



HAL
open science

On k-abelian palindromes

Julien Cassaigne, Juhani Karhumaki, Svetlana Puzynina

► **To cite this version:**

Julien Cassaigne, Juhani Karhumaki, Svetlana Puzynina. On k-abelian palindromes. *Information and Computation*, 2018, 260, pp.89-98. 10.1016/j.ic.2018.04.001 . hal-02060198

HAL Id: hal-02060198

<https://hal.science/hal-02060198>

Submitted on 2 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On k -abelian palindromes

Julien Cassaigne^a, Juhani Karhumäki^b, Svetlana Puzynina^{c,d,1}

^a*Institut de mathématiques de Marseille, France*

^b*Department of Mathematics and Statistics, University of Turku, Finland*

^c*LIP, ENS de Lyon, Université de Lyon, France*

^d*Sobolev Institute of Mathematics, Novosibirsk, Russia*

Abstract

A word is called a *palindrome* if it is equal to its reversal. In the paper we consider a k -abelian modification of this notion. Two words are called *k -abelian equivalent* if they contain the same number of occurrences of each factor of length at most k . We say that a word is a *k -abelian palindrome* if it is k -abelian equivalent to its reversal. A question we deal with is the following: how many distinct palindromes can a word contain? It is well known that a word of length n can contain at most $n + 1$ distinct palindromes as its factors; such words are called *rich*. On the other hand, there exist infinite words containing only finitely many distinct palindromes as their factors; such words are called *poor*. It is easy to see that there are no 1-abelian poor words, and there exist words containing $\Theta(n^2)$ distinct 1-abelian palindromes. We analyze these notions with respect to k -abelian equivalence. Our main results concern poor words: We show that in the k -abelian case there exist infinite words containing finitely many distinct k -abelian palindromic factors. We also make some observation concerning rich words, namely, we show that there exist finite rich words containing $\Theta(n^2)$ distinct k -abelian palindromes as their factors. Therefore, for poor words the situation resembles that of usual palindromes, while for rich words it is similar to the 1-abelian case.

Keywords: infinite words, palindromes, rich words, k -abelian equivalence

Email addresses: julien.cassaigne@math.cnrs.fr (Julien Cassaigne),
karhumak@utu.fi (Juhani Karhumäki), s.puzynina@gmail.com (Svetlana Puzynina)

¹Corresponding author.

1. Introduction

The palindromicity of words is a widely studied area in formal languages. When a model of a computation is introduced, among the first questions is to ask whether the set of palindromes (or some infinite subset of it) can be recognized by the model. In other words, can the model identify whether it is irrelevant if words are read from left to right or from right to left? It is folklore that deterministic finite automata cannot do that. On the other hand it is among the simplest tasks for push-down automata, or on-line log-space Turing machines. A slightly different approach is to look at palindromic factors of words. They can be viewed as measuring how much the word is locally independent of the reading direction of a factor. The notion of palindromic complexity for infinite words was introduced in [12], and further formalized in [2, 9]. It has been studied extensively ever since.

A problem related to our question of counting palindromes in a word is the problem of counting maximal repetitions in a word of length n , that is, runs in a word. It was shown in [23] that the maximal number of runs in a word is linear in n . Subsequently, there was a lot of research performed to find the bound [11], which led to a conjecture that this number is less than n . Recently, the conjecture has been proved with a remarkably simple argument, considering numerous attempts to solve it [7]. Not only runs, but also various other questions concerning counting squares in a word have been considered, see, e.g., [17, 18, 24].

We recall that a word is a *palindrome* if it is equal to its reversal. It is well known that the maximal number of palindromes a word of length n can contain is equal to $n + 1$, and such words are called *rich* in palindromes [14]. In some papers the same class of words was called *full* words (see, e.g., [4, 9]). Lately, there is an extensive number of papers devoted to the study of rich words and their generalizations (see, e.g., [10, 16]). This notion can be extended to infinite words: an infinite word is rich if each of its factors is rich. For example, Sturmian words are known to be rich. Note also that Sturmian words can be characterized via palindromic closures [13].

Recently the notion of palindromic poorness has been considered in [8, 15]. Namely, an infinite word is called *poor* in palindromes if it contains only finitely many distinct palindromes. In particular, it has been shown that there exist poor words with the set of factors closed under reversal. Besides that, in [15] the authors found the minimal number of palindromes an infinite word satisfying different conditions (uniform recurrence, closed

under reversal, etc.) can contain. In a related paper [27] words avoiding reversed subwords were studied.

In this paper the k -abelian version of the notion of a palindrome is studied. Two words are called *abelian equivalent* if they contain the same number of occurrences of each letter, or, equivalently, if they are permutations of each other. In the recent years there is a growing interest in abelian properties of words, as well as modifications of the notion of abelian equivalence [1, 6, 22, 26, 29]. One such modification is the notion of k -abelian equivalence: two words are called *k -abelian equivalent* if they contain the same number of occurrences of each factor of length at most k . For $k = 1$, the notion of k -abelian equivalence coincides with the notion of abelian equivalence, and when k is greater than half of the length of the words, k -abelian equivalence means equality. Therefore, the notion of k -abelian equivalence is an intermediate notion between abelian equivalence and equality of words. For more on k -abelian equivalence we refer to [20, 21].

In analogy with normal palindromes, we say that a word v is a *k -abelian palindrome* if its reversal is k -abelian equivalent to v . For example, the word *aabaaabbaa* is a 3-abelian palindrome. We are interested in the maximal and minimal numbers of k -abelian palindromes a word can contain.

For $k = 1$, clearly, each word is an abelian palindrome, since it is abelian equivalent to its reversal. Therefore, there are no infinite 1-abelian poor words. But for $k > 1$ this no longer holds. We build infinite k -abelian poor words for $k > 1$ and sufficiently large alphabets. In fact, we provide a complete characterization of pairs (k, Σ) for which k -abelian poor words over the alphabet Σ exist (see Theorem 1).

Since a word of length n contains at most $1 + \frac{n(n+1)}{2}$ factors in total, a k -abelian rich word of length n cannot contain more than $\Theta(n^2)$ k -abelian palindromes. However, we show that it can indeed contain $\Theta(n^2)$ inequivalent k -abelian palindromes, where the constants in $\Theta(n^2)$ depend on k (see Theorem 2).

The minimal and maximal numbers of inequivalent palindromes in the case of equality, k -abelian equality and abelian equality are summarized in Table 1 (here C is a constant). We remark that in the minimal case, that is for poor words, infinite words are considered, while in the maximal case, that is for rich words, only finite words are considered. The message of the table is that in the big picture k -abelian equivalence behaves like equality for poor words, while it behaves like abelian equivalence for rich words.

	equality	k -abelian	abelian
poor	C	C	∞
rich	$n + 1$	$\Theta(n^2)$	$\Theta(n^2)$

Table 1: Minimal and maximal numbers of palindromes in the case of equality, abelian and k -abelian equivalence.

2. Definitions and notation

Given a finite non-empty set Σ (called the alphabet), we let Σ^* and Σ^ω , respectively, denote the set of finite words and the set of (right) infinite words over the alphabet Σ . We will always assume $|\Sigma| \geq 2$. A word v is a *factor* (resp., a *prefix*, resp., a *suffix*) of a word w , if there exist words x, y such that $w = xvy$ (resp., $w = vy$, resp., $w = xv$). The set of factors of a finite or infinite word w is denoted by $F(w)$. The prefix and suffix of length k of w are denoted by $\text{pref}_k(w)$ and $\text{suff}_k(w)$, respectively. When the length is not important, we use the notation $\text{pref}(w)$ and $\text{suff}(w)$ to denote any prefix or suffix of w . Given a finite word $u = u_1u_2 \cdots u_n$ with $n \geq 1$ and $u_i \in \Sigma$, we let $|u| = n$ denote the length of u . The empty word is denoted by ε and we set $|\varepsilon| = 0$. An infinite word is called *recurrent* if each of its factors occurs infinitely often in it. An infinite word w is called *uniformly recurrent* if for each $v \in F(w)$ there exists N such that $v \in F(w_i \cdots w_{i+N})$ for every i . In other words, in a uniformly recurrent word each factor occurs with bounded gaps.

For each $v \in \Sigma^*$, we let $|u|_v$ denote the number of occurrences of the factor v in u . Two words u and v in Σ^* are said to be *abelian equivalent*, denoted $u \sim_{\text{ab}} v$, if and only if $|u|_a = |v|_a$ for all $a \in \Sigma$. For example, the words aba and aab are abelian equivalent. Clearly, abelian equivalence is an equivalence relation on Σ^* .

Let k be a positive integer. Two words u and v are *k -abelian equivalent*, denoted by $u \sim_k v$, if $|u|_t = |v|_t$ for every word t of length at most k . This is equivalent to the following conditions:

- $|u|_t = |v|_t$ for every word t of length k ,
- $\text{pref}_{k-1}(u) = \text{pref}_{k-1}(v)$ and $\text{suff}_{k-1}(u) = \text{suff}_{k-1}(v)$ (or $u = v$, if $|u| < k - 1$ or $|v| < k - 1$).

For instance, $aabab \sim_2 abaab$, but $aabab \not\sim_2 aaabb$. It is easy to see that k -abelian equivalence implies k' -abelian equivalence for every $k' < k$. In particular, it implies abelian equivalence, that is, 1-abelian equivalence.

For a finite word $v = v_1 \cdots v_n$ we let $v^R = v_n \cdots v_1$ denote its reversal. A word v is a *palindrome* if $v = v^R$. A word is a *k -abelian palindrome* (or briefly *k -palindrome*) if $v \sim_k v^R$. The empty word ε is considered as a palindrome and a k -palindrome.

An infinite word is *k -abelian palindromic poor* (briefly *k -poor*) if it contains finitely many k -abelian palindromes.

There are obvious constructions of k -poor words. For instance, one can take $(abc)^\omega$ that contains only four k -palindromes (ε, a, b, c) for $k \geq 2$. It is more interesting to restrict to certain families of infinite words. Here we consider two such restrictions, that apply on the set of factors of the infinite word: closed under reversal and closed under k -abelian reversal (a set of words L is said to be closed under k -abelian reversal if for every $u \in L$, there exists $u' \in L$ such that $u' \sim_k u^R$).

A word of length n is called *k -abelian palindromic rich* (briefly *k -rich*), if it contains at least $n^2/4k$ inequivalent k -abelian palindromes. Notice that the total number of factors contained in a word of length n is equal to $1 + \frac{n(n+1)}{2}$. Therefore, for a fixed k , a k -abelian rich word contains the number of k -palindromes of the same order as the total number of factors when n tends to infinity.

We emphasize that for poor words we consider infinite words, and for rich words we consider finite ones, and this is caused by the nature of the problem. Indeed, for poor words, since there exist infinite words containing only finitely many palindromes, all their factors have a uniformly bounded number of palindromes. On the other hand, the closed under reversal condition is not applicable to finite words, since it would imply a growing number of palindromes. Concerning rich words, an infinite word could easily contain infinitely many palindromes, so we are interested in maximal number of palindromes in finite ones. In the next two sections we consider k -abelian poor and rich words, respectively.

3. k -abelian poor words

In this section we show that there exist k -abelian palindromic poor words. This holds for almost all values of k and $|\Sigma|$, and we characterize those.

$k \setminus \Sigma $	2	3	4	...
1	-	-	-	-
2	-	\ominus	+	+
3	-	+	+	+
4	\ominus	+	+	+
5	+	+	+	+
...	+	+	+	+

Table 2: The classification of $(k, |\Sigma|)$ for the existence of k -poor words.

- Theorem 1.** Let $S_1 = \{(1, l) | l \in \mathbb{N}\} \cup \{(2, 2), (3, 2)\}$, $S_2 = \{(2, 3), (4, 2)\}$.
- I. For $(k, |\Sigma|) \notin S_1 \cup S_2$ there exist uniformly recurrent k -abelian palindromic poor words over Σ having a set of factors that is closed under reversal.
 - II. For $(k, |\Sigma|) \in S_1$ there are no k -abelian palindromic poor words over Σ .
 - III. For $(k, |\Sigma|) \in S_2$ there exist k -abelian poor words, but there are no k -abelian palindromic poor words over Σ having a set of factors that is closed under k -abelian reversal.

We emphasize that the theorem gives complete classification of pairs (k, l) for existence of k -poor words for the three cases: the set of factors closed under reversal, the set of factors closed under k -abelian reversal, and without any restrictions.

The results can be summarized in Table 2. Here + means that there exist k -abelian poor words having a set of factors that is closed under reversal over an alphabet Σ , - indicates that there are no k -abelian poor words over Σ , and \ominus means that there exist k -abelian poor words, but only with a set of factors that is not closed under k -abelian reversal. In what follows, we will write (k, l) -poor words for k -abelian poor words over an alphabet of cardinality l for brevity.

Proof. First we prove Part I of the theorem by providing constructions of poor words, and then prove the non-existence for Parts II and III of the theorem.

3.1. Proof of Part I

We remark that the existence of a (k, l) -poor word implies the existence of a (k', l') -poor word for each $k' \geq k$ and $l' \geq l$. Indeed, for $l' > l$ to build a (k, l') -poor word from a (k, l) -poor word one could split any letter into several letters in any way (i.e., for a chosen letter a , some occurrences of a are substituted by one of the $l' - l$ new letters). The word remains k -poor, and closure under reversal condition and uniform recurrence can be preserved. For $k' > k$, the statement follows from the fact that every k' -abelian palindrome is also a k -abelian palindrome for any $k \leq k'$. Therefore, it is enough to build $(5, 2)$ -, $(3, 3)$ - and $(2, 4)$ -poor words.

To construct uniformly recurrent k -abelian poor words we will use a self-avoiding fractal curve (see Fig. 1).

Let $\Delta = \{A, B, C, D\}$. Each letter in this alphabet is assigned a drawing instruction; here the arrow denotes a line segment of length 1 in the direction of the arrow:

$$\begin{aligned} A & : \rightarrow \\ B & : \uparrow \\ C & : \leftarrow \\ D & : \downarrow \end{aligned}$$

An infinite word $w = w_1 w_2 \cdots \in \Delta^\omega$ can be translated into a polygonal line visiting points of the lattice \mathbb{Z}^2 by interpreting letters of w as drawing instructions. We start at the origin $(x_0, y_0) = (0, 0)$. At step n , we are at a point (x_{n-1}, y_{n-1}) and we draw a line segment corresponding to the letter w_n , so that we come to a point (x_n, y_n) . For example, for the letter A , we draw a horizontal line segment of length one “to the right”, so that $(x_n, y_n) = (x_{n-1} + 1, y_{n-1})$.

It will be convenient to represent points with complex numbers. Let $z_n = x_n + iy_n$. Define the morphisms $\rho : \Delta^* \rightarrow \Delta^*$ and $Z : \Delta^* \rightarrow (\mathbb{Z}[i], +)$ by $\rho(A) = B$, $\rho(B) = C$, $\rho(C) = D$, $\rho(D) = A$, $Z(A) = 1$, $Z(B) = i$, $Z(C) = -1$, $Z(D) = -i$. Then $Z(\rho(u)) = iZ(u)$: the morphism ρ rotates a curve by 90 degrees counterclockwise. Now the points z_n can be expressed as $z_n = Z(\text{pref}_n(w))$.

Lemma 1. *There exists a uniformly recurrent word $w \in \Delta^\omega$ with the following properties:*

- (i) $F(w)$ is closed under the map $u \mapsto \rho^2(u^R)$;
- (ii) consecutive letters of w correspond to orthogonal segments, i.e., for all n , $w_{n+1} \in \{\rho(w_n), \rho^{-1}(w_n)\}$;
- (iii) the curve associated to w is self-avoiding, i.e., the points z_n are distinct.

Proof. We construct one such word w as the fixed point starting with A of a morphism φ :

$$\varphi: \begin{array}{l} A \mapsto ABA \\ B \mapsto BCB \\ C \mapsto CDC \\ D \mapsto DAD \end{array} .$$

Alternatively, w can be obtained as the limit of a sequence of finite words (u_n) defined by a recurrence formula:

$$\begin{array}{l} u_0 = A \\ u_{n+1} = u_n \rho(u_n) u_n \end{array} .$$

This follows from the identity $u_n = \varphi^n(A)$, which is obtained by a straightforward induction, using the fact that φ and ρ commute.

One has

$$w = ABABC BABABCBCDCBCBABABC BABABCBCDCBCB \dots$$

and a part of the curve associated with w is represented on Fig. 1.

As φ is a primitive morphism, the word w is uniformly recurrent [3, Theorem 10.9.5]. It is clear from the recurrence formula that the words u_n are palindromes, and this can also be deduced from the fact that φ preserves palindromicity. Therefore the set of factors of w is closed under reversal. It is also closed under ρ , as $\rho(u_n)$ occurs in u_{n+1} . Property (i) follows. Property (ii) is easily proved by induction. We now prove Property (iii) by contradiction.

Assume that $z_m = z_n$ for some $m < n$, and take such a pair (m, n) with $n - m$ as small as possible. It is clear that $n - m \neq 1$, as $|z_{m+1} - z_m| = 1$, and also that $n - m \neq 2$, as $|z_{m+2} - z_m| = \sqrt{2}$ by Property (ii). Therefore $n - m \geq 3$.

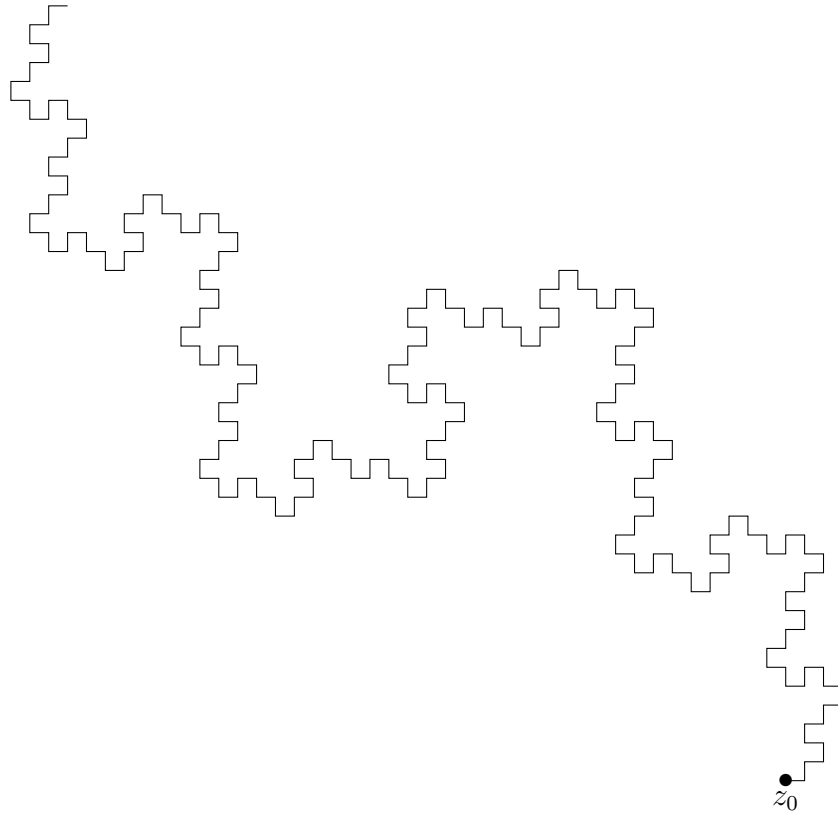


Figure 1: The self-avoiding curve corresponding to the word u_5 with the drawing instructions $A: \rightarrow, B: \uparrow, C: \leftarrow, D: \downarrow$

Let m' and n' be the integers closest to $m/3$ and $n/3$, respectively. Then $|m - 3m'| \leq 1$ and $|n - 3n'| \leq 1$, so that $|z_m - z_{3m'}| \leq 1$ and $|z_n - z_{3n'}| \leq 1$. We have:

$$|z_{3n'} - z_{3m'}| = |(z_{3n'} - z_n) + (z_m - z_{3m'})| \leq |z_{3n'} - z_n| + |z_m - z_{3m'}| \leq 2.$$

Now observe that $Z(\varphi(u)) = (2+i)Z(u)$ (it is sufficient to check it when u is a letter). It follows that $z_{3m'} = (2+i)z_{m'}$ and $z_{3n'} = (2+i)z_{n'}$, so that

$$|z_{n'} - z_{m'}| = \frac{|z_{3n'} - z_{3m'}|}{|2+i|} \leq \frac{2}{\sqrt{5}} < 1.$$

As $z_{m'}$ and $z_{n'}$ are lattice points, it follows that $z_{m'} = z_{n'}$. But

$$0 < \frac{n-m-2}{3} \leq n' - m' \leq \frac{n-m+2}{3} < n-m,$$

in contradiction with the minimality of $n-m$. \square

Now, in order to build k -abelian poor words, we apply morphisms to the word w constructed in Lemma 1 (for each of the cases (5, 2), (3, 3) and (2, 4) we provide a morphism):

$$\tau_{2,4} : \begin{cases} A \mapsto ab \\ B \mapsto cd \\ C \mapsto ba \\ D \mapsto dc \end{cases} \quad \tau_{3,3} : \begin{cases} A \mapsto abcca \\ B \mapsto abbca \\ C \mapsto accba \\ D \mapsto acbba \end{cases} \quad \tau_{5,2} : \begin{cases} A \mapsto aabbaabaa \\ B \mapsto aabbabaaa \\ C \mapsto aabaabbaa \\ D \mapsto aaababbaa \end{cases}.$$

Since w is uniformly recurrent, the words $\tau_{k,l}(w)$ are uniformly recurrent. Observe that, in all three cases, $\tau_{k,l}(C) = (\tau_{k,l}(A))^R$ and $\tau_{k,l}(D) = (\tau_{k,l}(B))^R$. It follows that $\tau_{k,l}(\rho^2(u^R)) = (\tau_{k,l}(u))^R$ for every word $u \in \Delta^*$. Since, by Property (i) of Lemma 1, $F(w)$ is closed under $u \mapsto \rho^2(u^R)$, we conclude that the sets of factors of the words $\tau_{k,l}(w)$ are closed under reversal.

We will now prove that $\tau_{k,l}(w)$ is k -abelian poor over an l -letter alphabet. The proofs are similar for the three cases.

For the case (2, 4), suppose that $\tau_{2,4}(w)$ contains a 2-abelian palindrome v of length at least 3. Take a factor v' inside v which is a full image of $\tau_{2,4}$ (i.e., possibly you have to crop a prefix and a suffix of v of length at most 1). By Property (ii) of Lemma 1, consecutive letters of w are never equal, so that the

factors ab , ba , cd and dc occur only as images of letters of w (and not across two images), so v' contains the same number of occurrences of ab (resp., ba , cd , dc) as v . Since v is a 2-palindrome, then v' contains the same number of occurrences of factors ab and ba , and the same number of occurrences of factors cd and dc . So, the corresponding factor $\tau_{2,4}^{-1}(v')$ of w contains the same number of occurrences of A and C , and the same number of occurrences of B and D . This gives a self-intersection of the corresponding curve, which contradicts Property (iii) of Lemma 1 (note that $v' \neq \varepsilon$ as $|v| \geq 3$).

For the case (3, 3), suppose that $\tau_{3,3}(w)$ contains a 2-abelian palindrome v of length at least 7. Take a factor v' of v which is a full image of $\tau_{3,3}$ and contains the same middle 3-factors bcc , bbc , cbb , ccb as v (i.e., possibly you have to crop a prefix and a suffix of v of length at most 3 or extend it with a letter a to the left or to the right). As in the proof of (4, 2) case, the factors bcc , bbc , ccb and cbb occur only inside the images of letters of w , so v' contains the same number of occurrences of bcc (resp., bbc , ccb , cbb) as v . Since v is a 3-palindrome, then v' contains the same number of occurrences of factors bcc and ccb , and the same number of occurrences of factors bbc and cbb . So, the corresponding factor $\tau_{3,3}^{-1}(v')$ of w contains the same number of occurrences of A and C , and the same number of occurrences of B and D . This gives a self-intersection of the corresponding curve, which contradicts Property (iii) of Lemma 1 (note that $v' \neq \varepsilon$ as $|v| \geq 7$).

The proof for the case (5, 2) is essentially the same as the proof for the case (3, 3), using middle 5-factors instead of 3-factors, so we omit the details.

3.2. Proof of Part II

Assume now that $(k, |\Sigma|) \in S_1 = \{(1, l) | l \in \mathbb{N}\} \cup \{(2, 2), (3, 2)\}$. Here we should prove that there are no k -abelian poor words. For $k = 1$ (i.e., the abelian equivalence) each word is an abelian palindrome, since every word is abelian equivalent to its reversal. Therefore, all factors of any infinite word are abelian palindromes, and hence there are no abelian palindromic poor words.

In the 2-abelian binary case, 2-palindromes are exactly the words starting and ending in the same letter. Indeed, the condition is necessary. Conversely, without loss of generality let a word v start and end with a , and let it contain m blocks of b 's. Then v contains m occurrences of the factor ab and m occurrences of the factor ba . Factors aa and bb do not affect 2-abelian palindromicity; hence v is a 2-palindrome. Since any infinite binary word

contains infinitely many factors starting and ending with the same letter, there are no 2-abelian poor binary infinite words.

In the 3-abelian binary case the proof is similar, just a bit more technical. We omit the details of the proof.

3.3. Proof of Part III

Finally, assume that $(k, |\Sigma|) \in \{(2, 3), (4, 2)\}$. Here we should prove that there exist k -abelian poor words, but there are no k -abelian poor words with the set of factors closed under k -abelian reversal.

The proofs of the two cases are similar, although the case $(4, 2)$ requires a more thorough analysis. We start with the case $(2, 3)$.

First we introduce rewriting rules which do not affect the 2-palindromicity:

- (1) for $x \in \Sigma$, substitute $xx \rightarrow x$,
- (2) for $x, y \in \Sigma$, substitute $xyx \rightarrow x$.

Claim (i). Let v be a word, and let v' be obtained from v by applying a rewriting rule (1) or (2). Then v is a 2-palindrome if and only if v' is a 2-palindrome.

Indeed, after applying the rewriting rule (1), the multiset (the set with multiplicities) of factors of length 2 of v' is obtained from the multiset of factors of length 2 of v by removing one factor xx . Clearly, the resulting set coincides with its reversal if and only if the original set does. After applying the rewriting rule (2), the multiset of factors of length 2 of v' is obtained from the multiset of factors of length 2 of v by removing two factors xy and yx . Again, the resulting set coincides with its reversal if and only if the original set does. The claim follows.

Now take a ternary word v and apply rewriting rules until the word does not contain factors of the form xx and xyx . We call the resulting word the *reduced form* of v . We note that the reduced form of v is unique.

The following claim is straightforward:

Claim (ii). 1. The reduced form of any ternary word is a factor of $(abc)^\infty$ or $(cba)^\infty$.

2. If u is the reduced form of a ternary word v , then the reduced form of v^R is u^R .

3. If two ternary words are 2-abelian equivalent, then they have the same reduced form.

4. A ternary word is a 2-palindrome if and only if its reduced form is empty or a letter.

Now assume that an infinite ternary word w with its set of factors closed under 2-abelian reversal does not contain 2-palindromes of length greater than N for some integer N . Take a factor $v = w_i \cdots w_{i+N}$ of length $N + 1$. Since the set of factors of w is closed under 2-abelian reversal, there exists an occurrence of $v'^R = w_j \cdots w_{j+N}$ for some $v' \sim_2 v$. Without loss of generality we can assume that $j > i$ and that the reduced form of v is a word u of the form $(abc)^m \text{pref}(abc)$ for some $m \geq 0$. Then the reduced form of v'^R equals u^R . Now consider the factor $w_i \cdots w_{j+N}$; it starts and ends with a , and so does its reduced form. Again without loss of generality, we can assume that its reduced form is of the form $(abc)^r a$ for some $r \geq 0$. We will now show that the factor $w_i \cdots w_{j+N}$ has a suffix $w_s \cdots w_{j+N}$, $s < j$, which is reduced to a , and hence is a 2-palindrome. To see that, first note that the reduced form of any non-empty binary word over $\{b, c\}$ is one of the words b, c, bc, cb . Then, consider all the suffixes of $w_i \cdots w_{j+N}$ beginning with a . The reduced form of any two consecutive such suffixes is of the form $(abc)^l a$, and the values of l differ by at most 1 (notice that there can be some suffixes of the form $(acb)^t a$, these can be considered as negative powers). So, considering consecutive prefixes starting with a , we go from nonnegative $l = r \geq 0$ to $l = -m < 0$, each time increasing or decreasing l by at most 1. So, at some point we have $l = 0$, which corresponds to a 2-palindrome. The length of this 2-palindrome is greater than N , a contradiction.

On the other hand, clearly, the word $(abc)^\omega$ is 2-abelian palindromic poor, although its set of factors is not closed under 2-abelian reversal.

Now we continue with the case (4, 2). First we introduce rewriting rules which do not affect 4-palindromicity:

for $x, y \in \Sigma$, substitute

- (1) $xxx \rightarrow xx$ (when this occurrence of xxx is not a prefix or suffix in v)
- (2) $xyyxx \rightarrow xxyxx$
- (3) $xyxyx \rightarrow xyx$
- (4) $xyxxyx \rightarrow xyx$
- (5) $xyxxy \rightarrow xxy$
- (6) $yxxyxx \rightarrow yxx$

Claim (iii). Let v be a binary word, and let v' be obtained from v by applying one of the rewriting rules (1)–(6). Then v is a 4-palindrome if and only if v' is a 4-palindrome.

Proof of Claim (iii). To prove the claim for rule 1, we need to consider a few cases. If xxx occurs as a factor of $yxxxx$, $xxxxy$ or $xxxxx$, then after

applying the rewriting rule (1), the multiset of factors of length 4 of v' is obtained from the multiset of factors of length 4 of v by removing one factor xxx . Clearly, the resulting set coincides with its reversal if and only if the original set does. If xxx occurs as a factor of $yxxy$, then after applying the rewriting rule (1), the multiset of factors of length 4 of v' is obtained from the multiset of factors of length 4 of v by removing factors $yxxy$ and $xyxy$ and adding a factor $xyxy$. Again, the resulting multiset coincides with its reversal if and only if the original one does.

The proofs for the rules (2)–(6) are simpler than the proof for the rule (1), since we do not have to consider any cases, so we just write how the multiset of factors of length 4 changes. It is straightforward to see that the resulting multiset coincides with its reversal if and only if the original one does. Besides that, those reductions keep the prefix and the suffix of length 3. Denote the multiset of factors of length 4 of a word u by $MF_4(u)$. Then:

$$(2) \quad MF_4(v') = MF_4(v) - xxyy - yyxx - xyxy + xxyx + xyxx$$

$$(3) \quad MF_4(v') = MF_4(v) - xyxy - yxyx$$

$$(4)–(6) \quad MF_4(v') = MF_4(v) - xxyx - yxyx - yxyx$$

Here, e.g., after applying the rule (2) the multiset of factors changes as follows: we remove one copy of each of the factors $xxyy$, $yyxx$, $xyxy$, and add one copy of each of the factors $xxyx$ and $xyxx$.

In all the cases the resulting set coincides with its reversal if and only if the original set does. The claim follows.

Now take a binary word v and apply rewriting rules until the word does not contain factors of the forms of the lefthandside parts of the rules (1)–(6). We call the resulting word the *reduced form* of v (we will prove later that the reduced form is indeed unique).

Claim (iv). 1. The reduced form of any binary word v is a factor of $(aababb)^\infty$ or $(bbabaa)^\infty$ (possibly starting or ending with xxx instead of xx), or a short word of the form $aaabaa$, $aabaa$, $aaabaaa$ or $aabaaaa$, up to renaming letters.

2. A binary word v is a 4-palindrome if and only if its reduced form is a palindrome of length not greater than 7.

Proof of Claim (iv). 1. The proof is summarized on Fig. 2. In this tree we show all possible words starting with a (starting with b is symmetric), and on each edge we mark the rule applied. We find out that the tree has only finitely many branches corresponding to the reduced forms. Now all possible reductions are obtained as follows: each branch can be cut in any place, and if terminated with a factor of the form xx for some $x \in \{a, b\}$,

then one extra copy of x can be added. So, the reduced form of any binary word v is a factor of $(aababb)^\infty$ or $(bbabaa)^\infty$, possibly starting or ending with xxx instead of xx , or one of the short words from the list corresponding to terminating branches.

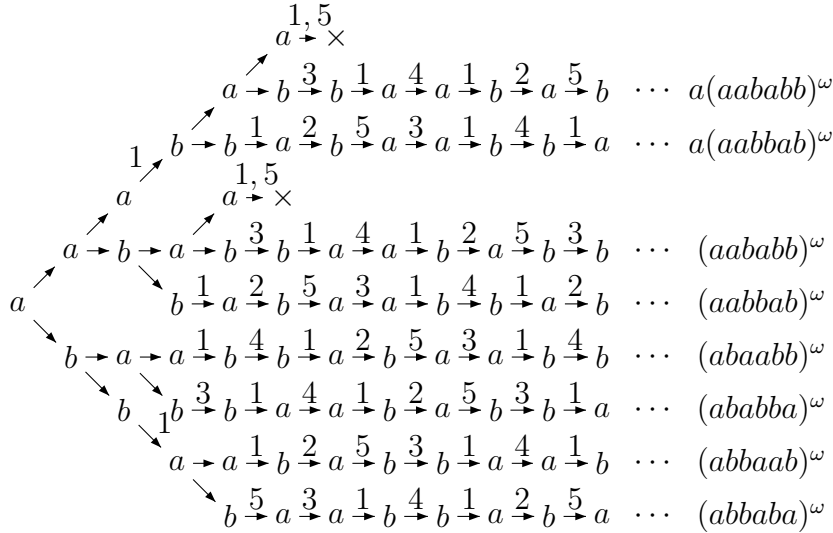


Figure 2: Reductions of words for $(k, l) = (4, 2)$.

2. Consider a 4-palindromic factor v . Without loss of generality its reduced form u is a factor of $(aababb)^\infty$, possibly starting or ending with xxx instead of xx . 4-palindromic factors of $(aababb)^\infty$ have maximal length 4, since each factor of length greater than 4 contains one of the factors $aaba$, $abab$, $babb$, $bbaa$, and no factor of $(aababb)^\infty$ contains the reversals of these factors. Possible starting or ending with xxx instead of xx increases the length of the 4-palindrome by at most 2. Considering the two terminating branches from Fig. 2, we get the maximal length of the 4-palindrome 7 (corresponding to the first terminating branch extended by a : $aaabaaa$). Any 4-palindrome of length at most 7 is a palindrome, since the suffix of length 3 is the reversal of the prefix of length 3. Claim (iv) is proved.

Claim (v). If $v \sim_4 u$, then the reduced forms of v and u coincide.

Proof of Claim (v). To prove the claim it is enough to notice that none of the rewriting rules (1)–(6) changes the prefix and the suffix of length 3, nor the difference between the number of each of factors of length 4 and their reversals. And these define the reduced form uniquely. In fact, these

differences define the power of $aababb$ in the reduced form. The claim is proved.

As a direct consequence of the claim above one gets that the reduced form of v is unique and does not depend on how the rewriting rules were applied.

Now we will prove that the word has to contain long palindromes. The proof is similar to the proof for the case (2,3), although in the case (4,2) there are a few technical details to verify via case study.

Assume that an infinite binary word w with its set of factors closed under 4-abelian reversal does not contain 4-palindromes of length greater than N for some integer N . Take a factor $v = w_i \cdots w_{i+N}$ of length $N + 1$. For technical reasons we assume that it does not start or end with xxx for $x \in \Sigma$. Since the set of factors of w is closed under 4-abelian reversal, there exists an occurrence of a factor v' , $v' = w_j \cdots w_{j+N} \sim_4 v^R$. Without loss of generality we can assume that $j > i$ and that the reduced form u of v is a factor of $(aababb)^\infty$, i.e., $u = z^m \text{pref}(z)$ for some $m \geq 0$ and some conjugate z of $aababb$. We will give the proof for the case $z = aababb$, the proofs for other conjugates are symmetric. Now consider the factor $w_i \cdots w_{j+N}$; its reduced form is $(aababb)^r aa$ for some $r \geq 0$.

Similarly to the case (2,3), it is not hard to see that there exists a suffix $w_s \cdots w_{j+N}$, $s < j$ which begins with aab (the same length 3 prefix as $w_i \cdots w_{j+N}$) and which is reduced to a word without factors of length 6 of $(aababb)^\infty$ or $(bbabaa)^\infty$. To see this, consider the consecutive suffices beginning with aab , then the reduced form of these suffices are of the form $(aababb)^l aa$ with the values of l differing by at most 1 (possibly with negative values corresponding to factors of $(bbabaa)^\infty$). Therefore, the suffix $w_s \cdots w_{j+N}$ corresponds to the power $l = 0$. With not too technical case study one can show that the reduced form of $w_s \cdots w_{j+N}$ is a palindrome of length at most 5, and hence $w_s \cdots w_{j+N}$ is a 4-palindrome. The length of this 4-palindrome is greater than N , a contradiction.

On the other hand, clearly, the word $(aababb)^\omega$ is 4-abelian palindromic poor, although its set of factors is not closed under 4-abelian reversal. So, there exists a binary 4-abelian palindromic poor word, but not with the set of factors closed under (4-abelian) reversal. \square

Remark 1. We emphasize that the word $\tau_{2,4}(w)$ from our construction in the case (2,4) contains only five 2-palindromes: ε , a , b , c , d , which is clearly the minimal possible number.

Remark 2. In the conference version [19] of this paper other examples of k -abelian poor words have been provided. Those examples are based on a different idea; namely, our constructions are modifications of the so-called sesquipowers, see, e.g., [25, Chapter 4]. The examples are recurrent, but not uniformly recurrent. Here we provide the constructions without proofs, and for the proofs we refer to [19].

We construct an infinite recurrent $(2, 4)$ -poor word as follows:

$$\begin{aligned} U_0 &= abca abda acda, \\ U_n &= U_{n-1}(abca)^{2^{2^n}}(abda)^{2^{2^n}}(acda)^{2^{2^n}}U_{n-1}^R. \end{aligned} \tag{1}$$

The required word is obtained as the limit $u = \lim_{n \rightarrow \infty} U_n$:

$$u = abca abda acda(abca)^4(abda)^4(acda)^4adca adba acba(abca)^{16}(abda)^{16} \dots$$

To prove that it contains only finitely many 2-abelian palindromes, we showed that each factor of length greater than 12 contains either unequal numbers of occurrences of factors bc and cb , or unequal numbers of occurrences of factors bd and db , or unequal numbers of occurrences of factors cd and dc .

An infinite recurrent $(3, 3)$ -poor word can be constructed as follows:

$$\begin{aligned} V_0 &= bbacc aabcc bbcaa, \\ V_n &= V_{n-1}(bbacc)^{2^{2^n}}(aabcc)^{2^{2^n}}(bbcaa)^{2^{2^n}}V_{n-1}^R. \end{aligned}$$

The word is given by the limit $v = \lim_{n \rightarrow \infty} V_n$:

$$v = bbacc aabcc bbcaa(bbacc)^4(aabcc)^4(bbcaa)^4aacbb cbaa ccabb(bbacc)^{16} \dots$$

The proof is based on the fact that each sufficiently long factor contains either unequal numbers of occurrences of factors bac and cab , or unequal numbers of occurrences of factors abc and cba , or unequal numbers of occurrences of factors bca and acb , and hence is not a 3-palindrome. In other words, two letter factors of the case $(2, 4)$ are now replaced by suitable three-letter factors over ternary alphabet.

An infinite recurrent $(5, 2)$ -poor word can be constructed as follows:

$$\begin{aligned} W_0 &= bbbabaaabbb bbbabbaabbb bbbabaabbbb, \\ W_n &= W_{n-1}(bbbabaaabbb)^{2^{2^n}}(bbbabbaabbb)^{2^{2^n}}(bbbabaabbbb)^{2^{2^n}}W_{n-1}^R. \end{aligned}$$

The word is given by the limit $w = \lim_{n \rightarrow \infty} W_n$. The proof is similar to the previous two examples; here the specific factors are five-letter binary words $abaaa$, $abbaa$, $abaab$ and their reversals.

4. k -abelian rich words

In this section we show that there exist words of length n which have the number of inequivalent k -abelian palindromic factors of the same order as the total number of their factors $\Theta(n^2)$. In this sense these words contain “many” k -palindromes and hence are considered as rich.

Theorem 2. *Let k be a natural number, $k \geq 2$. There exists a positive constant C such that for each $n \geq k$ there exists a word of length n containing at least Cn^2 k -abelian palindromes. Actually, we can choose $C = 1/4k$.*

Proof. The word is defined by

$$v = a^l (ba^{k-1})^m,$$

where $l \geq k - 1$ and m are chosen to give maximal number of k -palindromes among words of this type. We let $\lfloor r \rfloor$ denote the closest integer to r , we can take $m = \lfloor \frac{n-k+1}{2k} \rfloor$. Let us count the number of inequivalent k -palindromes in the word $v = v_1 \cdots v_n$, $n = km + l$. The k -palindromes are the following:

- Starting from position 1, we get the following k -palindromes
 - ε (empty word)
 - $v_1, v_1v_2, \dots, v_1 \cdots v_l$ (l k -palindromes consisting of only a 's)
 - $v_1 \cdots v_{l+k}, v_1 \cdots v_{l+2k}, \dots, v_1 \cdots v_{l+mk}$ (m k -palindromes starting with a^{k-1} , of length $l + ik$ and containing i letters b , $i = 1, \dots, m$)
- Starting from each position j , $j = 2, \dots, l - k + 2$, we get the following new k -palindromes: $v_j \cdots v_{l+k}, v_j \cdots v_{l+2k}, \dots, v_j \cdots v_{l+mk}$ (m k -palindromes starting with a^{k-1} , of length $l - j + 1 + ik$ and containing i letters b , $i = 1, \dots, m$)
- Starting from each position j , $j = l - k + 3, \dots, l + 1$, we get the following new k -palindromes: $v_j \cdots v_{2l-j+2}, v_j \cdots v_{2l-j+2+k}, \dots, v_j \cdots v_{2l-j+2+(m-1)k}$ (m k -palindromes starting with a^{l+1-j} , of length $2l - 2j + 3 + (i - 1)k$ and containing i letters b , $i = 1, \dots, m$)

It is not hard to see that all the above k -palindromes are distinct up to k -abelian equivalence; in fact, they are abelian inequivalent. So, in total we have $(l+1)(m+1) = (n - mk + 1)(m+1)$ distinct k -palindromes. Considering

this as a function of m , we get that this function takes a maximal value when $m = \frac{n-k+1}{2k}$. Since all numbers are integer there, the actual maximal number of k -palindromes given by this construction is given by taking the closest integer value, i.e., $m = \lfloor \frac{n-k+1}{2k} \rfloor$ (since the function is quadratic in m). Note that the condition $l \geq k - 1$ is always satisfied then. Taking these values and taking into account the condition $n \geq k$, we derive that the number of k -palindromes is $(l + 1)(m + 1) \geq n^2/4k$. \square

We remark that in the $\Theta(n^2)$ number of k -palindromic factors the constant actually depends on k , so it makes sense when n is large relatively to k .

5. Conclusions

We have considered the numbers of k -abelian palindromes in finite and infinite words. These numbers are always between a constant and a quadratic bound, corresponding to so-called poor and rich words. Our main result was a construction of infinite words containing only finitely many k -abelian palindromes. This construction was based on a self-avoiding fractal curve and could be modified for different pairs (k, l) , where k was a constant in k -abelian equivalence and l was the size of the alphabet. For the remaining pairs, to show that such an infinite poor word does not exist, we used a different approach, based on rewriting rules preserving k -abelian palindromicity. We also gave an example showing the existence of rich finite words, that is words containing the maximal number of k -abelian palindromes up to a constant multiplicative factor. The bound we found is $n^2/4k$, that is of order Cn^2 , where C is a constant independent of n .

6. Acknowledgements

The second author is supported in part by the Academy of Finland under grant 257857. The third author is supported by the LABEX MILYON (ANR-10-LABEX-0070) of Université de Lyon, within the program “Investissements d’Avenir” (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

- [1] B. Adamczewski: Balances for fixed points of primitive substitutions. *Theor. Comput. Sci.* 307 (2003) 47–75.

- [2] J.-P. Allouche, M. Baake, J. Cassaigne, D. Damanik: Palindrome complexity. *Theor. Comput. Sci.* 292 (2003) 9–31.
- [3] J.-P. Allouche, J. Shallit: *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge UK, Cambridge University Press, 2003.
- [4] P. Ambrož, C. Frougny, Z. Masáková, E. Pelantová: Palindromic complexity of infinite words associated with simple Parry numbers. *Ann. Inst. Fourier (Grenoble)* 56 (2006) 2131–2160.
- [5] S. Avgustinovich, J. Karhumäki, S. Puzynina: On abelian versions of Critical Factorization Theorem. *RAIRO - Theoret. Inf. Appl.* 46 (2012) 3–15.
- [6] S. Avgustinovich, S. Puzynina: Weak Abelian Periodicity of Infinite Words. *CSR 2013, LNCS 7913* (2013) 258–270.
- [7] H. Bannai, T. Shunsuke, I. Nakashima, M. Takeda, K. Tsuruta: The “Runs” Theorem. *Arxiv: 1406.0263* (2014).
- [8] J. Berstel, L. Boasson, O. Carton, I. Fagnot: Infinite words without palindrome. *Arxiv: 0903.2382* (2009).
- [9] S. Brlek, S. Hamel, M. Nivat, C. Reutenauer: On the palindromic complexity of infinite words. *Internat. J. Found. Comput. Sci.* 15 (2004) 293–306.
- [10] M. Bucci, A. De Luca, A. Glen, L.Q. Zamboni: A new characteristic property of rich words. *Theoret. Comput. Sci.* 410 (2009) 2860–2863.
- [11] M. Crochemore, L. Ilie, L. Tinta: The “runs” conjecture. *Theor. Comput. Sci.* 412 (2011) 2931–2941.
- [12] D. Damanik, D. Zare: Palindrome complexity bounds for primitive substitution sequences. *Discrete Math.* 222 (2000) 259–267.
- [13] A. de Luca: Sturmian words: structures, combinatorics and their arithmetics. *Theoret. Comput. Sci.* 183 (1997) 45–82.
- [14] X. Droubay, J. Justin, G. Pirillo: Episturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci.* 255 (2001) 539–553.

- [15] G. Fici, L. Q. Zamboni: On the least number of palindromes contained in an infinite word. *Theor. Comput. Sci.* 481 (2013) 1–8.
- [16] A. Glen, J. Justin, S. Widmer, L.Q. Zamboni: Palindromic richness. *European J. Combin.* 30 (2009) 510–531.
- [17] T. Harju, T. Kärki, D. Nowotka: The number of positions starting a square in binary words. *Electron. J. Combin.* 18 (2011) no. 1.
- [18] L. Ilie: A note on the number of squares in a word. *Theor. Comput. Sci.* 380 (2007) 373–376.
- [19] J. Karhumäki, S. Puzynina: On k -abelian palindromic rich and poor words. *DLT 2014, LNCS 8633* (2014) 191–202.
- [20] J. Karhumäki, S. Puzynina, A. Saarela: Fine and Wilf’s Theorem for k -Abelian Periods. *Int. J. Found. Comput. Sci.* 24 (2013) 1135–1152.
- [21] J. Karhumäki, A. Saarela, L. Q. Zamboni: On a generalization of Abelian equivalence and complexity of infinite words. *J. Comb. Theory, Ser. A* 120 (2013) 2189–2206.
- [22] V. Keränen: Abelian squares are avoidable on 4 letters. In: Kuich, W. (ed.) *ICALP 1992. LNCS 623*, Springer, Heidelberg (1992) 41–52.
- [23] R. Kolpakov, G. Kucherov: Finding Maximal Repetitions in a Word in Linear Time. *FOCS 1999*, 596–604.
- [24] G. Kucherov, P. Ochem, M. Rao: How Many Square Occurrences Must a Binary Sequence Contain? *Electr. J. Comb.* 10 (2003).
- [25] M. Lothaire: *Algebraic Combinatorics on Words*. Cambridge UK, Cambridge University Press, 2002.
- [26] S. Puzynina, L. Q. Zamboni: Abelian returns in Sturmian words. *J. Comb. Theory, Ser. A* 120 (2013) 390–408.
- [27] N. Rampersad, J. Shallit: Words avoiding reversed subwords. *J. Combin. Math. Combin. Comput.* 54 (2005) 157–164.
- [28] G. Richomme, K. Saari, L.Q. Zamboni: Abelian Complexity of Minimal Subshifts. *J. London Math. Soc.* 83 (2011) 79–95.

- [29] M. Rigo, P. Salimov: Another Generalization of Abelian Equivalence: Binomial Complexity of Infinite Words. WORDS 2013, LNCS 8079 (2013) 217–228.