  $u$ .

A  $K$ -linear representation of dimension  $d$  of a series  $S$  is a triple  $(\lambda, \mu, \gamma)$  where  $\lambda \in K^{1 \times d}$ ,  $\mu$  is a morphism from  $A^*$  to  $K^{d \times d}$ , and  $\gamma \in K^{d \times 1}$ , such that

$$\langle S, u \rangle = \lambda \mu(u) \gamma.$$

A series  $s$  is  $K$ -rational if it has a  $K$ -linear representation. Its *rank* on  $K$  is the minimal dimension of all its linear representations.

For example, the series  $S$  defined on  $\{a, b\}^*$  by  $\langle S, u \rangle = |u|_a - |u|_b$  is rational of rank 2. The triple  $(\lambda, \mu, \gamma)$  defined by

$$\lambda = [1 \ 0], \mu(a) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \mu(b) = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}, \gamma = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

is a  $\mathbb{Z}$ -linear representation of  $S$  of dimension 2.

Černý's conjecture gives an upper bound on the size of a shortest synchronizing word in a synchronized automaton.

**Conjecture 1 (Černý 1964)** *A synchronized  $n$ -state deterministic complete automaton has a synchronizing word of length at most  $(n - 1)^2$ .*

The conjecture was proved by Dubuc for circular automata.

**Proposition 1 (Dubuc 1998)** *A circular synchronized  $n$ -state deterministic complete automaton has a synchronizing word of size at most  $(n - 1)^2$ .*

### 3. One-cluster automata

In the sequel  $\mathcal{A} = (Q, E)$  denotes an  $n$ -state deterministic and complete automaton over an alphabet  $A$ . We fix a particular letter  $a \in A$ .

Let  $\mathcal{R}$  be the subgraph of the graph of  $\mathcal{A}$  made of the edges labeled by  $a$ . The graph  $\mathcal{R}$  is a disjoint union of connected component called *a-clusters*. Since each state has exactly one outgoing edge in  $\mathcal{R}$ , each  $a$ -cluster contains a unique cycle, called an *a-cycle*, with trees attached to the cycle at their root. For each state  $p$  of an  $a$ -cluster, we define the *level* of  $p$  as the distance between  $p$  and the root of the tree containing  $p$ . If  $p$  belongs to the cycle, its level is null. The *level* of the automaton is the maximal level of its states.

A *one-cluster automaton* with respect to a letter  $a$  is a complete deterministic automaton which has only one  $a$ -cluster. Equivalently, an automaton is one-cluster if it satisfies the following condition: for any pair of states  $p, q$ , one has  $p \cdot a^r = q \cdot a^s$  for some integers  $r, s$ .

Note that a one-cluster automaton whose level is null is circular.

Let  $C$  be a cycle of  $\mathcal{A}$  and  $P$  be a subset of  $C$ . A word  $u$  is said to be *P-augmenting* if

$$\text{card}(Pu^{-1} \cap C) > \text{card}(P),$$

where we denote  $Pu^{-1} = \{q \in Q \mid q \cdot u \in P\}$ .

We now prove the existence of a quadratic upper bound on the size of a shortest synchronizing word in a synchronized automaton.

Let  $\mathcal{A}$  be a synchronized  $n$ -state deterministic complete automaton. If  $\mathcal{A}$  is one-cluster, then it has a synchronizing word of length at most  $2n^2 - 7n + 7$ .

We prove the proposition for irreducible automata. The case of reducible automata easily reduces to this one. Let  $\mathcal{A} = (Q, E)$  be a deterministic complete and irreducible  $n$ -state automaton.

Since Černý's conjecture is proved for circular automata, we may assume that the level  $\ell$  of the automaton is greater than or equal to 1.

Let  $C$  be the  $a$ -cycle and let  $m$  be the length of  $C$ . Let  $P$  be a subset of  $C$ . Note that a word  $u$  is a  $P$ -augmenting word if and only if

$$C M_u P^t > C P^t,$$

where  $P, C$  denote the characteristic row vectors of the sets  $P, C$ . Indeed,

$$C M_u P^t = \sum_{r \in C, s \in P} (M_u)_{rs} = \text{card}\{r \in C \mid r \cdot u \in P\} = \text{card}(P u^{-1} \cap C).$$

Similarly,  $C P^t = \text{card}(P)$ .

We denote by  $S$  the series defined by:

$$\langle S, u \rangle = C M_u P^t - C P^t,$$

By definition, one has  $\langle S, u \rangle > 0$  if and only if  $u$  is a  $P$ -augmenting word.

**Lemma 3.** *The series  $S$  has rank on  $\mathbb{Q}$  at most  $n$ .*

**Proof.** The series  $S$  is  $\mathbb{Z}$ -rational since it is the difference of two  $\mathbb{Z}$ -rational series (the second one is actually a constant). It has the following linear representation  $(\lambda, \mu, \gamma)$  with  $\lambda \in \mathbb{Z}^{1 \times 2n}$ ,  $\mu : A^* \rightarrow \mathbb{Z}^{2n \times 2n}$ ,  $\gamma \in \mathbb{Z}^{2n \times 1}$ ,

$$\lambda = (C, -C), \mu(u) = \begin{bmatrix} M_u & 0 \\ 0 & I \end{bmatrix}, \gamma = \begin{bmatrix} P^t \\ P^t \end{bmatrix},$$

since  $\langle S, u \rangle = \lambda \mu(u) \gamma$ . The rank of  $S$  on  $\mathbb{Q}$  is bounded above by the dimension of the  $\mathbb{Q}$ -vector space generated by the row vectors  $(C M_u, -C)$ . This space is included in the vector space generated by the vectors  $(C M_u - C, \mathbf{0})$  and the row vector  $(C, -C)$ , where  $\mathbf{0}$  is the null column vector of size  $n$ . Thus the rank of  $S$  is at most equal to the dimension of the vector space  $V$  generated by the vectors  $C(M_u - I)$ , plus one. We now show that the dimension of  $V$  is at most  $n - 1$ . Since the automaton  $\mathcal{A}$  is complete deterministic, for any  $u \in A^*$ ,  $M_u \mathbf{1} = \mathbf{1}$ , where  $\mathbf{1}$  is the column vector with coefficients 1 of size  $n$ . This implies that  $C(M_u - I) \cdot \mathbf{1} = 0$ . Thus the vectors of  $V$  are orthogonal to the vector  $\mathbf{1}$ . The dimension of  $V$  is thus at most  $n - 1$ . This proves that the rank of  $S$  on  $\mathbb{Q}$  is at most  $n$ .  $\square$

We denote by  $T$  the  $\mathbb{Z}$ -rational series defined by

$$\langle T, u \rangle = \langle S, u a^\ell \rangle,$$

where  $\ell$  denotes the level of the automaton  $\mathcal{A}$ . If  $(\lambda, \mu, \gamma)$  is a  $\mathbb{Q}$ -linear representation of  $S$  of dimension  $n$ , then  $(\lambda, \mu, \mu(a^\ell)\gamma)$  is a representation of  $T$ . Thus the rank of  $T$  on  $\mathbb{Q}$  is at most  $n$ .

**Lemma 4.** *For any subset  $P$  of  $C$  such that  $P \neq \emptyset$  and  $P \neq C$ , there is a  $P$ -augmenting word of length at most  $2(n-1)$ .*

**Proof.** Since  $\mathcal{A}$  is synchronized and irreducible, there is a synchronizing word  $u$  such that  $Q \cdot u$  is a single state  $r$  belonging to  $P$ . Let  $k$  be a positive integer such that  $km \geq \ell$ , where  $m$  is the length of the cycle  $C$ . We have  $Q \cdot ua^{km-\ell}a^\ell = r \cdot a^{km} = r$ . Hence  $ua^{km-\ell}a^\ell$  also is a synchronizing word focusing to  $r$ . Let  $R$  denote the characteristic row vector of  $r$ . Since  $q \cdot u = r$  for all  $q \in Q$  and since  $C$  has  $m$  elements, we have  $CM_u = mR$ . Moreover, since  $r \in C$ ,  $RM_{a^m} = R$ . We have

$$\begin{aligned} \langle T, ua^{km-\ell} \rangle &= \langle S, ua^{km} \rangle \\ &= CM_{ua^{km}} P^t - CP^t \\ &= CM_u M_{a^{km}} P^t - CP^t \\ &= mRM_{a^{km}} P^t - CP^t \\ &= mRP^t - CP^t \\ &= m - \text{card}(P) \neq 0. \end{aligned}$$

As a consequence  $T$  is non null.

Since  $T$  has rank at most  $n$  on  $\mathbb{Q}$ , there is a word  $v$  of length at most  $n-1$  such that  $\langle T, v \rangle \neq 0$  (see [11], [20] or [21]).

If there is word  $v$  of length at most  $n-1$  such that  $\langle T, v \rangle > 0$ , then  $va^\ell$  is a  $P$ -augmenting word and the claim is proved. Otherwise, there is a word  $v$  of length at most  $n-1$  such that  $\langle T, v \rangle < 0$ .

Since  $\ell$  is the level of  $\mathcal{A}$ , the vector  $CM_{va^\ell}$  is a linear combination of elements of  $C$  and the sum of its coefficients is equal to  $m$ .

We have

$$\begin{aligned} \sum_{i=0}^{m-1} \langle T, va^i \rangle &= \sum_{i=0}^{m-1} \langle S, va^\ell a^i \rangle = \sum_{i=0}^{m-1} C(M_{va^\ell a^i} - I)P^t, \\ &= \left( \sum_{i=0}^{m-1} CM_{va^\ell} M_{a^i} - \sum_{i=0}^{m-1} C \right) P^t, \\ &= (CM_{va^\ell} \left( \sum_{i=0}^{m-1} M_{a^i} \right) - mC) P^t, \\ &= \left( \sum_{r \in C} r M_{va^\ell} \sum_{i=0}^{m-1} M_{a^i} - mC \right) P^t, \\ &= \left( \sum_{r \in C} C - mC \right) P^t = 0. \end{aligned}$$

Indeed, for any  $r$  in  $C$ , the state  $q = r \cdot va^\ell$  is in  $C$  and for any state  $q$  in  $C$ , the row of index  $q$  of the matrix  $\sum_{i=0}^{m-1} M_{a^i}$  is the row vector  $C$ .

As a consequence, there is a word  $w$  of length at most  $n + m - 2$  such that  $\langle T, w \rangle > 0$ . Hence there is a  $P$ -augmenting word of length at most  $n + m - 2 + \ell$ .

Thus, in all cases, there is a word of length at most  $n + m - 2 + \ell$  which is  $P$ -augmenting.  $\square$

To prove Proposition 2, we show that  $\mathcal{A}$  has a synchronizing word of length at most  $1 + 2m(n - 2)$ . Indeed, let  $P_1$  be reduced to an arbitrary state of  $C$ . If  $P_1 = C$  (that is to say if  $m = 1$ ), then  $Q \cdot a^\ell \subseteq P_1$ , and thus  $a^\ell$  is a synchronizing word.

Otherwise, by Lemma 4, there exists a word  $u_1$  of length at most  $n + m - 2 + \ell$  which is  $P_1$ -augmenting. Set  $P_2 = P_1 u_1^{-1} \cap C$ . If  $P_2 \neq C$ , there is a word  $u_2$  of length at most  $n + m - 2 + \ell$  which is  $P_2$ -augmenting, and so on. In this way, we build a sequence  $u_1, \dots, u_{t-1}$  of words and a sequence  $P_1, \dots, P_t$  of sets of states, with  $t \leq m$ , such that, for  $1 \leq i < t$ ,

- $u_i$  is a  $P_i$ -augmenting word of length at most  $n + m - 2 + \ell$ ;
- $P_{i+1} = P_i u_i^{-1} \cap C$ ;
- $P_t = C$ .

Then the word  $a^\ell u_{t-1} \dots u_1$  is a synchronizing word of length at most  $\ell + (m - 1)(n + m - 2 + \ell)$ . Indeed,  $Q \cdot a^\ell u_{t-1} \dots u_1 \subseteq C \cdot u_{t-1} \dots u_1 \subseteq P_{t-1} \cdot u_{t-2} \dots u_1 \subseteq \dots \subseteq P_2 \cdot u_1 \subseteq P_1$ .

Since  $m \leq n - \ell$  and  $m \leq n - 1$ , we have

$$\begin{aligned} \ell + (m - 1)(n + m + \ell - 2) &\leq \ell + (m - 1)(2n - 2) \\ &\leq n - m + 2mn - 2n - 2m + 2 \\ &\leq 2mn - n - 3m + 2 \\ &= (n - 2)(2m - 1) + m \\ &\leq 1 + 2m(n - 2) \\ &\leq 1 + 2(n - 1)(n - 2) = 2n^2 - 6n + 5. \end{aligned}$$

We slightly improve this bound by giving a better upper bound on the size of the first augmenting word  $u_1$ .

Let  $q$  be a state of  $C$ . We denote by  $R_q$  the  $\mathbb{Z}$ -rational series defined by

$$\langle R_q, u \rangle = \langle S_q, ua^{(n-m)} \rangle,$$

where  $\langle S_q, u \rangle = C M_u q^t - C q^t$  and  $q$  is the row-characteristic vector of the state  $q$ . The series  $R_q$  has rank at most  $n$ . One can show, as for  $T$ , that  $R_q$  is non null.

If there a state  $q$  of  $C$  and a word  $u$  of length at most  $n - 1$  such that  $\langle R_q, u \rangle > 0$ , then  $ua^{(n-m)}$  is a  $\{q\}$ -augmenting word of length  $2n - m - 1$ . Otherwise,  $\langle R_q, u \rangle \leq 0$  for any word  $u$  of length at most  $n - 1$  and any state  $q$  of  $C$ .

Let  $u \in A^*$  of length at most  $n - 1$ .

$$\begin{aligned}
 \sum_{q \in C} \langle R_q, u \rangle &= \sum_{q \in C} \langle S_q, ua^{(n-m)} \rangle \\
 &= \sum_{q \in C} (CM_u M_{a^{(n-m)}} q^t - Cq^t) \\
 &= CM_u M_{a^{(n-m)}} C^t - C \cdot C^t \\
 &= C \cdot C^t - C \cdot C^t = 0.
 \end{aligned}$$

The last but one equality results from  $n - m \geq l$  which implies  $CM_u M_{a^{(n-m)}} = C$ .

As a consequence,  $\langle R_q, u \rangle = 0$  for any state  $q$  of  $C$ . Since  $R_q$  has rank at most  $n$ ,  $R_q$  is null, a contradiction. Thus there a state  $q$  of  $C$  and a word  $u$  of length at most  $n - 1$  such  $ua^{(n-m)}$  is a  $\{q\}$ -augmenting word of length  $2n - m - 1$ . Thus we get a  $\{q\}$ -augmenting word of length at most  $2n - m - 1$ .

With this improvement, we obtain a synchronizing word of length at most  $\ell + (2n - m - 1) + (m - 2)(n + m - 2 + \ell)$ . We have

$$\begin{aligned}
 &\ell + (2n - m - 1) + (m - 2)(n + m - 2 + \ell) \\
 &\leq n - m + 2n - m - 1 + (m - 2)(2n - 2) \\
 &\leq n - m + 2n - m - 1 + 2nm - 4n - 2m + 4 \\
 &\leq 2mn - n - 4m + 3 \\
 &= m(2n - 4) - n + 3 \\
 &\leq (n - 1)(2n - 4) - n + 4 \\
 &\leq 2n^2 - 7n + 7,
 \end{aligned}$$

which completes the proof.

#### 4. Application to finite prefix codes

In this section we show how the previous result can be applied to the automaton associated to a finite prefix code.

A *prefix code* on the alphabet  $A$  is a set  $X$  of words on  $A$  such that no element of  $X$  is a prefix of another word of  $X$ .

A prefix code is *maximal* if it is not contained in another prefix code on the same alphabet. As an equivalent definition, a prefix code  $X$  is maximal if for any word  $u$  in  $A^*$  has a prefix in  $X$  or is a prefix of a word of  $X$ .

For a deterministic automaton  $\mathcal{A}$  and an initial state  $i$ , the set  $X_{\mathcal{A}}$  of labels of first return paths from  $i$  to  $i$  is a prefix code. If the automaton is complete, the prefix code is maximal.

Conversely, for any finite prefix code  $X$ , there exists a deterministic automaton  $\mathcal{A}$  such that  $X = X_{\mathcal{A}}$ . Moreover, the automaton  $\mathcal{A}$  can be supposed to be irreducible. If  $X$  is a maximal prefix code, the automaton  $\mathcal{A}$  is complete.

The automaton  $\mathcal{A}$  can be chosen as follows. The set of states is the set  $Q$  of prefixes of the words of  $X$ . The transitions are defined for  $p \in Q$  and  $a \in A$  by  $p \cdot a = pa$  if  $pa$  is a prefix of a word of  $X$ , and by  $p \cdot a = \varepsilon$  if  $pa \notin X$ . This automaton, denoted  $\mathcal{A}_X$  is a *decoder* of  $X$ . Let indeed  $\alpha$  be a one-to-one map from a source alphabet  $B$  onto  $X$ . Let us add an output label to each edge of  $\mathcal{A}_X$  in the following way. The output label of  $(p, a, q)$  is  $\varepsilon$  if  $q \neq \varepsilon$  and is equal to  $\alpha^{-1}(pa)$  if  $q = \varepsilon$ . With this definition, for any word  $x \in X^*$ , the output label of the path  $i \xrightarrow{x} i$  is the word  $\alpha^{-1}(x)$ .

Let us show that, as a consequence of the fact that  $X$  is finite, the automaton  $\mathcal{A}$  is additionally one-cluster with respect to any letter.

Indeed, let  $a$  be a letter and let  $C$  be the set of states of the form  $i \cdot a^j$ . For any state  $q$ , there exists a path  $i \xrightarrow{u} q \xrightarrow{v} i$ . We may suppose that  $i$  does not occur elsewhere on this path. Thus  $uv \in X$ . Since  $X$  is a finite maximal code, there is an integer  $j$  such that  $ua^j \in X$ . Then  $q \cdot a^j = i$  belongs to  $C$ . This shows that  $\mathcal{A}$  is one-cluster with respect to  $a$ .

A maximal prefix code  $X$  is *synchronized* if there is a word  $x \in X^*$  such that for any word  $w \in A^*$ , one has  $wx \in X^*$ . Such a word  $x$  is called a *synchronizing word* for  $X$ .

Let  $X$  be a synchronized prefix code. Let  $\mathcal{A}$  be an irreducible deterministic automaton with an initial state  $i$  such that  $X_{\mathcal{A}} = X$ . The automaton  $\mathcal{A}$  is synchronized. Indeed, let  $x$  be a synchronizing word for  $X$ . Let  $q$  be a state of  $\mathcal{A}$ . Since  $\mathcal{A}$  is irreducible, there is a path  $i \xrightarrow{u} q$  for some  $u \in A^*$ . Since  $x$  is synchronizing for  $X$ , we have  $ux \in X^*$ , and thus  $q \cdot x = i$ . This shows that  $x$  is a synchronizing word for  $\mathcal{A}$ .

Conversely, let  $\mathcal{A}$  be an irreducible complete deterministic automaton. If  $\mathcal{A}$  is a synchronized automaton, the prefix code  $X_{\mathcal{A}}$  is synchronized. Indeed, let  $x$  be a synchronizing word for  $\mathcal{A}$ . We may assume that  $q \cdot x = i$  for any state  $q$ . Then  $x$  is a synchronizing word for  $X$ .

**Proposition 5.** *Let  $X$  be a maximal synchronized prefix code with  $n$  codewords on an alphabet of size  $k$ . The decoder of  $X$  has a synchronizing word of length at most  $O((\frac{n}{k})^2)$ .*

**Proof.** The automaton  $\mathcal{A}_X$  is one-cluster. The number  $N$  of its states is the number of prefixes of the words of  $X$ . Thus  $N = (n - 1)/(k - 1)$  since a complete  $k$ -ary tree with  $n$  leaves has  $(n - 1)/(k - 1)$  internal nodes. By Proposition 2, there is a synchronizing word of length at most  $1 + 2(N - 1)(N - 2)$ , whence  $O((\frac{n}{k})^2)$ .  $\square$

**Example 6.** *Let us consider the following Huffman code  $X = (00 + 01 + 1)(0 + 10 + 11)$  corresponding to a source alphabet  $B = \{a, b, c, d, e, f, g, h, i\}$  with a probability distribution  $(1/16, 1/16, 1/8, 1/16, 1/16, 1/8, 1/8, 1/8, 1/4)$ . The Huffman tree is pictured in the left part of Figure 6 while the decoder automaton  $\mathcal{A}_X$  is given in its right part. The word 010 is a synchronizing word of  $\mathcal{A}_X$ .*

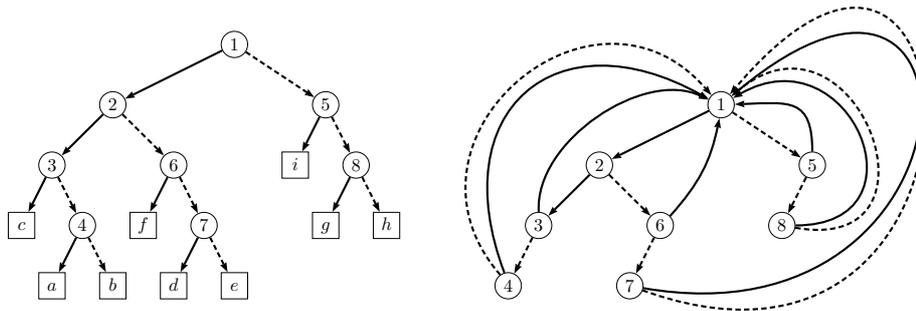


Fig. 2. A synchronized Huffman code  $X$  on the left and its decoder  $\mathcal{A}_X$  on the right.

When the lengths of the codewords in  $X$  are not relatively prime, the automaton  $\mathcal{A}_X$  is never synchronized (see Example of Figure 6). When the lengths of the codewords in  $X$  are relatively prime, the code  $X$  is not necessarily synchronized. However, there is always another Huffman code  $X'$  corresponding to the same length distribution which is synchronized by a result of Schützenberger [22]. One can even choose  $X'$  such that the underlying graph of  $\mathcal{A}_X$  and  $\mathcal{A}_{X'}$  are the same. This is a particular case of the road coloring theorem of due to Trahtman [26] (see also [2]). The particular case corresponding to finite prefix codes was proved before in [15].

Our result guarantees that the Huffman decoder has a synchronizing word of length at most quadratic in the number of nodes of the Huffman tree.

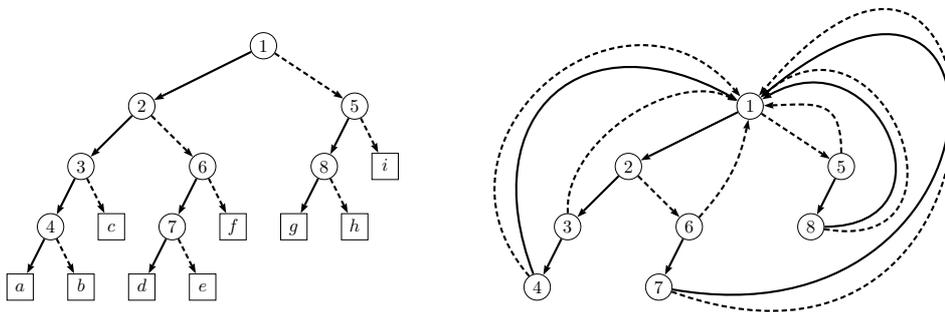


Fig. 3. A non synchronized Huffman code  $X$  on the left and its decoder on the right. The automaton on the right is not synchronized. Indeed, for any word  $w$ , the set of states reachable by  $w$  is either  $\{1, 3\}$ ,  $\{2, 4\}$ ,  $\{1, 5\}$ , or  $\{1, 6\}$ .

**Notes<sup>a</sup>**

The present work recently inspired follow-up results. Recently, Steinberg showed how to recover our result and the bounds from [19], [7],[8] using probability arguments. He also obtained the Černý bound for one-cluster automata with prime length cycle in [23].

**References**

- [1] M.-P. BÉAL, *A note on Černý's conjecture and rational series*. preprint IGM 2003-05. Unpublished, 2003.
- [2] M.-P. BÉAL AND D. PERRIN, *A quadratic algorithm for road coloring*, CoRR, abs/0803.0726 (2008).
- [3] ———, *A quadratic upper bound on the size of a synchronizing word in one-cluster automata*, in *Developments in Language Theory, 2009*, pp. 81–90.
- [4] M. V. BERLINKOV, *On Carpi and D'Alessandro conjecture*, CoRR, abs/0909.3790 (2009).
- [5] M. T. BISKUP, *Shortest synchronizing strings for Huffman codes*, in *MFCS, 2008*, pp. 120–131.
- [6] A. CARPI AND F. D'ALESSANDRO, *The synchronization problem for strongly transitive automata*, in *Developments in Language Theory, 2008*, pp. 240–251.
- [7] ———, *Strongly transitive automata and the Černý conjecture*, *Acta Inf.*, 46 (2009), pp. 591–607.
- [8] ———, *The synchronization problem for locally strongly transitive automata*, in *MFCS, 2009*, pp. 211–222.
- [9] J. ČERNÝ, *Poznámka k. homogénnym experimentom s konečnými automatmi*, *Mat. fyz. čas SAV*, 14 (1964), pp. 208–215.
- [10] L. DUBUC, *Sur les automates circulaires et la conjecture de Černý*, *RAIRO Inform. Théor. Appl.*, 32 (1998), pp. 21–34.
- [11] S. EILENBERG, *Automata, languages, and machines. Vol. A*, Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York, 1974. *Pure and Applied Mathematics*, Vol. 58.
- [12] C. F. FREILING, D. S. JUNGREIS, F. THÉBERGE, AND K. ZEGER, *Almost all complete binary prefix codes have a self-synchronizing string*, *IEEE Transactions on Information Theory*, 49 (2003), pp. 2219–2225.
- [13] J. KARI, *A counter example to a conjecture concerning synchronizing words in finite automata*, *EATCS Bulletin*, 73 (2001), p. 146.
- [14] ———, *Synchronizing finite automata on Eulerian digraphs*, in *MFCS'2001*, no. 2136 in LNCS, Berlin, 2001, Springer, pp. 432–438.
- [15] D. PERRIN AND M.-P. SCHÜTZENBERGER, *Synchronizing words and automata and the road coloring problem*, in *Symbolic Dynamics and its Applications*, P. Walters, ed., American Mathematical Society, 1992, pp. 295–318. *Contemporary Mathematics*, vol. 135.
- [16] J.-E. PIN, *Le problème de la synchronisation et la conjecture de Černý*, thèse de 3ème cycle, Université Paris VI, 1978.
- [17] ———, *Sur un cas particulier de la conjecture de Černý*, in *5th ICALP*, no. 62 in LNCS, Berlin, 1978, Springer, pp. 345–352.

<sup>a</sup>Notes added for the final version after the reviewing process

- [18] ———, *On two combinatorial problems arising from automata theory*, *Annals of Discrete Mathematics*, 17 (1983), pp. 535–548.
- [19] I. K. RYSTSOV, *Quasioptimal bound for the length of reset words for regular automata*, *Acta Cybern.*, 12 (1995), pp. 145–152.
- [20] J. SAKAROVITCH, *Éléments de théorie des automates*, Éditions Vuibert, 2003.
- [21] J. SAKAROVITCH, *Elements of automata theory*, Cambridge University Press, Cambridge, 2009. Translated from the 2003 French original by Reuben Thomas.
- [22] M.-P. SCHÜTZENBERGER, *On synchronizing prefix codes*, *Inform. and Control*, 11 (1967), pp. 396–401.
- [23] B. STEINBERG, *The Černý conjecture for one-cluster automata with prime length cycle*, CoRR, abs/0910.0410 (2009).
- [24] A. N. TRAHMAN, *Synchronization of some DFA*, in TAMC, 2007, pp. 234–243.
- [25] A. N. TRAHMAN, *Some aspects of synchronization of DFA*, *J. Comput. Sci. Technol.*, 23 (2008), pp. 719–727.
- [26] A. N. TRAHMAN, *The road coloring problem*, *Israel J. Math.*, 172 (2009), pp. 51–60.
- [27] M. V. VOLKOV, *Synchronizing automata preserving a chain of partial orders*, in CIAA, 2007, pp. 27–37.
- [28] ———, *Synchronizing automata and the Černý conjecture*, in LATA, 2008, pp. 11–27.