Synchronizing Automata with Large Reset Lengths

Andrzej Kisielewicz* and Marek Szykuła**

Department of Mathematics and Computer Science, University of Wrocław
andrzej.kisielewicz@math.uni.wroc.pl, msz@ii.uni.wroc.pl

Abstract. We study synchronizing automata with the shortest reset words of relatively large length. First, we refine the Frankl-Pin result on the length of the shortest words of rank $m$, and the Béal, Berlinkov, Perrin, and Steinberg results on the length of the shortest reset words in one-cluster automata. The obtained results are applied to computation aimed in extending the class of small automata for which the Černý conjecture is verified and discovering new automata with special properties regarding synchronization. In particular, a new class of slowly synchronizing automata on a ternary alphabet is constructed and a conjecture on $cn$-extendable sets in synchronizing automata is disproved.

1 Introduction

We deal with deterministic finite automata $A = \langle Q, \Sigma, \delta \rangle$, where $Q$ is the set of the states, $\Sigma$ is the input alphabet, and $\delta : Q \times \Sigma \rightarrow Q$ is the (complete) transition function. The cardinality $n = |Q|$ is the size of $A$, and if $k = |\Sigma|$ then $A$ is called $k$-ary. The rank of a word $w \in \Sigma'$ is $|Qw|$, and the rank of $A$ is the minimal rank of a word over $A$. For a nonempty subset $\Sigma' \subseteq \Sigma$, we may define the automaton $A' = \langle Q, \Sigma', \delta' \rangle$, where $\delta'$ is the natural restriction of $\delta$ to $\Sigma'$. In such a case $A$ is called an extension of $A'$. The automata of rank 1 are called synchronizing, and each word $w$ with $|Qw| = 1$ is called a synchronizing (or reset) word for $A$.

We are interested in the length of a shortest reset word for $A$ (there may be more than one word of the same shortest length). We call it the reset length of $A$. The famous Černý conjecture states that every synchronizing automaton $A$ with $n$ states has a reset word of length $\leq (n - 1)^2$. This conjecture was formulated by Černý in 1964 [9], and is considered the longest-standing open problem in combinatorial theory of finite automata. So far, the conjecture has been proved only for a few special classes of automata and a cubic upper bound has been established (see Volkov [23] for an excellent survey). In [18] we have verified the conjecture for all binary automata with $n < 12$ states.

In this paper we prove some new results improving known bounds and extending the class of automata for which the Černý conjecture is verified. In particular, we strengthen the Frankl-Pin result on the length of the shortest words of rank $m$, and the Béal, Berlinkov, Perrin [6,5], and Steinberg [21] results on one-cluster automata. These are refinements of a rather technical nature. The motivation for these refinements is to make computations in this area more effective. This allows to extend the studies reported in [22,23,18]. In particular, we search for synchronizing automata with relatively large reset length. We improve the algorithm from [18] which takes a set of $(k - 1)$-ary automata with $n$ states and generates all their nonisomorphic one-letter extensions. To perform an exhaustive search over the $k$-ary automata with $n$ states with some property, we need to

* Supported in part by Polish MNiSZW grant IP2012 052272.
** Supported in part by Polish NCN grant DEC-2013/09/N/ST6/01194.
progressively run the algorithm \( k - 1 \) times starting from the complete set of non-isomorphic unary automata. However, in each run, if we know that any extension of an automaton \( \mathcal{A} \) cannot have the desired property, we can safely drop \( \mathcal{A} \) from further computations. Since the number of generated automata grows rapidly, suitable knowledge saves a lot of computational time and extends the class of the automata investigated. In this study, we concentrate on automata of arity \( k > 2 \).

2 Theoretical Base for Computation

Through the paper, if not stated otherwise, \( \mathcal{A} \) denotes a (deterministic, finite) automaton, \( \Sigma \) its alphabet, \( Q \) its set of states, and \( n = |Q| \) its size. Words are always the words over \( \Sigma \) (that is, elements of \( \Sigma^* \)). A word \( w \) is said to compress a set \( M \subseteq Q \) if \(|Mw| < |M|\). In such a case \( M \) itself is called compressible.

2.1 Frankl-Pin Sequences

Suppose that \( M \subseteq Q \) and \( u = a_1 \ldots a_{\ell} \) is a shortest word compressing \( M \), that is, such that \(|Mu| < |M|\). Let \( M_i = M_{a_1} \ldots a_{i-1} \) for \( 1 \leq i \leq \ell + 1 \). Then, there are \( x_i, y_i \in M_i \) such that \( x_i a_{\ell} = y_i a_i \). For \( \ell > i \geq 1 \), we define \( x_i, y_i \in M_i \) by \( x_i a_i = x_{i+1} \) and \( y_i a_i = y_{i+1} \). This defines \( x_i \) and \( y_i \) uniquely, since otherwise \( u \) would not be a shortest word compressing \( M \). By the same reason the subsets \( M_i \) are of the same cardinality for \( i \leq \ell \), and together with the pairs \( R_i = \{ x_i, y_i \} \) they satisfy the following conditions:

1. \( R_i \subseteq M_i \) for \( 1 \leq i \leq \ell \);
2. \( R_i \not\subseteq M_j \) for \( 1 \leq j < i \leq \ell \).

In [20], J.-E. Pin used this observation to bound the length of a word \( u \) compressing \( M \). He suggested a certain combinatorial estimation that was proved subsequently by Frankl. (We quote the result in a restricted form sufficient for our aims).

**Theorem 1.** (P. Frankl [11]) Let \( Q \) be an \( n \)-element subset, \( M_1, \ldots, M_\ell \) be a sequence of its \( m \)-subsets (for some \( 1 < m < n \)), and \( R_1, \ldots, R_\ell \) be a sequence of pairs contained in \( Q \). If the conditions 1 and 2 above are satisfied, then

\[
\ell \leq \binom{n - m + 2}{2}.
\]

We say that a sequence \( (M_i, R_i), (1 \leq i \leq \ell) \) of \( m \)-subsets \( M_i \) and pairs \( R_i \) satisfying conditions 1 and 2 is an \( m \)-subset Frankl-Pin sequence. If all the pairs \( R_i \) belong to a set \( P \) of pairs, we will say that this sequence is over \( P \). From what we said it follows that a shortest word compressing \( M \) cannot be longer than the length of the Frankl-Pin sequence starting from \( M \). Hence summing up the binomial coefficient in Theorem 1 we obtain, in particular, the bound \((n^3 - n)/6\) for the length of a shortest reset word. In spite of many efforts to improve it, it is still the best bound known in the literature.

In order to improve it, we introduce the following technical notions. Let \( P \) be an arbitrary set of compressible pairs in \( \mathcal{A} \). By a synchronizing height \( h(P) \) of \( P \) we mean the minimal \( h \) such that for each pair \( \{ x, y \} \in P \) there exists a word \( w \) of length \( h \) such that \( xw = yw \). We make use of the observation that if the synchronizing height is smaller than the maximal length of a Frankl-Pin sequence over \( P \), then we can improve the Pin’s estimation from [20].
Proposition 1. Let $P$ be a set of compressible pairs in $A$, $h(P)$ the synchronizing height of $P$, and $p(P)$ the maximal length of a Frankl-Pin sequence over $P$. Then, for every compressible $m$-subset $M$ of $Q$ ($2 \leq m \leq n$), there is a word compressing $M$ whose length does not exceed

$$\left(\frac{n - m + 2}{2}\right) - p(P) + h(P).$$

Proof. (In the first part of the proof we modify the Pin’s argument mentioned above; see [20, Proposition 3.1]). Let $u = a_1 \ldots a_t$ be a shortest word such that either $|Mu| < |M|$ or $\{x_k, y_k\} \in Mu$ for some $\{x_k, y_k\} \in P$. First observe, that if $|u| = 0$, then it means that $M$ contains a pair from $P$, and consequently, there is a word $w$ compressing $M$ of length $|w| \leq h(P)$. Since, by Theorem 1, $(n - m + 2) \geq p(P)$, $w$ has the required length.

Thus, we may assume that $|u| \geq 1$. Let $M_i = Ma_1 \ldots a_{i-1}$ for $1 \leq i \leq t + 1$. Since $|M_{i+1}| < |M|$ or there are $\{x, y\} \in M_{i+1}$, we have that $M_i$ contains two distinct states $x, y$ such that either $x_i a_t = y_i a_t$, or $x_i a_t = x$ and $y_i a_t = y$. For $\ell \geq i \geq 1$ we define $R_i = \{x_i, y_i\} \subseteq M_i$ by $x_i a_i = x_{i+1}$ and $y_i a_i = y_{i+1}$. The sequence $(M_i, R_i)$ is a Frankl-Pin sequence. Indeed, condition 1 holds by definition. For condition 2, assume that $R_i \subseteq M_j$ for some $1 \leq j < i \leq \ell$. Then, for word $u' = a_1 \ldots a_{i-1} a_i \ldots a_t$ we have $x u' = x u$ and $y u' = y u$, and $u'$ is shorter than $u$, which is a contradiction.

We extend the sequence $(M_i, R_i)$ as follows. Let $(T_i, P_i)$ be a Frankl-Pin sequence over $P$ with $1 \leq i \leq p(P)$. Define

- $M'_i = M_i$ and $R'_i = R_i$ for $1 \leq i \leq \ell$;
- $M'_{i+1} = T_{i+1} - \ell$ and $R'_{i+1} = P_{i+1} - \ell$ for $\ell + 1 \leq i \leq \ell + p(P)$.

Then $(M'_i, R'_i)$ is a Frankl-Pin sequence of length $\ell + p(P)$. Indeed, condition 1 trivially holds. For condition 2 it is enough to observe that, by the definitions of $M_i$ and $(T_i, P_i)$, for $i > \ell$, $R'_i = P_{i-\ell}$ is not in any $M'_j$ with $j < i$.

Now, by Theorem 1, $\ell + p(P) \leq (n - m + 2)$. Thus $|u| \leq (n - m + 2) - p(P)$. If $|Mu| < |M|$ we are done. Otherwise, $\{x', y'\} \in Mu$ for some $\{x', y'\} \in P$ and we must append to $u$ a word compressing $\{x', y'\}$, which has length at most $h(P)$. As a result we obtain a word $w$ of length at most $(n - m + 2) - p(P) + h(P)$, as required. \hfill \Box

Our result is to be applied in concrete situations, when we can find a relatively large set of compressible pairs with small synchronizing height. In order to estimate the minimal length of Frankl-Pin sequence we have the following auxiliary result. Given words $w_1, \ldots, w_k$ we define $P(w_1, \ldots, w_k)$ to be the set $P$ of pairs $(x, y)$ such that $x w_i = y w_i$ for some $1 \leq i \leq k$. Given $k$, choose words $w_1, \ldots, w_k$ so that $P(w_1, \ldots, w_k)$ is of maximal cardinality. Denote this cardinality by $p(k)$.

Proposition 2. If $A = (Q, \Sigma, \delta)$ is an $n$-state automaton of rank $r$, then for each $2 \leq m \leq r$ there exists an $m$-subsets Frankl-Pin sequence of length $p = \lfloor (r - m + 3)/2 \rfloor$.

Proof. Fix $m \leq r$, and let $c = \lfloor (r - m + 3)/2 \rfloor$. Choose $w_1, \ldots, w_c$ so that $P = P(w_1, \ldots, w_c)$ has cardinality $p = p(c)$. Fix some ordering of pairs in $P$ denoting $P = (P_i), 1 \leq i \leq p$. $P_i = \{x_i, y_i\}$. We proceed to define corresponding $(M_i)$. To this end, let $w$ be a word of rank $r$, and let $S = Qw$. Then $S$ is not compressible, and in particular, no pair $(x, y) \in P$ is contained in $S$.

Given a pair $P_i = \{x_i, y_i\}$, let $T_i$ consists of all elements $x \in S$ such that $x w_k = x_i w_k$ or $x w_k = y_i w_k$ for some $1 \leq k \leq c$. Note that, since $S$ is not compressible, for every $k$, there is
at most one \( x \) such that \( xw_k = x_iw_k \). Similarly, there is at most one \( x \) such that \( xw_k = y_iw_k \). Moreover, since \( x_iw_k = y_iw_k \) for some \( k \), the cardinality \( |T_i| \leq 2c - 1 \), and consequently the \( |S \setminus T_i| \geq r - 2c + 1 \geq m - 2 \).

We define \( M_i = S'_i \cup \{x_i, y_i\} \), where \( S'_i \) is an arbitrary \((m - 2)\)-subset of \( S \setminus T_i \). Now, consider a pair \( P_j = \{x_j, y_j\} \) with \( j \neq i \). Since \( S'_i \) is not compressible, \( P_j \not\subseteq S'_i \). Since \( P_j \neq P_i \), the remaining case for \( P_j \subseteq M_i \) is when \( |P_j \cap S'_i| = 1 \) and one of \( x_i \) or \( y_i \) belongs to \( P_j \). Then, take \( k \) with \( x_jw_k = y_jw_k \). It follows that there is \( x \in S'_i \) such that either \( xw_k = x_iw_k \) or \( xw_k = y_iw_k \), which contradicts the fact that \( S'_i \subseteq S \setminus T_i \). Consequently, \( P_j \not\subseteq M_i \). This proves that \((P_i, M_i)\) is a Frankl-Pin sequence over \( P \) of required length.

\[ \square \]

In the case of \(|\Sigma| = 1\) we have the following more specific result, which may be used to estimate the reset length, when one letter of the automaton is known.

**Corollary 1.** If \( A \) is a unary automaton of rank \( r \), and \( P \) is the set of all compressible pairs in \( A \), then for each \( 2 \leq m \leq r \) there exists an \( m \)-subsets Frankl-Pin sequence over \( P \) of length \( p(P) = |P| \). Moreover, \( |P| \geq \frac{2}{3}n(\frac{r}{2} - 1) \), and \( h(P) = n - r \).

**Proof.** We have \( \Sigma = \{a\} \) in this case, and each word is of the form \( w = a^h \). It follows that for each set \( w_1, \ldots, w_k \), there is \( h \leq n-r \) such that \( P(w_1, \ldots, w_k) = P(a^h) \). Note that \( a^{n-r} \) compresses each compressible pair. Consequently, taking \( w_1 = a^{n-r} \), we see that for every \( k \), \( p(k) \) is the cardinality of the set \( P \) of all compressible pairs, and the first part of the result follows from Proposition \( \ref{prop:compressible_pairs} \).

The set \( P \), in this case, is the union of equivalence classes of the relation determined by the letter \( a \) with the condition: \( q, p \in Q \) are in the relation if and only if \( qa^{n-r} = pa^{n-r} \). If \( c_1, \ldots, c_r \) are the cardinalities of the equivalence classes, then \( |P| \geq \sum_{i=1}^{r} \frac{c_i(c_i-1)}{2} \). The expression achieves its minimum when all \( c_i \) are equal. Hence, \( |P| \geq r \frac{(r^2 - 1)}{2} \), as claimed. Obviously, \( h(P) \) is given by the fact that \( a^{n-r} \) compresses each compressible pair.

\[ \square \]

Of course, the bound above is the worst case. Having a concrete letter \( a \) one may compute the exact value of \(|P|\), which in particular cases may be as large as \( \binom{n+r}{2} \).

Proposition \( \ref{prop:compressible_pairs} \) should be combined and compared with the following result proved by J.-E. Pin:

**Proposition 3.** (J.-E. Pin, \[\ref{pin:19} \] Proposition 5) Let \( A = \langle Q, \Sigma, \delta \rangle \) be an automaton with \( n = |Q| \) states, and \( u \in \Sigma^* \) be a word of rank \( m \). If there is a word of rank \( \leq m - 1 \), then there is such a word of length at most \( 2|u| + n - m + 1 \).

In our computation, given an automaton \( A \) of rank \( m \), we can bound the reset length of a synchronizing extension by applying successively \( m - 1 \) times either Proposition \( \ref{prop:compressible_pairs} \) or Proposition \( \ref{prop:compressible_pairs} \) depending on which bound is smaller for the given rank. It can be demonstrated that the best results are achieved when each time we take a word of the minimal rank \( m \), and check which of the propositions gives a better bound. It turned out that first, for larger ranks, Proposition \( \ref{prop:compressible_pairs} \) gives a better bound, and then Proposition \( \ref{prop:compressible_pairs} \) becomes more effective. Also note that, when applied to get a bound on the length of a shortest reset word, the Pin’s result gives an exponential estimate, while our result gives a polynomial bound.

### 2.2 One-cluster Automata

Our next aim is to generalize the results on one-cluster automata obtained by Béal, Berlinkov, Perrin \[\ref{beal:05} \] and Steinberg \[\ref{steinberg:21} \]. Recall that an automaton \( A = \langle Q, \Sigma, \delta \rangle \) is one-cluster, if it has a
letter \( a \in \Sigma \) such that for every pair \( q, s \in Q \) there are \( k, m \geq 1 \) such that \( qa^k = sa^m \). This means that the graph of transformation induced by \( a \) is connected. In particular, it has a unique cycle \( C \subseteq Q \) with the property \( Ca^m = C \) for every \( m \geq 0 \), and there is \( \ell \geq 0 \) such that \( Qa^\ell = C \). The least such \( \ell \) is called the level of \( A \). Steinberg proved the following:

**Theorem 2.** (B. Steinberg [21]) Let \( A \) be a synchronizing one-cluster automaton with \( n \) states, level \( \ell \) and cycle \( C \) of prime length \( p \). Then \( A \) has a reset word of length at most

\[
 n - p + 1 + 2\ell + (p - 2)(n + \ell).
\]

We generalize this result for non-prime lengths of the cycle. Our proof is similar to that of Steinberg, but we use a different crucial lemma that does not use the fact that \( p \) is prime. Assuming that \( \mathcal{A} = \langle Q, \Sigma, \delta \rangle \) is a synchronizing one-cluster automaton with \( n \) states, level \( \ell \) and cycle \( C \) of length \( m \), we use here the notions and notations of [21]. In particular, we consider the matrix representation \( \pi: \Sigma' \to \text{M}_n(Q) \) defined by \( \pi(w)_{q,r} = 1 \) if \( qw = r \), and 0, otherwise. Given \( S \subseteq Q \) we define \([S]\) to be the characteristic row vector of \( S \) in \( Q \), \([S]^T\) its transpose, and

\[
\gamma_S = [S]^T - ([S] / |C|)[Q]^T.
\]

In fact, we will consider only subsets \( S \subseteq C \), so \( |S| / |C| = k / m \) for some \( k \leq m \), and \( \gamma_S \) is a column vector of \( \mathbb{Q}^n \) with entries \( 1 - k / m \) and \(-k / m\) only. Now, by \([C]^w\gamma_S \) we denote the product of corresponding matrices; in particular, the word \( w \) represents in this notation the matrix \( \pi(w) \), and the whole product is an element of \( \mathbb{Q} \). In [21] and earlier papers, the following fact (not difficult to see) is used

\[
[C]^w\gamma_S = |C \cap Sw^{-1}| - |S|.
\]

If this difference is larger than zero then the preimage of \( S \) by the word \( w \) has more elements in the cycle \( C \) than \( S \) itself, and in consequence \( w \) compresses \( C \). So, in general, we are looking for short words \( w \) for which \([C]^w\gamma_S > 0 \).

In order to formulate our crucial lemma we define the number \( D(m,k) \) as follows. Let \([S] = (a_1, \ldots, a_m) \), \( a_i \in \{0, 1\} \). By the cyclic transforms of \([S] \) we mean the following vectors: \((c_1, \ldots, c_m)\), \((c_2, \ldots, c_m, c_1)\), \ldots, \((c_m, c_1, \ldots, c_{m-1})\). Given \( 1 \leq k \leq m \), by \( D(m,k) \) we denote the minimal dimension of the subspace of \( \mathbb{Q}^m \) generated by the cyclic transforms of a vector \([S] \) with \(|S| = k \) (that is, \([S] \) runs here over all vectors with exactly \( k \) ones and \( m-k \) zeros). Obviously, \( D(m,1) = D(m, m-1) = m \). Yet, for example, \( D(2k,k) = 2 \). More information about \( D(m,k) \) can be inferred from [13], where in particular the rank of the matrix generated by cyclic transforms of a vector is considered.

**Lemma 1.** Let \( \mathcal{A} = \langle Q, \Sigma, \delta \rangle \) be a synchronizing one-cluster automaton with \( n \) states, level \( \ell \) and cycle \( C \) of length \( m \). Let \( S \) be a subset of \( Q \) of cardinality \(|S| = k > 0\), and \( \mathcal{W} = \text{Span}\{a^\ell \gamma_S \in \mathbb{Q}^n | 0 \leq j \leq m - 1 \} \). If \( S \subseteq C \), then \( \dim \mathcal{W} \geq D(m,k) - 1 \) and

\[
\sum_{0 \leq j \leq m - 1} a^{\ell + j} \gamma_S = 0.
\]

**Proof.** First note, that by definition,

\[
a^{\ell + j} \gamma_S = a^{\ell + j}[S]^T - a^{\ell + j} \frac{k}{m}[Q]^T.
\]
The summands are (as it is easy to check; see [21]) the characteristic vectors of preimages

\[ [S(a^{\ell+j})^{-1}]^T - \frac{k}{m}[Q(a^{\ell+j})^{-1}]^T = [S(a^{\ell+j})^{-1}]^T - \frac{k}{m}[Q]^T. \]

Hence,

\[ \sum_{0 \leq j \leq m-1} a^{\ell+j}\gamma_S = \left( \sum_{0 \leq j \leq m-1} [S(a^{\ell+j})^{-1}]^T \right) - m[Q]^T. \]  \tag{3}

In order to compute the sum on the right hand side, we observe that, for each \( 0 \leq j \leq m, \)

\[ \sum_{q \in S} [q(a^{\ell+j})^{-1}]^T = [Q]^T. \]

This is so, because for every \( q \in Q \) and \( s \in C \), there is \( 0 \leq j \leq m-1 \) such that \( qa^{\ell+j} = s \), and for all \( 0 \leq i, j \leq m-1 \), and \( q \in Q \), if \( qa^{\ell+j} = qa^{\ell+i} \) then \( i = j \). It follows that

\[ \sum_{0 \leq j \leq m-1} [S(a^{\ell+j})^{-1}]^T = \sum_{0 \leq j \leq m-1} \sum_{q \in S} [q(a^{\ell+j})^{-1}]^T = \sum_{0 \leq j \leq m-1} [Q]^T = m[Q]^T. \]

Combining this with (3) yields (2).

It remains to estimate the dimension of \( W \). Let us denote \( w_j = [S(a^{\ell+j})^{-1}]^T \), and for \( c \in Q \), \( \hat{c} = c[Q]^T \). Then, \( a^{\ell+j}\gamma_S = v_j - \hat{c} \), for \( c = k/m \). We consider the restriction \( V \subset \mathbb{Q}^m \) of \( W \) to the coordinates corresponding to \( [C] \), which formally is the image in the orthogonal projection \( \phi \) of \( W \) on the orthogonal complement of \( [Q \setminus C] \). Then, of course, \( \dim W \geq \dim V \), and it is enough to estimate \( \dim V \) from below.

Let \( v_i = \phi(w_i) \in V \) be the image of \( w_i \in W \) (i.e. restriction of \( v_i \) to \( m \) coordinates corresponding to \( C \)). Then \( v_0 \) is the characteristic vector of \( S \) in \( C \), and \( v_0, \ldots, v_{m-1} \) are simply the cyclic transforms of \( v \) in \( \mathbb{Q}^m \). Consequently, for \( U = \text{Span}\{v_0, \ldots, v_{m-1}\} \), \( \dim U \geq D(m,k) \). Moreover, we have \( V = \text{Span}\{v_0 - \bar{d}, \ldots, v_{m-1} - \bar{d}\} \), where \( \bar{d} \in \mathbb{Q}^m \) denotes \( \bar{d} = d[C]^T \) with \( d = k/m \). Since \( \sum_{0 \leq j \leq m-1} v_j = k \),

\[ U = \text{Span}\{v_0, \ldots, v_{m-1}, \bar{d}\} = \text{Span}\{v_0 - \bar{d}, \ldots, v_{m-1} - \bar{d}, \bar{d}\} = \text{Span}\{V, \bar{d}\}. \]

Now, since the sum of the coordinates in each \( v_i - \bar{d} \) is equal to

\[ k(1 - k/m) - (m - k)k/m = 0, \]

it follows that \( \bar{d} \notin V \). Consequently

\[ \dim V \geq \dim U - 1 \geq D(m,k) - 1, \]

as required.

\[ \__ \]

**Theorem 3.** Let \( A \) be a synchronizing one-cluster automaton with \( n \) states, level \( \ell \) and cycle \( C \) of length \( m > 3 \). Then \( A \) has a reset word of length at most

\[ (n + \ell + m - 3)(m - 1) + \ell - 2 - \sum_{k=2}^{m-2} D(m,k). \]
Proof. We modify suitably the argument used in the proof in [21]. Let $S$ be a proper subset of $Q$ with $0 < k = |S| < m$. We wish to show first that there exists a short word $w \in \Sigma^*$ with $[C]w\ell^{+j}\gamma_S \neq 0$ for some $0 \leq j \leq m - 1$. If this holds for the empty word, we are done. Otherwise, $[C]w\ell^{+j}\gamma_S = 0$ for all $0 \leq j \leq m - 1$. This means $a^{\ell+j}\gamma_S \in [C]^\perp$ (the orthogonal complement of $[C]$). Since $A$ is synchronizing, there exists a word $u$ resetting to a state in $C \cap S(a^{\ell})^{-1}$. Then, by (1),

$$[C]u\ell\gamma_S = |C \cap S(a^{\ell})^{-1}| - |S| = |C \cap Q| - |S| \neq 0. \tag{4}$$

To find $u$ short enough with this property, let $W = \text{Span}\{a^{\ell+j}\gamma_S \in Q^n \mid 0 \leq j \leq m - 1\}$. We have $W \subseteq [C]^\perp$, yet by (1), $aW \not\subseteq [C]^\perp$. By the standard ascending chain condition (see [21] Lemma 2; cf. also [19,10,16]), we infer that there exists a word $w$ satisfying $[C]w\ell^{+j}\gamma_S \neq 0$ for some $j$, whose length $|w| \leq \dim C^\perp - \dim W + 1$. By Lemma 1 we get $|w| \leq n - D(m,k) + 1$.

Moreover, by the same lemma,

$$\sum_{0 \leq j \leq m-1} |C|w\ell^{+j}\gamma_S = |C|w \sum_{0 \leq j \leq m-1} a^{\ell+j}\gamma_S = 0.$$

Since $|C|w\ell^{+j}\gamma_S \neq 0$ for some $j$, there must be $j$ such that $|C|w\ell^{+j}\gamma_S > 0$. This means, by (1), that $|C \cap S(wa^{\ell+j})^{-1}| - |S| > 0$, for some $0 \leq j \leq m - 1$ and $|w| \leq n - D(m,k) + 1$. Since $j \leq m - 1$, it means that there exists a word $u_k$ of length at most $n + \ell + m - D(m,k)$ such that $|C \cap S(u_k^{-1})| > |S|$.

Thus, as in [21], we may find a sequence of words $u_1, u_2, \ldots, u_{m-1}$ such that starting from an arbitrary one-element set $S_1 = \{q\}$, we have $|S_k| > |S_{k-1}|$ for $S_k = C \cap S_{u_k^{-1}}$, $1 \leq k \leq m - 1$, and the length of $|u_k| \leq n + \ell + m - D(m,k)$. In particular, the word $u = u_{m-1} \ldots u_2u_1$ has the property $Cu = \{q\}$, and since $Qa^\ell = C$, the word $a^\ell u$ is synchronizing. Since the suffix of this word is $u_1 = wa^\ell$, we may replace $u_1$ by a shorter word $u'_1 = wa^\ell$, and $v = a^\ell u_{m-1} \ldots u_2u'_1$ is also synchronizing. For the length (using the inequalities obtained in the preceding paragraph) we have

$$v \leq \ell + (m-2)(n+\ell+m) - \left(\sum_{2 \leq k \leq m-1} D(m,k)\right) + n - D(m,1) + 1 + \ell.$$

Taking into account that $D(m,1) = D(m,m - 1) = m$ we obtain the required result. \hfill \Box

We note that it is only the last line why we need to assume $m > 3$. For $m < 3$ our proof yields exactly the bounds of Theorem 2 $2n + 3\ell - 2$ for $m = 3$, and $n + 2\ell - 1$ for $m = 2$ (which are of course very useful in our computations).

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|c|c|c|c|c|c|}
\hline
$m \setminus k$ & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & $\sum_{k=2}^{m-2} D(m,k)$ \\
\hline
4 & 1 & 2 & 4 & 1 & & & & & & & & & 2 \\
6 & 6 & 3 & 2 & 3 & 6 & 1 & & & & & & & 8 \\
8 & 8 & 4 & 8 & 2 & 8 & 4 & 1 & & & & & & 26 \\
9 & 9 & 9 & 3 & 9 & 3 & 9 & 1 & & & & & & 42 \\
10 & 10 & 5 & 10 & 5 & 2 & 5 & 10 & 5 & 10 & 1 & & & 42 \\
12 & 12 & 6 & 4 & 3 & 8 & 2 & 8 & 3 & 4 & 6 & 12 & 1 & 44 \\
\hline
\end{tabular}
\caption{The values of $D(m,k)$ and $\sum_{k=2}^{m-2} D(m,k)$ for non-prime $m$.}
\end{table}
Concerning the number $D(m, k)$, Ingleton [14] showed that the dimension of the vector space generated by the transforms of a vector $S = (c_1, \ldots, c_m)$ is exactly $m - d$, where $d$ is the degree of the polynomial $g(x) = \gcd(c_1 + c_2x + \ldots + c_m x^{m-1}, x_m - 1)$ in $\mathbb{Q}[x]$. In case, when $m$ is prime, $g(x) = 1$, and $D(m, k) = m$ for all $k$. Substituting this we obtain the result of Steinberg [21]. In general, what makes our estimation better than \((n + m + \ell - 2)(m - 2) + (2n - m - 1) + \ell\) obtained by Béal, Berlinkov and Perrin [5], and \((2n - 3)(m - 2) + n - m + 2 + 2\ell\) announced in Steinberg [21] is the summand $-\sum_{k=2}^{m-2} D(m, k)$. For small $m$, this summand can be easily computed, and a number of letters can be omitted while proving the Černý conjecture for the automata with a fixed number of states (see Table 1). In particular, if $A$ is a synchronizing one-cluster automaton with cycle $C$ of length $m < 8$, then $A$ satisfy the Černý conjecture. But also for larger cycles many cases can be settled just by using Theorem 3.

While computing bounds for reset length for automata with more than 2 letters it happens that the partial automaton is not one-cluster, but it has a one-cluster transformation. The proof of Theorem 3 does not require $a$ to be a single letter. Hence, taking into account the length of a one-cluster transformation we may generalize our result as follows:

**Corollary 2.** Let $A = \langle Q, \Sigma \rangle$ be a synchronizing automaton with $n$ states, such that there exists a word $w$ of length $s$ inducing a one-cluster transformation with level $\ell$ and cycle $C$ of length $m > 2$. Then $A$ has a reset word of length at most

$$s(\ell + m - 2)(m - 1) + (n - 1)(m - 1) + s\ell - 2 - \sum_{k=2}^{m-2} D(m, k).$$

3 The Algorithm

We describe some details of the algorithm that we use in searching for interesting examples of synchronizing automata. Note that the number of automata grows very quickly with the number of letters (see [3]). To overcome this, we exclude a priori a large part of automata from generation process. To do so, we apply the results developed in the previous section.

Such algorithm is useful not only for verifying the Černý conjecture, but for investigating various conjectures in the area. Our goal is to exhaustively search over the automata with given size $n$ and arity $k$ and to report all those with a long reset length, say with a reset length longer that a predefined threshold. For example, setting $\text{threshold} = (n - 1)^2$ verifies the Černý conjecture for the fixed $n$ and $k$.

Obviously, we are interested only in irreducible synchronizing automata, that is those for which removing any letter results with a non-synchronizing automaton. Also, it is well known [23] that to verify the Černý conjecture it is enough to consider only strongly connected automata (those with the underlying digraph strongly connected), so a special attention is paid to this class.

3.1 Sieving Procedure

As we have already mentioned we generate all $k$-ary automata with $n$ states running successively the algorithm described in [18] for arities $i = 2, \ldots, k$. For each automaton $A$ generated by the algorithm, we apply the sieving procedure (Algorithm 1). The procedure checks if an automaton should be reported, and whether it should be kept for the next $(i + 1)$-th run of the algorithm.
Algorithm 1 Sieving Procedure

Require: $A$ – a generated automaton with $n$ states on $k$ letters.
Require: threshold – the bound restricting reset length.

1: procedure Sieve($A$)
2: if $A$ is synchronizing then
3: if $A$ is strongly connected and irreducible synchronizing then
4: $\ell \leftarrow$ the reset length of $A$
5: if $\ell \geq$ threshold then Report $A$ end if
6: end if
7: else
8: Compute the minimal rank $m$, a word $u$ of this rank, and the set of all compressible pairs $P$.
9: Compute the bound from successive applications of Proposition 1 and Proposition 3 for $A$.
10: if the bound is not larger than threshold then return end if
11: $T_A \leftarrow$ the transition semigroup of $A$
12: for all $t \in T_A$ such that $t$ is a one-cluster transformation do
13: if the bound from Corollary 2 is not larger than threshold then return end if
14: end for
15: if $\{t_a : a \in \Sigma\}$ is a reducible set of generators of $T_A$ then return end if
16: Store $A$ for the next run
17: end if
18: end procedure

First we check if $A$ is synchronizing (line 2). If so, we check if it is irreducible synchronizing and if its reset length is larger than threshold (lines 3-6), so we could report it.

If $A$ is not synchronizing, then it is a potential candidate for further processing in the next run of the algorithm. Using the methods from Section 2 we check if all its irreducible synchronizing extensions of $A$ have reset length not larger than threshold.

We compute the minimal rank $m$ of $A$, a word $u$ of this rank, and the set of all compressible pairs $P$ (line 8). This can be done by a standard BFS algorithm on the power automaton, in the same manner as computing a shortest reset word. It can be seen (as we have mentioned in the remark following Proposition 3) that using the minimal rank and a shortest word of this rank yields the best bound from Proposition 1 and Proposition 3. Hence we apply successively both the propositions for $r = m, \ldots, 2$ (line 9). To use Proposition 1 we need also to know $h(P)$, which is the maximum synchronizing height over the pairs from $P$, and $p(P)$, which is the length of a Frankl-Pin sequence over $P$. The value of $h(P)$ is easily computable having $P$. However, there is no known effective algorithm finding a longest $r$-subsets Frankl-Pin sequence over $P$, and a brute-force algorithm has potentially at least double exponential running time. Hence we use the following greedy algorithm to compute such a sequence: We pick a pair $P_i \in P$ whose states are involved in the least number of the other pairs from $P$. We remove $P_i$ from $P$. Then we try to find $M_i \subseteq Q$ in a similar greedy manner. If $M_i$ of the size $r$ is found, then we append $(P_i, M_i)$ to the sequence. We continue this process until $P$ is empty. If the bound is not larger than threshold, then we skip the automaton (line 10).

For further estimations, we compute the transition semigroup $T_A$ of $A$ (line 11). For each transformation in $T_A$ we store the length of the shortest words inducing it. Then for each one-cluster transformation we check the bound from Corollary 2 (lines 12-14).
Finally, since we are interested only in irreducible synchronizing automata, we skip also $A$ if the set of the transformations induced by the letters $a \in \Sigma$ is a reducible set of generators. This is done at the end, since this procedure has high computational cost.

Table 2 illustrates savings in computation resulted due to using various exclusions in the sieving procedure. It contains the numbers of automata remaining after exclusions based on the results named in the first column. The second and the third columns contain numbers for the two selected cases of computation with different threshold, $n$, and $k$.

|                | I            | II           |
|----------------|--------------|--------------|
| No exclusions  | 7,864,973    | 187,138,741  |
| Proposition 1  | 1,033,590    | 1,372,878    |
| Proposition 3  | 1,804,727    | 3,644,756    |
| Proposition 1 with Proposition 3 | 916,354 | 1,206,910  |
| Corollary 2    | 4,081,378    | 74,650,059   |
| Irreducible generators of $T_A$ | 7,864,012 | 177,187,304 |
| All exclusions  | 817,894      | 447,089      |

3.2 Further Exclusions

In the last $k$-th run we do not need to store automata for the next run. Also the number of automata in the last run is usually the largest compared to the previous runs. Hence we may use further additional checks to reduce the number of generated automata. In this subsection we present some facts that are useful to this aim.

**Proposition 4.** For a binary synchronizing automaton $A$, if both transformations $t^2 = a$ or $t^2$ is the identity transformation, then $A$ has a reset word of length at most $2n - 2$.

**Proof.** Let $\Sigma = \{a, b\}$. Consider a shortest reset word $w$. It has no two the same consecutive letters, so it is of the form $w = (ab)^m$ or $w = (ab)^m a$. Obviously, since $w$ is synchronizing, either $a$ or $ab$ collapses $Q$, that is $|Qab| < |Q|$, and the same can be said for any $Q' = Q(ab)^t$. It follows that $m \leq n - 1$ and $|w| \leq 2n - 2$. □

**Proposition 5.** Let $A = (Q, \Sigma, \delta)$ be a $k$-ary synchronizing automaton of size $n \geq 5$. If the Černý conjecture is true for all $k$-ary automata of size less than $n$, and $A$ is not strongly connected, then $A$ has a reset word of length at most $n^2 - 4n + 5$.

**Proof.** Let $X \subseteq Q$ be the sink component of $A$. There is a word $w$ which maps every state from $Q \setminus X$ to $X$ such that $|w| \leq 1 + 2 + \cdots + |X| = |X|(|X| + 1)/2$. Let $v$ be the shortest reset word synchronizing $X$, $|v| \leq (n - 1 - |X|)^2$. Then $wv$ resets $A$ and

$$|wv| \leq \frac{|X|(|X| + 1)}{2} + (n - 1 - |X|)^2 = n^2 - 2n|X| - 2n + \frac{3}{2}|X|^2 + 5|X| + 1.$$

For $1 \leq |X| \leq n - 1$ and $n \geq 5$ the expression yields the maximum $n^2 - 4n + 5$ for $|X| = 1$. □
We note that in our algorithm we are able to utilize those special results on the Černý conjecture that refer to a part of the automaton. The results referring to the structure properties not preserved by restrictions of the alphabet are difficult to use. This concerns, in particular, the general result of [13]. Nevertheless, we are able to apply successfully a very special case of [13], which we describe now.

A pair of states \( x, y \) \( \in Q \) is called twin in \( A = \langle Q, \Sigma, \delta \rangle \), if \( xa = ya \) or \( \{xa, ya\} \subseteq \{x, y\} \) for each letter \( a \in \Sigma \). In such a case the equivalence relation whose the only nontrivial block is \( \{x, y\} \) is a congruence, and the factor automaton is synchronizing if \( A \) is synchronizing. Formally we define \( A' = \langle Q', \Sigma, \delta' \rangle \) to be the automaton obtained by identifying the twin states \( x \) and \( y \) as follows:

\[
Q' = (Q \cup \{z\}) \setminus \{x, y\},
\]

\[
\delta'(s, a) = \begin{cases} 
z, & \text{if } s \neq z \text{ and } \delta(s, a) \in \{x, y\}, \\
\delta(s, a), & \text{if } s \neq z \text{ and } \delta(s, a) \in Q \setminus \{x, y\}, \\
\delta(x, a), & \text{if } s = z.
\end{cases}
\]

Then we have

**Proposition 6.** Let \( A \) be a strongly connected synchronizing automaton. If \( A \) has two twin states \( x, y \), and a reset word of length \( r \), then the factor automaton \( A' \) is strongly connected, synchronizing, and has a shortest reset word of length \( r - 1 \) or \( r \).

**Proof.** Any state \( p \in Q \) is reachable from \( x \), and \( x \) is reachable from \( p \). Hence the same holds for \( p \in Q' \) and \( z \), and so, \( A' \) is strongly connected. Obviously, a word synchronizing \( A \) also synchronizes \( A' \). Let \( w \) be a shortest reset word for \( A' \). If \( w \) is not a reset word for \( A \), then \( Qw = \{x, y\} \). Since \( A \) is synchronizing there exists \( a \in \Sigma \) such that \( xa = ya \), as otherwise \( \{xa, ya\} = \{x, y\} \) for all \( a \in \Sigma \). So \( wa \) is a reset word for \( A \) of length \( |w| + 1 \).

Thus we can skip the automata with a twin pair (which is easily recognizable). Similarly, as in Proposition 5 one can observe that if \( A \) is an automaton of size \( n \) with a twin pair, and the Černý conjecture is true for all automata of size less than \( n \), then \( A \) has the reset length at most \( n^2 - 4n + 5 \).

**4 Results**

We have verified that the Černý conjecture is true for automata with \( n \leq 5 \) states, regardless of the alphabet size. For \( n = 6 \) we checked automata up to \( k = 5 \) letters, for \( n = 7 \) up to 4 letters, and for \( n = 8 \) up to 3 letters. Our computation confirms, in particular, all results stated in [22].

It is especially important, since it remains unclear whether the results announced in [22] are based on partial computation or on unknown ideas. Anyway, no clue is given how and on what base the generation of automata is restricted, while it is apparent that it is impossible to perform this computation with brute force approach. For \( n = 8 \) and \( k = 3 \) we checked 20,933,723,139 automata. Without using parallelism, the computation would take 1.25 years of one CPU core. Compare this with the number 572,879,126,392,178,688 of ICFA automata that one would need to generate applying the technique described in [2]. Our computation was focused on discovering new interesting examples automata and gaps in reset lengths sequences. Some of the results are presented below.
4.1 Slowly Synchronizing Automata on a Ternary Alphabet

Automata with reset lengths close to the Černý bound \((n - 1)^2\) are referred to as slowly synchronizing. The authors of [3] constructed a few series of such automata over binary alphabets. Yet, until now, no series of synchronizing automata with reset length \(n^2 + O(n)\) have been presented for larger alphabets. We have discovered two new series of ternary irreducible slowly synchronizing automata. They turn out to be related to the digraph \(W_n\) (see [3]).

Let \(Q_n = \{v_1, \ldots, v_n\}, n \geq 3\), and \(\Sigma = \{a, b, c\}\). Let \(\mathcal{M}_n = \langle Q_n, \Sigma, \delta_n \rangle\) and \(\mathcal{M}'_n = \langle Q_n, \Sigma, \delta'_n \rangle\) be the automata shown in Figure 2. The transition functions \(\delta_n\) and \(\delta'_n\) are defined as follows:

\[
\delta(v_i, a) = \delta'(v_i, a) = \begin{cases} 
v_{i+1}, & \text{if } 1 \leq i \leq n - 1, \\
v_2, & \text{if } i = n,
\end{cases}
\]

\[
\delta(v_i, c) = \begin{cases} 
v_n, & \text{if } i = 1, \\
v_i, & \text{if } 2 \leq i \leq n - 1, \\
v_1, & \text{if } i = n,
\end{cases}
\]

\[
\delta'(v_i, b) = \delta'(v_i, b) = \begin{cases} 
v_2, & \text{if } i = 1, \\
v_i, & \text{if } 2 \leq i \leq n - 1, \\
v_1, & \text{if } i = n.
\end{cases}
\]

\[
\delta'(v_i, c) = \begin{cases} 
v_i, & \text{if } 1 \leq i \leq n - 1, \\
v_1, & \text{if } i = n.
\end{cases}
\]

In determining the reset length of the series we applied a technique, which is alternative to that of [23]. Our method is based on analyzing the behavior of the inverse BFS algorithm finding the length of the shortest reset words (see [17]) and is suitable, in general, for a more mechanical way to establish the reset length of concrete automata. (In particular, it works fine for automata defined in [3]). The description of the method and proofs based on it will be presented in the extended version of the paper.

For a synchronizing automaton \(\mathcal{A} = \langle Q, \Sigma, \delta \rangle\) with \(n > 1\) states, we define the sequence of families \((L_i)\) of the subsets of \(Q\). Let \(L_0\) be the family of all the singletons, whose are a common end of more than one edge with the same label. We will define \(L_i\) inductively for \(i \geq 1\). Let \(L_i' = \{S a^{-1} : S \in L_{i-1}, a \in \Sigma\}\). A set \(S \in L_i'\) is called visited, if \(|S| = 1\) or there is \(T \in L_j, T \supseteq S\) for \(j < i\), or there is \(T \in L_i, T \supseteq S\). We define \(L_i\) to be the set of all non-visited sets from \(L_i'\).
The proof of the following lemma in a more general form can be found in [17, Theorem 1], but we include it here for the sake of completeness.

**Lemma 2.** There exists a shortest reset word \( w \) such that for any suffix \( u \) of \( w \) of length \( i \), \( \{ q \} u^{-1} \in L_i \) for some \( q \in Q \). The smallest \( i \) such that \( Q \in L_i \) is the reset length of the automaton.

**Proof.** If \( w \) is a shortest reset word, then it synchronizes the automaton in a state \( q \), which is a common end of more than one edge with the same label. Hence the statement holds for \( i = 0 \) with any of the shortest reset words.

Let \( i \geq 1 \) be the smallest integer such that the statement does not hold. Get a shortest reset word \( w \) satisfying the statement for \( i - 1 \). So \( w = uav \), where \( |v| = i - 1 \). By the assumption, there is \( S = \{ q \} v^{-1} \in L_{i-1} \) for some \( q \in Q \). By the definition of \( L_i \), it follows that \( Sa^{-1} \) must be visited, since it is not in \( L_i \). Since \( w \) is a shortest reset word, clearly \( Sa^{-1} \) cannot be a singleton, so there is \( T \geq Sa^{-1}, T \in L_j \) for \( j \leq i \); we can get \( v' \) such that \( \{ q' \} v'^{-1} = T \) for some \( q' \in Q \). Then \( w' = uv' \) is also a synchronizing word. If \( j < i \) then \( |v'| < |v| \) and \( w' \) is a shorter word — a contradiction. Otherwise \( \{ q' \} v'^{-1} \in L_i \), and the statement holds for \( i \) with \( w' \) — a contradiction again.

If \( Q \in L_i \) then it means that there exists a shortest reset word \( w \) of length \( i \). Hence \( i \) is the reset length. \( \square \)

**Theorem 4.** For \( n \geq 3 \), the automata \( \mathcal{M}_n \) and \( \mathcal{M}'_n \) are irreducible synchronizing. The first has reset length \( n^2 - 3n + 3 \), and the second has reset length \( n^2 - 3n + 2 \).

**Proof.** One easily verifies that the word \( acb(a^{n-2}cb)^{n-3} \) synchronizes \( \mathcal{M}_n \), and has the length \( 3 + (n - 2 + 2)(n - 3) = n^2 - 3n + 3 \). Also the word \( cb(a^{n-2}cb)^{n-3} \) synchronizes \( \mathcal{M}'_n \), and has the length \( n^2 - 3n + 2 \).

We show that there is no shorter reset word for \( \mathcal{M}_n \). By Lemma 2 it is sufficient to show what is the smallest \( i \) such that \( Q_n \in L_i \), which is the length of the shortest reset words. Here only \( v_2 \) is a common end of more than one edge with the same label, so \( L_0 = \{ \{ v_2 \} \} \).

We claim that \( L_{in} = \{ \{ v_2, \ldots, v_{2+i} \} \} \) for \( i = 0, \ldots, n - 3 \). Moreover \( L_{in+4} = \{ \{ v_{n-2}, v_{n-1} \} \} \) for \( i = 0 \) and \( L_{in+4} = \{ \{ v_{n-2}, v_{n-1}, v_n, v_1, \ldots, v_{i-1} \} \} \) for \( 1 \leq i \leq n - 4 \). The proof follows by induction. Clearly for \( i = 0 \) the claim holds. So consider \( L_{in} = \{ \{ v_2, \ldots, v_{2+i} \} \} \) and assume that the claim holds for \( i' < i \).

The action of \( a^{-1} \) for the set from \( L_{in} \) results in the same set, so we have \( L_{in+1} = \{ S_1, T_1 \} \), where \( S_1 = \{ v_{n-1}, v_1, \ldots, v_{i+1} \} \) was obtained by the action of \( a^{-1} \) and \( T_1 = \{ v_1, \ldots, v_{2+i} \} \) by the action of \( b^{-1} \). Consider the sets obtained from \( S_1 \). Observe that if a set contains \( \{ v_n, v_1 \} \) then the action of \( a^{-1} \) can result in a non-visited set. If \( i = 0 \) then \( S_1 a^{-1} = \{ v_n, v_1 \} a^{-1} = \{ v_{n-1} \} \) is a visited singleton. If \( i \geq 1 \) then \( S_2 = S_1 a^{-1} = \{ v_{n-1}, v_n, v_1, \ldots, v_i \} \). But \( S_2 a^{-1} = \{ v_{n-2}, v_{n-1} \} \) if \( i = 1 \), or \( \{ v_{n-2}, v_{n-1}, v_n, v_1, \ldots, v_i \} \) if \( i \geq 2 \); so \( S_2 a^{-1} \) is visited by the assumption of \( L_{(i-1)n+4} \). Consider the sets obtained from \( T_1 \). Only \( T_2 = T_1 c^{-1} = \{ v_n, v_2, \ldots, v_{2+i} \} \) is non-visited. Then let \( T_3 = T_2 a^{-1} = \{ v_{n-1}, v_1, v_1+1 \} \). For \( T_3 b^{-1} = \{ v_n, v_1, \ldots, v_{2+i} \} \) observe that in the next step it results either in \( T_3 \) or in itself. Only the action of \( a^{-1} \) applied to \( T_3 \) results in a non-visited set, so \( L_{in+4} = \{ \{ v_{n-2}, v_{n-1} \} \} \) if \( i = 0 \) and \( L_{in+4} = \{ \{ v_{n-2}, v_{n-1}, v_n, v_1, \ldots, v_i \} \} \) if \( i \geq 1 \).

Now, if \( i = 0 \) then only the action of \( a^{-1} \) results in a non-visited set over the next \( n - 4 \) steps resulting in \( L_n = \{ v_2, v_3 \} \). Similarly, if \( i \geq 1 \) then after the next \( i - 1 \) steps by the action of \( a^{-1} \) we have \( \{ v_{n-1}, \ldots, v_n \} \). Now the action of \( c^{-1} \) may result in a non-visited set containing \( \{ v_1 \} \), but in the next step all of the three actions produce visited sets from it. So after the next \( n - 3 - i \) steps by the action of \( a^{-1} \) we finally have \( L_{(i+1)n} = \{ \{ v_2, \ldots, v_{3+i} \} \} \).
From the claim it follows that $Q_n$ does not appear in $L$ for $i \leq (n-3)n$, and $L_{(n-3)n} = \{(v_2, \ldots, v_{n-1})\}$. Then applying $(acb)^{-1}$ or $(bcb)^{-1}$ results in $Q_n$, and there is no shorter such word as is easily verified. Hence $L_{(n-3)n+3} = \{\{Q_n\}\}$ and so $i = (n-3)n + 3 = n^2 - 3n + 3$ is the such that $Q_n \in L_i$.

The proof for the automaton $\mathcal{M}_n'$ follows exactly in the same way, with the two following exceptions: $T_2 = \{v_n, v_1, \ldots, v_{2+i}\}$, and finally we apply $(cb)^{-1}$ to the set from $L_{(n-3)n}$, resulting in $Q_n$.

It remains to show that removing any letter in $\mathcal{M}_n$ ($\mathcal{M}_n'$) results in a non-synchronizing automaton. Indeed, removing the letter $a$ results in unconnected states $v_3, \ldots, v_{n-1}$. The only synchronizing pairs of states are $(v_1, v_n)$ under $a$, and $(v_1, v_2)$ under $b$. Observe that it is not possible to map the pair $(v_{n-1}, v_n)$ to $(v_1, v_n)$: Only $a$ maps a state from $Q_n \setminus \{v_1, v_n\}$ to $\{v_1, v_n\}$, and it maps exactly one such state. However both $v_1$ and $v_n$ are mapped by $a$ to $Q_n \setminus \{v_1, v_n\}$. Hence removing the letter $b$ results in a non-synchronizing automaton. Removing the letter $c$ makes $v_1$ unreachable from the other states, hence no pairs can be synchronized except $(v_1, v_n)$ and $(v_1, v_2)$. \hfill \square

Our computation did not show any other possible series of slowly synchronizing automata with such large reset lengths. So we formulate the following:

\textit{Conjecture 1.} Up to isomorphism, there exist exactly two synchronizing irreducible ternary automata of size $n \geq 9$ with the reset length $\geq n^2 - 3n + 2$, namely $\mathcal{M}_n$ and $\mathcal{M}_n'$. For $k > 3$, each synchronizing irreducible $k$-ary automaton of size $n \geq 9$ has the reset length less than $n^2 - 3n + 2$.

We note that, for $n < 9$, $k > 2$ there exist other irreducible synchronizing automata with the reset length equal or greater than $n^2 - 3n + 2$; they however do not generalize to a series.

### 4.2 Gaps in the Sequence of Possible Reset Lengths

Trahtman \cite{22} has observed that for each $n = 6, \ldots, 10$ there is an integer $m < (n-1)^2$ such that no binary automaton of with $n$ states has a reset length in $[m, (n-1)^2 - 1]$. He observed also that for $n \leq 7$ automata on 3 or 4 letters do not fill this gap. In \cite{18} Ananichev, Gusev and Volkov gave arguments allowing to conjecture that for $n \geq 7$ it is the interval $[n^2 - 3n + 5, n^2 - 2n]$ forming a gap in the sequence of possible reset lengths of binary automata. They have observed also that for $n = 9$ the second gap appears: \cite{53, 55}. Recently \cite{18} we have confirmed the existence of both two gaps for $n \leq 11$ and discovered the existence of the third gap, \cite{78, 79}, for binary automata with $n = 11$ states. It seems, on the one hand, that with growing $n$ there should be more and more gaps in the sequence of possible reset lengths for binary automata. But it is still far from clear what can be the role of automata with more than 2 letters here.

We have partial experimental results concerning gaps for irreducible automata of arity $k \geq 3$. They suggest that the maximal reset length for $k$-ary automata decreases as $k$ is growing (for example, see Table \cite{3}). Yet, we have discovered an automaton with $n = 9$ states on the ternary alphabet with reset length 53, while there is no binary automaton with the same reset length; it falls in the second gap discovered in \cite{2}. Also we have found a ternary automaton with $n = 11$ states with reset length 79, falling in the third gap discovered in \cite{18}. Moreover, for any considered arity for $n \leq 8$ we found that there is only one gap in the sequence of possible reset lengths of irreducible synchronizing automata (of course, in this case the sequence starts from $k$). It seems we need more experiments to state any reasonable conjecture here.
Many partial results towards the solution of the Černý conjecture use the so-called extension method \[10,16,21,8\]. This method is to look for a sequence of (short) words $w_1, \ldots, w_{n-1}$ such that $S_1 = \{q\}$ for some $q \in Q$, and $S_{k+1} = S_k w_k^{-1}$ with $|S_{k+1}| \geq |S_k|$ for $k \geq 1$. Then, necessarily, $S_n = Q$, and the word $w_{n-1} \ldots w_2 w_1$ is synchronizing. In other words, we try to extend subsets of $Q$, starting from a singleton, rather than to compress subsets starting from $Q$. In connection with this method we say that $S \subseteq Q$ is $m$-extendable if there is a word $w$ of length at most $m$ with $|S w^{-1}| \geq |S|$. For example, for automata whose underlying digraph is Eulerian, Kari \[10\] proved that any subset is $n$-extendable. This implies that such automata satisfy the Černý conjecture.

Unfortunately, it is not true, in general, that each subset in each synchronizing automaton is $n$-extendable. In \[7\] Berlinkov constructed a series of synchronizing automata with subsets that are not $cn$-extendable for any constant $c < 2$. However, it remained open whether each subset (in a strongly connected synchronizing automaton) is $2n$-extendable. Importance of this question comes from the fact, that it would imply a quadratic upper bound for the length of the shortest reset words, and many proofs of the Černý conjecture for special cases rely on the fact that subsets are extendable by short words. Here we present a series of automata, whose subsets are not $cn$-extendable for any constant $c$.

For $m \geq 3$ let $n = 2m - 1$. Let $\mathcal{A}_n = (Q_n, \{a, b\}, \delta_n)$ be the automaton shown in Figure 2. $Q_n = \{v_1, \ldots, v_{2m-1}\}$ and $\delta_n$ is defined as follows:

$$
\delta(v_i, a) = \begin{cases} 
v_1, & \text{if } i = m, \\
v_{m+1}, & \text{if } i = 2m - 1, \\
v_{i+1}, & \text{otherwise,}
\end{cases}$$

$$
\delta(v_i, b) = \begin{cases} 
v_i, & \text{if } 1 \leq i \leq m - 1, \\
v_{2m-1}, & \text{if } i = m, \\
v_i-m, & \text{otherwise.}
\end{cases}
$$

$\mathcal{A}_n$ is strongly connected as is easily verified. Define $Q_U = \{v_{m+1}, \ldots, v_{2m-1}\}$ and $Q_D = \{v_1, \ldots, v_m\}$. We show that $\mathcal{A}_n$ is synchronizing by a word of length $2m^2 - 2m + 2 = (n^2 + 3)/2$.

**Proposition 7.** The word $aba^m b(a^{m-1} ba^m b)^{m-2}$ synchronizes $\mathcal{A}_n$ to the state $v_1$.

**Proof.** First observe that $Qba^m b = Q_D \setminus \{v_m\}$. It suffices to ensure that $S = \{v_1, \ldots, v_i\}$ for $i = 2, \ldots, m - 1$ then $Sa^{m-1} ba^m b = \{v_1, \ldots, v_{i-1}\}$. Indeed, $Sa^{m-1} = \{v_1, \ldots, v_{i-1}, v_m\}$; then $v_m b a^m b = v_1$ and $\{v_1, \ldots, v_{i-1}\}$ are mapped by the action of $ba^m b$ to itself.

We prove that there are subsets, for which the shortest words extending them have length $\Omega(m^2)$. For any $T \subseteq Q$ we say that $v_i \in T \cap Q_U$ is covered if $v_i = v_{2m-1}$ and $v_m \in T$, or $v_i \neq v_{2m-1}$ and $v_{i-m} \in T$.

**Lemma 3.** Let $S \subseteq Q$ be a non-empty set, whose states from $S \cap Q_U$ are all covered. Then a shortest reset word $w$ such that $|Sw^{-1} \cap Q_D| > |S|$ has length at least $m$.

| $k \setminus \ell$ | 27  | 28  | 29  | 30  | 31  | 32  | 33  | 34  | 35  | 36  |
|-----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 2               | 123 | 75  | 39  | 31  | 3  | 0  | 0  | 0  | 0  | 0  |
| 3               | 2752| 951 | 257 | 29  | 2  | 0  | 0  | 0  | 0  | 0  |
| 4               | 3835| 829 | 145 | 8   | 0  | 0  | 0  | 0  | 0  | 0  |
Proof. Let \( w = a_1 \ldots a_k \) be a shortest word of length \( k \) such that \( |S w^{-1} \cap Q_D| > |S| \). Let \( S_i = S a_k^{-1} \ldots a_i^{-1} \) for \( i = 1, \ldots, k + 1 \). So \( S_1 = S w^{-1} \) and \( S_{k+1} = S \).

Clearly \( a_1 = b \). Since the transformation of \( b \) maps only one state from \( Q_D \) to \( Q_U \), namely \( v_m b = v_{2m-1} \), \( S_{k-1} \) must contain \( v_{2m-1} \) and must not contain \( v_m \).

Also we have that \( S_2 \neq S \) and has non-covered state \( v_{2m-1} \). For any \( T \subseteq Q, T b^{-2} \subseteq T \); from \( w \) is a shortest word it follows that it does not have subword \( b b \). Hence \( a_2 = a \). Now \( S_3 \) contains non-covered state \( v_{m+1} \); hence \( a_2 = a \), as any preimage under \( b \) does not contain a uncovered state other than \( v_{2m-1} \). So \( S_4 \) contains uncovered state \( v_{m+2} \). Following this argument again \( 2m - 1 - (m + 2) = m - 3 \) times we have that \( a_i = a \) for \( 2 \leq i \leq m - 1 \). Then \( S_{m+1} \) contains \( v_{2m-1} \). Hence \( |w| = k \geq m \).

\[ \square \]

**Theorem 5.** Let \( A_n \) be the automaton defined above, \( n = 2m - 1 \) and \( T = Q_U \). If \( u \) is a word of the shortest length such that \( |T u^{-1}| > |T| \), then \( |u| \geq 2 + m[(m - 4)/2] \). In particular, for any \( c > 0 \), if \( n \) is large enough, then \( T \) is not \( c \)-extendable.

\[ \square \]

Proof. Let \( u = a_1 \ldots a_k \) be a shortest word such that \( |T u^{-1}| > |T| \). Let \( T_i = T a_k^{-1} \ldots a_i^{-1} \) for \( i = 1, \ldots, k + 1 \). Obviously \( u \) ends with \( b \): \( a_k = b \). Then \( T_k = T b^{-1} = \{v_m\} \).

For a subset \( X \subseteq Q \) we define \( d_X = |\{X \cap Q_D\}| \). Observe that \( |X a^{-1}| = |X| \) and \( |X b^{-1}| \leq 2d_X + 1 \). Also \( d_{X a^{-1}} = d_X \) and \( d_{X b^{-1}} \leq d_X + 1 \).

From \( |T_1| = |T| = m - 1 \) we have that \( m - 1 \leq 2d_{T_2} + 1 \), and so \( d_{T_2} \geq (m - 2)/2 \). Hence the word \( v = a_2 \ldots a_{k-1} \) is so that \( T_k v^{-1} = T_2 \).

We can split \( v \) into the shortest parts \( v_1, \ldots, v_\ell \) extending the number of states in \( Q_D \), that is \( d_{T_i} > d_{T_{i+1}} \) for \( i = 1, \ldots, \ell \), where \( T'_i = T_k v^{-1} \ldots v_i^{-1} \). Since they are the shortest, each \( v_i \) ends with \( b \) and extends the number of states in \( Q_D \) by 1. Hence all the states in each \( T'_i \cap Q_U \) are covered.

By applying Lemma 3 for each \( T'_i \) we have that \( |v_i| \geq m \). Since \( v \) extends the number of states in \( Q_D \) by at least \( (m - 2)/2 - 1 \), we have that \( \ell \geq [(m - 4)/2] \). Thus \( |u| \geq 2 + m[(m - 4)/2] \).

\[ \square \]

From Theorem 5 we have that \( Q_U \) requires a word of length \( \Omega(m^2) \) to be extended, and by Proposition 2 a word of length \( \Theta(m^2) \) does the job. Hence, if the Černý conjecture is true, then this is asymptotically optimal lower bound. In fact, any subset \( S \) is \( \Theta(m^2) \)-extendable if \( |S \cap Q_U| - |S \cap Q_D| \) is \( \Omega(m) \). A similar series can be defined for each even \( n \geq 6 \) (see Figure 3).
4.4 The Generalized Černý Conjecture

Pin [20] proposed the following generalization of the Černý Conjecture: For every $0 < k, n$, if there is a word of rank $\leq n - k$, then there is such a word of length $\leq k^2$. Pin proved this for $k \leq 3$. However, Kari [15] found a celebrated counterexample (Figure 4) to this conjecture for $k = 4$, which is a binary automaton $K$ with 6 states. In view of this, a modification of this generalized conjecture was proposed restricting it to $k$ being the rank of the considered automaton (see for example [1]). However this seems to be a quite radical restriction.

In our computations, we have found no other counterexample to Pin’s conjecture except for trivial extensions and modifications. This may suggest that Kari construction works due to the number of involved states small enough, and is, in fact, an exception. By a trivial extension of an
automaton over alphabet $\Sigma$ we mean one obtained by adding letters to $\Sigma$ that acts either as the identity transformation or as any letter in $\Sigma$. So a trivial extension has the same number of the states and the transition semigroup, and trivial extensions of the Kari automaton $K$ are counterexamples to the Pin’s conjecture, for $k = 4$, as well. By a disjoint union of two automata $A = (Q, \Sigma, \delta)$ and $A' = (Q', \Sigma, \delta')$ we mean the construction when the automata have the same alphabet $\Sigma$, and disjoint sets of states $Q, Q'$, and the union is simply $A = (Q \cup Q', \Sigma, \delta \cup \delta')$. If we take a disjoint union of $K$ with any permutation automaton (one whose letters act like permutations, or in other words, one of rank equal to its size), then again we get a counterexample to the Pin’s conjecture, for $k = 4$. Yet, in all these automata the failure is caused by the same Kari construction on the set of the 6 states. We have discovered no other counterexample. This may be treated as an experimental evidence for the conjecture we state below. Consider the smallest class of automata containing $K$ and closed on taking trivial extension and disjoint union with permutation automata. Let us call automata in this class Kari-like automata (for an example, see Figure 4). Then we have

Conjecture 2. For every $k$, if an automaton $A$ has a word of rank at most $n - k$, then there is such a word of length at most $k^2$, unless $A$ is a Kari-like automaton and $k = 4$ (in which case there is a word of rank $n - 4$ of length $k^2 + 1 = 17$).

By our computations, this conjecture has been verified for all $k$-ary automata of size $n$ for $n \leq 10$ and $k = 2$, $n \leq 8$ and $k \leq 3$, and $n = 6$, $k = 5$.

4.5 Aperiodic Synchronizing Automata

In [23] Volkov mentioned that although a quadratic upper bound for the reset length of aperiodic synchronizing automata has been proved, the largest reset length for known aperiodic automata does not exceed $n + \lfloor n/2 \rfloor - 2$ for $n \geq 7$. This length is reached by a series of binary automata constructed by Ananichev [4]. In this connection it is worth noting that we have verified that $n + \lfloor n/2 \rfloor - 2$ is the largest reset length of aperiodic automata for $1 < n \leq 10$, $k = 2$, and $1 < n \leq 7$, $k = 3$. Since this verification involves a huge number of aperiodic automata we are fairly sure that the following counterpart of the Černý conjecture holds.

Conjecture 3. Every synchronizing aperiodic automaton $A$ with $n > 1$ states has a reset word of length at most $n + \lfloor n/2 \rfloor - 2$.

In this connection, it may be also interesting to know that we have discovered a series of irreducible ternary aperiodic automata that reaches the bound in the conjecture. They have a quite simple definition: $Q = \{0, 1, \ldots, n - 1\}$, $\Sigma = \{a, b, c\}$, $\delta(0, a) = i + 1$ for $0 \leq i < n$, $\delta(i, b) = i - 1$ for $1 < i \leq n$, $\delta(\lfloor n/2 \rfloor, c) = 0$, and $\delta(i, x) = i$, otherwise ($x \in \Sigma$).

Acknowledgments. The computations were performed on grid that belongs to Institute of Computer Science of Jagiellonian University. We thank Jakub Kowalski for a help in parallelization of our computations, and Jakub Tarnawski for computing the values of $D(m, k)$ for small $m, k$.

References

1. J. Almeida and B. Steinberg. Matrix Mortality and the Černý-Pin Conjecture. In Developments in Language Theory, volume 5583 of LNCS, pages 67–80. Springer, 2009.
2. D. Ananichev, V. Gusev, and M. Volkov. Slowly synchronizing automata and digraphs. In *Mathematical Foundations of Computer Science 2010*, volume 6281 of LNCS, pages 55–65. Springer, 2010.

3. D. Ananichev, V. Gusev, and M. Volkov. Primitive digraphs with large exponents and slowly synchronizing automata. *Zapiski Nauchnyh Seminarov POMI [Kombinatorika i Teorija Grafov. IV]*, 402:9–39, 2012. In Russian.

4. D. S. Ananichev. The Mortality Threshold for Partially Monotonic Automata. In *Developments in Language Theory*, volume 3572 of LNCS, pages 112–121. Springer, 2005.

5. M.-P. Béal, M. V. Berlinkov, and D. Perrin. A quadratic upper bound on the size of a synchronizing word in one-cluster automata. *International Journal of Foundations of Computer Science*, 22(2):277–288, 2011.

6. M.-P. Béal and D. Perrin. A quadratic upper bound on the size of a synchronizing word in one-cluster automata. In *Developments in Language Theory*, volume 6281 of LNCS, pages 55–65. Springer, 2010.

7. M. V. Berlinkov. On a Conjecture by Carpi and D’Alessandro. *International Journal of Foundations of Computer Science*, 22(7):1565–1576, 2011.

8. M. V. Berlinkov. Synchronizing Quasi-Eulerian and Quasi-One-Cluster Automata. *International Journal of Foundations of Computer Science*, 24(6):729–745, 2013.

9. J. Černý. Poznámka k homogénym eksperimentom s konečnými automatami. *Matematicko-fyzikálny Časopis Slovenskej Akadémie Vied*, 14(3):208–216, 1964. In Slovak.

10. L. Dubuc. Sur les automates circulaires et la conjecture de Černý. *Informatique théorique et applications*, 32:21–34, 1998.

11. P. Frankl. An extremal problem for two families of sets. *European Journal of Combinatorics*, 3:125–127, 1982.

12. V. Froidure and J.-E. Pin. Algorithms for computing finite semigroups. In *Foundations of Computational Mathematics*, pages 112–126. Springer, 1997.

13. M. Grech and A. Kisielewicz. The Černý conjecture for automata respecting intervals of a directed graph. *Discrete Mathematics and Theoretical Computer Science*, 15(3):61–72, 2013.

14. A. W. Ingleton. The rank of circulant matrices. *Journal of the London Mathematical Society*, 31(4):445–460, 1956.

15. J. Kari. A counter example to a conjecture concerning synchronizing word in finite. *EATCS Bulletin*, 73:146–147, 2001.

16. J. Kari. Synchronizing finite automata on Eulerian digraphs. *Theoretical Computer Science*, 295(1-3):223–232, 2003.

17. A. Kisielewicz, J. Kowalski, and M. Szykuła. Computing the shortest reset words of synchronizing automata. *Journal of Combinatorial Optimization*, 2013.

18. A. Kisielewicz and M. Szykuła. Generating small automata and the Černý conjecture. In *Implementation and Application of Automata*, volume 7982 of LNCS, pages 340–348. Springer, 2013.

19. J.-E. Pin. Utilisation de l’algèbre linéaire en théorie des automates. In *Actes du 1er Colloque AFCET-SMF de Mathématiques Appliquées II*, AFCET, pages 85–92, 1972.

20. J.-E. Pin. On two combinatorial problems arising from automata theory. In *Proceedings of the International Colloquium on Graph Theory and Combinatorics*, volume 75 of *North-Holland Mathematics Studies*, pages 535–548, 1983.

21. B. Steinberg. The Černý conjecture for one-cluster automata with prime length cycle. *Theoretical Computer Science*, 412(39):5487–5497, 2011.

22. A. N. Trahtman. An efficient algorithm finds noticeable trends and examples concerning the Černý conjecture. In *Mathematical Foundations of Computer Science*, volume 4162 of LNCS, pages 789–800. Springer, 2006.

23. M. Volkov. Synchronizing automata and the Černý conjecture. In *Language and Automata Theory and Applications*, volume 5196 of LNCS, pages 11–27. Springer, 2008.