Specular sets

Valérie Berthé[1], Clelia De Felice[2], Vincent Delecroix[3], Francesco Dolce[4], Julien Leroy[5], Dominique Perrin[4], Christophe Reutenauer[6], Giuseppina Rindone[4]

[1]CNRS, Université Paris 7, [2]Università degli Studi di Salerno, [3]CNRS, Université de Bordeaux, [4]Université Paris Est, [5]Université de Liège, [6]Université du Québec à Montréal

Abstract. We introduce specular sets. These are subsets of groups which form a natural generalization of free groups. These sets are an abstract generalization of the natural codings of interval exchanges and of linear involutions. We prove several results concerning the subgroups generated by return words and by maximal bifix codes in these sets.

1 Introduction

We have studied in a series of papers initiated in [3] the links between minimal sets, subgroups of free groups and bifix codes. In this paper, we continue this investigation in a situation which involves groups which are not free anymore. These groups, named here specular, are free products of a free group and of a finite number of cyclic groups of order two. These groups are close to free groups and, in particular, the notion of a basis in such groups is clearly defined. It follows from Kurosh's theorem that any subgroup of a specular group is specular. A specular set is a subset of such a group which generalizes the natural codings of linear involutions studied in [9].

The main results of this paper are Theorem 5.1, referred to as the Return Theorem and Theorem 5.2, referred to as the Finite Index Basis Theorem. The first one asserts that the set of return words to a given word in a uniformly recurrent specular set is a basis of a subgroup of index 2 called the even subgroup. The second one characterizes the monoidal bases of subgroups of finite index of specular groups contained in a specular set S as the finite S-maximal symmetric bifix codes contained in S. This generalizes the analogous results proved initially in [3] for Sturmian sets and extended in [7] to the class of tree sets (this class contains both Sturmian sets and interval exchange sets).

There are two interesting features of the subject of this paper.

In the first place, some of the statements concerning the natural codings of interval exchanges and of linear involutions can be proved using geometric methods, as shown in a separate paper [9]. This provides an interesting interpretation of the groups playing a role in these natural codings (these groups are generated either by return words or by maximal bifix codes) as fundamental groups of some surfaces. The methods used here are purely combinatorial.

In the second place, the abstract notion of a specular set gives rise to groups called here specular. These groups are natural generalizations of free groups, and are free products of $\mathbb{Z}$ and $\mathbb{Z}/2\mathbb{Z}$. They are called *free-like* in [2] and appear at several places in [12].

The idea of considering recurrent sets of reduced words invariant by taking inverses is connected, as we shall see, with the notion of G-rich words of [18].

The paper is organized as follows. In Section 2, we recall some notions concerning words, extension graphs and bifix codes. In Section 3, we introduce specular groups, which form a family with properties very close to free groups. We prove properties of these groups extending those of free groups, like a Schreier's Formula (Formula (3.1)). In Section 4, we introduce specular sets. This family contains the natural codings of linear involutions without connection studied in [5]. We prove a result connecting specular sets with the family of tree sets introduced in [6] (Theorem 4.6). In Section 5, we prove several results concerning subgroups generated by subsets of specular groups. We first prove that the set of return words to a given word forms a basis of the even subgroup (Theorem 5.1 referred to as the Return Theorem). This is a subgroup defined in terms of particular letters, called even letters, that play a special role with respect to the extension graph of the empty word. We next prove the Finite Index Basis Theorem (Theorem 5.2).

Acknowledgments The authors thank Laurent Bartholdi and Pierre de la Harpe for useful indications. This work was supported by grants from Région Île-de-France and ANR project Eqinocs.

2 Preliminaries

A set of words on the alphabet A and containing A is said to be *factorial* if it contains the factors of its elements. An *internal factor* of a word x is a word v such that $x = uvw$ with u, w nonempty.

Let S be a set of words on the alphabet A. For $w \in S$, we denote $L_S(w) = \{a \in A \mid aw \in S\}$, $R_S(w) = \{a \in A \mid wa \in S\}$ and $E_S(w) = \{(a, b) \in A \times A \mid awb \in S\}$. Further $\ell_S(w) = \mathrm{Card}(L_S(w))$, $r_S(w) = \mathrm{Card}(R_S(w))$, $e_S(w) = \mathrm{Card}(E_S(w))$. We omit the subscript S when it is clear from the context. A word w is *right-extendable* if $r(w) > 0$, *left-extendable* if $\ell(w) > 0$ and *biextendable* if $e(w) > 0$. A factorial set S is called *right-extendable* (resp. *left-extendable*, resp. *biextendable*) if every word in S is right-extendable (resp. left-extendable, resp. biextendable).

A word w is called *right-special* if $r(w) \geq 2$. It is called *left-special* if $\ell(w) \geq 2$. It is called *bispecial* if it is both left-special and right-special.

For $w \in S$, we denote

$$m_S(w) = e_S(w) - \ell_S(w) - r_S(w) + 1.$$

The word w is called *weak* if $m_S(w) < 0$, *neutral* if $m_S(w) = 0$ and *strong* if $m_S(w) > 0$.

We say that a factorial set S is *neutral* if every nonempty word in S is neutral. The *characteristic* of S is the integer $1 - m_S(\varepsilon)$. Thus a neutral set of characteristic 1 is such that all words (including the empty word) are neutral. This is what is called a neutral set in [6].

A set of words $S \neq \{\varepsilon\}$ is *recurrent* if it is factorial and if for any $u, w \in S$, there is a $v \in S$ such that $uvw \in S$. An infinite factorial set is said to be *uniformly recurrent* if for any word $u \in S$ there is an integer $n \geq 1$ such that u is a factor of any word of S of length n. A uniformly recurrent set is recurrent.

The *factor complexity* of a factorial set S of words on an alphabet A is the sequence $p_n = \text{Card}(S \cap A^n)$. Let $s_n = p_{n+1} - p_n$ and $b_n = s_{n+1} - s_n$ be respectively the first and second order differences sequences of the sequence p_n.

The following result is from [11] (see also [10], Theorem 4.5.4).

Proposition 2.1 *Let S be a factorial set on the alphabet A. One has $b_n = \sum_{w \in S \cap A^n} m(w)$ and $s_n = \sum_{w \in S \cap A^n} (r(w) - 1)$ for all $n \geq 0$.*

Let S be a biextendable set of words. For $w \in S$, we consider the set $E(w)$ as an undirected graph on the set of vertices which is the disjoint union of $L(w)$ and $R(w)$ with edges the pairs $(a, b) \in E(w)$. This graph is called the *extension graph* of w. We sometimes denote $1 \otimes L(w)$ and $R(w) \otimes 1$ the copies of $L(w)$ and $R(w)$ used to define the set of vertices of $E(w)$.

If the extension graph $E(w)$ is acyclic, then $m(w) = 1 - c$, where c is the number of connected components of the graph $E(w)$. Thus w is weak or neutral.

A biextendable set S is called *acyclic* if for every $w \in S$, the graph $E(w)$ is acyclic. A biextendable set S is called a *tree set* of characteristic c if for any nonempty $w \in S$, the graph $E(w)$ is a tree and if $E(\varepsilon)$ is a union of c trees (the definition of tree set in [6] corresponds to a tree set of characteristic 1). Note that a tree set of characteristic c is a neutral set of characteristic c.

As an example, a Sturmian set is a tree set of characteristic 1 (by a Sturmian set, we mean the set of factors of a strict episturmian word, see [3]).

Let S be a factorial set of words and $x \in S$. A *return word* to x in S is a nonempty word u such that the word xu is in S and ends with x, but has no internal factor equal to x. We denote by $\mathcal{R}_S(x)$ the set of return words to x in S. The set of *complete return words* to $x \in S$ is the set $x\mathcal{R}_S(x)$.

Bifix codes. A *prefix code* is a set of nonempty words which does not contain any proper prefix of its elements. A *suffix code* is defined symmetrically. A *bifix code* is a set which is both a prefix code and a suffix code (see [4] for a more detailed introduction).

Let S be a recurrent set. A prefix (resp. bifix) code $X \subset S$ is S-maximal if it is not properly contained in a prefix (resp. bifix) code $Y \subset S$. Since S is recurrent, a finite S-maximal bifix code is also an S-maximal prefix code (see [3], Theorem 4.2.2). For example, for any $n \geq 1$, the set $X = S \cap A^n$ is an S-maximal bifix code.

Let X be a bifix code. Let Q be the set of words without any suffix in X and let P be the set of words without any prefix in X. A *parse* of a word w with respect to a bifix code X is a triple $(q, x, p) \in Q \times X^* \times P$ such that $w = qxp$. We denote by $d_X(w)$ the number of parses of a word w with respect to X. The

S-degree of X, denoted $d_X(S)$, is the maximal number of parses with respect to X of a word of S. For example, the set $X = S \cap A^n$ has S-degree n.

Let S be a recurrent set and let X be a finite bifix code. By Theorem 4.2.8 in [3], X is S-maximal if and only if its S-degree is finite. Moreover, in this case, a word $w \in S$ is such that $d_X(w) < d_X(S)$ if and only if it is an internal factor of a word of X.

3 Specular groups

We consider an alphabet A with an involution $\theta : A \to A$, possibly with some fixed points. We also consider the group G_θ generated by A with the relations $a\theta(a) = 1$ for every $a \in A$. Thus $\theta(a) = a^{-1}$ for $a \in A$. The set A is called a *natural* set of generators of G_θ.

When θ has no fixed point, we can set $A = B \cup B^{-1}$ by choosing a set of representatives of the orbits of θ for the set B. The group G_θ is then the free group on B. In general, the group G_θ is a free product of a free group and a finite number of copies of $\mathbb{Z}/2\mathbb{Z}$, that is, $G_\theta = \mathbb{Z}^{*i} * (\mathbb{Z}/2\mathbb{Z})^{*j}$ where i is the number of orbits of θ with two elements and j the number of its fixed points. Such a group will be called a *specular group* of type (i, j). These groups are very close to free groups, as we will see. The integer $\mathrm{Card}(A) = 2i + j$ is called the *symmetric rank* of the specular group $\mathbb{Z}^{*i} * (\mathbb{Z}/2\mathbb{Z})^{*j}$. Two specular groups are isomorphic if and only if they have the same type. Indeed, the commutative image of a group of type (i, j) is $\mathbb{Z}^i \times (\mathbb{Z}/2\mathbb{Z})^j$ and the uniqueness of i, j follows from the fundamental theorem of finitely generated Abelian groups.

Example 3.1. Let $A = \{a, b, c, d\}$ and let θ be the involution which exchanges b, d and fixes a, c. Then $G_\theta = \mathbb{Z} * (\mathbb{Z}/2\mathbb{Z})^2$ is a specular group of symmetric rank 4.

By Kurosh's Theorem, any subgroup of a free product $G_1 * G_2 * \cdots * G_n$ is itself a free product of a free group and of groups conjugate to subgroups of the G_i (see [17]). Thus, we have, replacing the Nielsen-Schreier Theorem of free groups, the following result.

Theorem 3.1. *Any subgroup of a specular group is specular.*

It also follows from Kurosh's theorem that the elements of order 2 in a specular group G_θ are the conjugates of the j fixed points of θ and this number is thus the number of conjugacy classes of elements of order 2.

A word on the alphabet A is *θ-reduced* (or simply reduced) if it has no factor of the form $a\theta(a)$ for $a \in A$. It is clear that any element of a specular group is represented by a unique reduced word.

A subset of a group G is called *symmetric* if it is closed under taking inverses. A set X in a specular group G is called a *monoidal basis* of G if it is symmetric, if the monoid that it generates is G and if any product $x_1x_2\cdots x_m$ of elements of X such that $x_kx_{k+1} \neq 1$ for $1 \leq k \leq m-1$ is distinct of 1. The alphabet A is a monoidal basis of G_θ and the symmetric rank of a specular group is the cardinality of any monoidal basis (two monoidal bases have the same cardinality since the type is invariant by isomorphism).

If H is a subgroup of index n of a specular group G of symmetric rank r, the symmetric rank s of H is

$$s = n(r-2) + 2. \tag{3.1}$$

This formula replaces Schreier's Formula (which corresponds to the case $j = 0$). It can be proved as follows. Let Q be a Schreier transversal for H, that is, a set of reduced words which is a prefix-closed set of representatives of the right cosets Hg of H. Let X be the corresponding Schreier basis, formed of the paq^{-1} for $a \in A$, $p, q \in Q$ with $pa \notin Q$ and $pa \in Hq$. The number of elements of X is $nr - 2(n-1)$. Indeed, this is the number of pairs $(p, a) \in Q \times A$ minus the $2(n-1)$ pairs (p, a) such that $pa \in Q$ with pa reduced or $pa \in Q$ with pa not reduced. This gives Formula (3.1).

Any specular group $G = G_\theta$ has a free subgroup of index 2. Indeed, let H be the subgroup formed of the reduced words of even length. It has clearly index 2. It is free because it does not contain any element of order 2 (such an element is conjugate to a fixed point of θ and thus is of odd length).

A group G is called *residually finite* if for every element $g \neq 1$ of G, there is a morphism φ from G onto a finite group such that $\varphi(g) \neq 1$.

It follows easily by considering a free subgroup of index 2 of a specular group that any specular group is residually finite. A group G is said to be *Hopfian* if any surjective morphism from G onto G is also injective. By a result of Malcev, any finitely generated residually finite group is Hopfian (see [16], p. 197). Thus any specular group is Hopfian.

As a consequence, one has the following result, which can be obtained by considering the commutative image of a specular group.

Proposition 3.2 *Let G be a specular group of type (i, j) and let $X \subset G$ be a symmetric set with $2i + j$ elements. If X generates G, it is a monoidal basis of G.*

4 Specular sets

We assume given an involution θ on the alphabet A generating the specular group G_θ. A *specular set* on A is a biextendable symmetric set of θ-reduced words on A which is a tree set of characteristic 2. Thus, in a specular set, the extension graph of every nonempty word is a tree and the extension graph of the empty word is a union of two disjoint trees.

The following is a very simple example of a specular set.

Example 4.1. Let $A = \{a, b\}$ and let θ be the identity on A. Then the set of factors of $(ab)^\omega$ is a specular set (we denote by x^ω the word x infinitely repeated).

The following result shows in particular that in a specular set the two trees forming $E(\varepsilon)$ are isomorphic since they are exchanged by the bijection $(a, b) \mapsto (b^{-1}, a^{-1})$.

Proposition 4.1 *Let S be a specular set. Let $\mathcal{T}_0, \mathcal{T}_1$ be the two trees such that $E(\varepsilon) = \mathcal{T}_0 \cup \mathcal{T}_1$. For any $a, b \in A$ and $i = 0, 1$, one has $(1 \otimes a, b \otimes 1) \in \mathcal{T}_i$ if and only if $(1 \otimes b^{-1}, a^{-1} \otimes 1) \in \mathcal{T}_{1-i}$*

Proof. Assume that $(1 \otimes a, b \otimes 1)$ and $(1 \otimes b^{-1}, a^{-1} \otimes 1)$ are both in $\mathcal{T}_0$. Since $\mathcal{T}_0$ is a tree, there is a path from $1 \otimes a$ to $a^{-1} \otimes 1$. We may assume that this path is reduced, that is, does not use consecutively twice the same edge. Since this path is of odd length, it has the form $(u_0, v_0, u_1, \ldots, u_p, v_p)$ with $u_0 = 1 \otimes a$ and $v_p = a^{-1} \otimes 1$. Since S is symmetric, we also have a reduced path $(v_p^{-1}, u_p^{-1}, \cdots, u_1^{-1}, v_0^{-1}, u_0^{-1})$ which is in $\mathcal{T}_0$ (for $u_i = 1 \otimes a_i$, we denote $u_i^{-1} = a_i^{-1} \otimes 1$ and similarly for v_i^{-1}). Since $v_p^{-1} = u_0$, these two paths have the same origin and end. But if a path of odd length is its own inverse, its central edge has the form (x, y) with $x = y^{-1}$ a contradiction with the fact that the words of S are reduced. Thus the two paths are distinct. This implies that $E(\varepsilon)$ has a cycle, a contradiction.

The next result follows easily from Proposition 2.1.

Proposition 4.2 *The factor complexity of a specular set on the alphabet A is $p_n = n(k-2) + 2$ for $n \geq 1$ with $k = \mathrm{Card}(A)$.*

Doubling maps. We now introduce a construction which allows one to build specular sets.

A *transducer* is a graph on a set Q of vertices with edges labeled in $\Sigma \times A$. The set Q is called the set of states, the set Σ is called the *input* alphabet and A is called the *output* alphabet. The graph obtained by erasing the ouput letters is called the *input automaton* (with an unspecified initial state). Similarly, the *ouput automaton* is obtained by erasing the input letters.

Let $\mathcal{A}$ be a transducer with set of states $Q = \{0, 1\}$ on the input alphabet Σ and the output alphabet A. We assume that

1. the input automaton is a group automaton, that is, every letter of Σ acts on Q as a permutation,
2. the output labels of the edges are all distinct.

We define two maps $\delta_0, \delta_1 : \Sigma^* \to A^*$ corresponding to initial states 0 and 1 respectively. Thus $\delta_0(u) = v$ (resp. $\delta_1(u) = v$) if the path starting at state 0 (resp. 1) with input label u has output label v. The pair $\delta = (\delta_0, \delta_1)$ is called a *doubling map* on $\Sigma \times A$ and the transducer $\mathcal{A}$ a *doubling transducer*. The *image* of a set T on the alphabet Σ by the doubling map δ is the set $S = \delta_0(T) \cup \delta_1(T)$.

If $\mathcal{A}$ is a doubling transducer, we define an involution $\theta_{\mathcal{A}}$ as follows. For any $a \in A$, let (i, α, a, j) be the edge with input label α and output label a. We define $\theta_{\mathcal{A}}(a)$ as the output label of the edge starting at $1 - j$ with input label α. Thus, $\theta_{\mathcal{A}}(a) = \delta_i(\alpha) = a$ if $i + j = 1$ and $\theta_{\mathcal{A}}(a) = \delta_{1-i}(\alpha) \neq a$ if $i = j$.

The *reversal* of a word $w = a_1 a_2 \cdots a_n$ is the word $\tilde{w} = a_n \cdots a_2 a_1$. A set S of words is closed under reversal if $w \in S$ implies $\tilde{w} \in S$ for every $w \in S$. As is well known, any Sturmian set is closed under reversal (see [3]). The proof of the following result can found in [5].

Proposition 4.3 *For any tree set T of characteristic 1 on the alphabet Σ, closed under reversal and for any doubling map δ, the image of T by δ is a specular set relative to the involution $\theta_{\mathcal{A}}$.*

We now give an example of a specular set obtained by a doubling map.

Example 4.2. Let $\Sigma = \{\alpha, \beta\}$ and let T be the Fibonacci set, which is the Sturmian set formed of the factors of the fixed point of the morphism $\alpha \mapsto \alpha\beta, \beta \mapsto \alpha$. Let δ be the doubling map given by the transducer $\mathcal{A}$ of Figure 4.1 on the left.

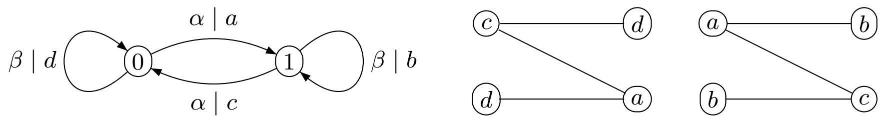

Fig. 4.1. A doubling transducer and the extension graph $E_S(\varepsilon)$.

Then $\theta_{\mathcal{A}}$ is the involution θ of Example 3.1 and the image of T by δ is a specular set S on the alphabet $A = \{a, b, c, d\}$. The graph $E_S(\varepsilon)$ is represented in Figure 4.1 on the right.

Note that S is the set of factors of the fixed point $g^\omega(a)$ of the morphism $g : a \mapsto abcab,\ b \mapsto cda,\ c \mapsto cdacd,\ d \mapsto abc$. The morphism g is obtained by applying the doubling map to the cube f^3 of the Fibonacci morphism f in such a way that $g^\omega(a) = \delta_0(f^\omega(\alpha))$.

Odd and even words. We introduce a notion which plays, as we shall see, an important role in the study of specular sets. Let S be a specular set. Since a specular set is biextendable, any letter $a \in A$ occurs exactly twice as a vertex of $E(\varepsilon)$, one as an element of $L(\varepsilon)$ and one as an element of $R(\varepsilon)$. A letter $a \in A$ is said to be *even* if its two occurrences appear in the same tree. Otherwise, it is said to be *odd*. Observe that if S is recurrent, there is at least one odd letter.

Example 4.3. Let S be the specular set of Example 4.2. The letters a, c are odd and b, d are even.

A word $w \in S$ is said to be *even* if it has an even number of odd letters. Otherwise it is said to be *odd*. The set of even words has the form $X^* \cap S$ where $X \subset S$ is a bifix code, called the *even code*. The set X is the set of even words without a nonempty even prefix (or suffix).

Proposition 4.4 *Let S be a recurrent specular set. The even code is an S-maximal bifix code of S-degree 2.*

Proof. Let us verify that any $w \in S$ is comparable for the prefix order with an element of the even code X. If w is even, it is in X^*. Otherwise, since S is recurrent, there is a word u such that $wuw \in S$. If u is even, then wuw is even and thus $wuw \in X^*$. Otherwise wu is even and thus $wu \in X^*$. This shows that

X is S-maximal. The fact that it has S-degree 2 follows from the fact that any product of two odd letters is a word of X which is not an internal factor of X and has two parses.

Example 4.4. Let S be the specular set of Example 4.2. The even code is $X = \{abc, ac, b, ca, cda, d\}$.

Denote by $\mathcal{T}_0, \mathcal{T}_1$ the two trees such that $E(\varepsilon) = \mathcal{T}_0 \cup \mathcal{T}_1$. We consider the directed graph $\mathcal{G}$ with vertices $0, 1$ and edges all the triples (i, a, j) for $0 \leq i, j \leq 1$ and $a \in A$ such that $(1 \otimes b, a \otimes 1) \in \mathcal{T}_i$ and $(1 \otimes a, c \otimes 1) \in \mathcal{T}_j$ for some $b, c \in A$. The graph $\mathcal{G}$ is called the *parity graph* of S. Observe that for every letter $a \in A$ there is exactly one edge labeled a because a appears exactly once as a left (resp. right) vertex in $E(\varepsilon)$.

Note that, when S is a specular set obtained by a doubling map using a transducer $\mathcal{A}$, the parity graph of S is the output automaton of $\mathcal{A}$.

Example 4.5. The parity graph of the specular set of Example 4.2 is the output automaton of the doubling transducer of Figure 4.1.

The proof of the following result can be found in [5].

Proposition 4.5 *Let S be a specular set and let $\mathcal{G}$ be its parity graph. Let $S_{i,j}$ be the set of words in S which are the label of a path from i to j in the graph $\mathcal{G}$.*

(1) The family $(S_{i,j} \setminus \{\varepsilon\})_{0 \leq i,j \leq 1}$ is a partition of $S \setminus \{\varepsilon\}$.
(2) For $u \in S_{i,j} \setminus \{\varepsilon\}$ and $v \in S_{k,\ell} \setminus \{\varepsilon\}$, if $uv \in S$, then $j = k$.
(3) $S_{0,0} \cup S_{1,1}$ is the set of even words.
(4) $S_{i,j}^{-1} = S_{1-j,1-i}$.

A *coding morphism* for a prefix code X on the alphabet A is a morphism $f : B^* \to A^*$ which maps bijectively B onto X. Let S be a recurrent set and let f be a coding morphism for an S-maximal bifix code. The set $f^{-1}(S)$ is called a *maximal bifix decoding* of S.

The following result is the counterpart for uniformly recurrent specular sets of the main result of [8, Theorem 6.1] asserting that the family of uniformly recurrent tree sets of characteristic 1 is closed under maximal bifix decoding. The proof can be found in [5].

Theorem 4.6. *The decoding of a uniformly recurrent specular set by the even code is a union of two uniformly recurrent tree sets of characteristic* 1*.*

Palindromes. The notion of palindromic complexity originates in [14] where it is proved that a word of length n has at most $n + 1$ palindrome factors. A word of length n is full if it has $n + 1$ palindrome factors and a factorial set is *full* (or rich) if all its elements are full. By a result of [15], a recurrent set S closed under reversal is full if and only if every complete return word to a palindrome in S is a palindrome. It is known that all Sturmian sets are full [14] and also all natural codings of interval exchange defined by a symmetric permutation [1]. In [18], this notion was extended to that of H-fullness, where H is a finite group of morphisms and antimorphisms of A^* (an antimorphism is the composition of a morphism and reversal) containing at least one antimorphism. As one of the equivalent definitions of H-full, a set S closed under H is H-full if for every $x \in S$, every complete return word to the H-orbit of x is fixed by a nontrivial element of H (a complete return word to a set X is a word of S which has exactly two factors in X, one as a proper prefix and one as a proper suffix).

The following result connects these notions with ours. If δ is a doubling map, we denote by H the group generated by the antimorphism $u \mapsto u^{-1}$ for $u \in G_\theta$ and the morphism obtained by replacing each letter $a \in A$ by $\tau(a)$ if there are edges (i, b, a, j) and $(1-i, b, \tau(a), 1-j)$ in the doubling transducer. Actually, we have $H = \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. The proof of the following result can be found in [5]. The fact that T is full generalizes the results of [14, 1].

Proposition 4.7 *Let T be a recurrent tree set of characteristic* 1 *on the alphabet Σ, closed under reversal and let S be the image of T under a doubling map. Then T is full and S is H-full.*

Example 4.6. Let S be the specular set of Example 4.2. Since it is a doubling of the Fibonacci set (which is Sturmian and thus full), it is H-full with respect to the group H generated by the map σ taking the inverse and the morphism τ which exchanges a, c and b, d respectively. The H-orbit of $x = a$ is the set $X = \{a, c\}$ and $\mathcal{CR}_S(X) = \{ac, abc, ca, cda\}$.

All four words are fixed by $\sigma\tau$. As another example, consider $x = ab$. Then $X = \{ab, bc, cd, da\}$ and $\mathcal{CR}_S(X) = \{abc, bcab, bcd, cda, dab, dacd\}$. Each of them is fixed by some nontrivial element of H.

5 Subgroup Theorems

In this section, we prove several results concerning the subgroups generated by subsets of a specular set.

The Return Theorem. By [6, Theorem 4.5], the set of return words to a given word in a uniformly recurrent tree set of characteristic 1 containing the alphabet A is a basis of the free group on A. We will see a counterpart of this result for uniformly recurrent specular sets.

Let S be a specular set. The *even subgroup* is the group generated by the even code. It is a subgroup of index 2 of G_θ with symmetric rank $2(\mathrm{Card}(A)-1)$ by (3.1). Since no even word is its own inverse (see Proposition 4.5), it is a free group. Thus its rank is $\mathrm{Card}(A) - 1$. The proof can be found in [5].

Theorem 5.1. *Let S be a uniformly recurrent specular set on the alphabet A. For any $w \in S$, the set of return words to w is a basis of the even subgroup.*

Note that this implies that $\mathrm{Card}(\mathcal{R}_S(x)) = \mathrm{Card}(A) - 1$.

Example 5.1. Let S be the specular set of Example 4.2. The set of return words to a is $\mathcal{R}_S(a) = \{bca, bcda, cda\}$. It is a basis of the even subgroup.

Finite Index Basis Theorem. The following result is the counterpart for specular sets of the result holding for uniformly recurrent tree sets of characteristic 1 (see [7, Theorem 4.4]). The proof can be found in [5].

Theorem 5.2. *Let S be a uniformly recurrent specular set and let $X \subset S$ be a finite symmetric bifix code. Then X is an S-maximal bifix code of S-degree d if and only if it is a monoidal basis of a subgroup of index d.*

Note that when X is not symmetric, the index of the subgroup generated by X may be different of $d_X(S)$.

Note also that Theorem 5.2 implies that for any uniformly recurrent specular set and for any finite symmetric S-maximal bifix code X, one has $\mathrm{Card}(X) = d_X(S)(\mathrm{Card}(A)-2)+2$. This follows actually also (under more general hypotheses) from Theorem 2 in [13].

The proof of the Finite Index Basis Theorem needs preliminary results which involve concepts like that of incidence graph which are interesting in themselves.

Saturation Theorem. The *incidence graph* of a set X, is the undirected graph $\mathcal{G}_X$ defined as follows. Let P be the set of proper prefixes of X and let Q be the set of its proper suffixes. Set $P' = P \setminus \{1\}$ and $Q' = Q \setminus \{1\}$. The set of vertices of $\mathcal{G}_X$ is the disjoint union of P' and Q'. The edges of $\mathcal{G}_X$ are the pairs (p, q) for $p \in P'$ and $q \in Q'$ such that $pq \in X$. As for the extension graph, we sometimes denote $1 \otimes P', Q' \otimes 1$ the copies of P', Q' used to define the set of vertices of $\mathcal{G}_X$.

Example 5.2. Let S be a factorial set and let $X = S \cap A^2$ be the bifix code formed of the words of S of length 2. The incidence graph of X is identical with the extension graph $E(\varepsilon)$.

Let X be a symmetric set. We use the incidence graph to define an equivalence relation γ_X on the set P of proper prefixes of X, called the *coset equivalence* of X, as follows. It is the relation defined by $p \equiv q \bmod \gamma_X$ if there is a path (of even length) from $1 \otimes p$ to $1 \otimes q$ or a path (of odd length) from $1 \otimes p$ to $q^{-1} \otimes 1$ in the incidence graph $\mathcal{G}_X$. It is easy to verify that, since X is symmetric, γ_X is indeed an equivalence. The class of the empty word ε is reduced to ε.

The following statement is the generalization to symmetric bifix codes of Proposition 6.3.5 in [3]. We denote by $\langle X \rangle$ the subgroup generated by X.

Proposition 5.3 *Let X be a symmetric bifix code and let P be the set of its proper prefixes. Let γ_X be the coset equivalence of X and let $H = \langle X \rangle$. For any $p, q \in P$, if $p \equiv q \bmod \gamma_X$, then $Hp = Hq$.*

We now use the coset equivalence γ_X to define the *coset automaton* $\mathcal{C}_X$ of a symmetric bifix code X as follows. The vertices of $\mathcal{C}_X$ are the equivalence classes of γ_X. We denote by $\hat{p}$ the class of p. There is an edge labeled $a \in A$ from s to t if for some $p \in s$ and $q \in t$ (that is $s = \hat{p}$ and $t = \hat{q}$), one of the following cases occurs (see Figure 5.1):

(i) $pa \in P$ and $pa \equiv q \bmod \gamma_X$
(ii) or $pa \in X$ and $q = \varepsilon$.

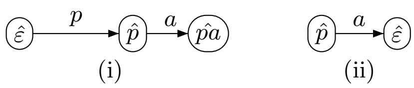

Fig. 5.1. The edges of the coset automaton.

The proof of the following statement can be found in [5].

Proposition 5.4 *Let X be a symmetric bifix code, let P be its set of proper prefixes and let $H = \langle X \rangle$. If for $p, q \in P$ and a word $w \in A^*$ there is a path labeled w from the class $\hat{p}$ to the class $\hat{q}$, then $Hpw = Hq$.*

Let A be an alphabet with an involution θ. A directed graph with edges labeled in A is called *symmetric* if there is an edge from p to q labeled a if and only if there is an edge from q to p labeled a^{-1}. If $\mathcal{G}$ is a symmetric graph and v is a vertex of $\mathcal{G}$, the set of reductions of the labels of paths from v to v is a subgroup of G_θ called the subgroup *described* by $\mathcal{G}$ with respect to v.

A symmetric graph is called *reversible* if for every pair of edges of the form $(v, a, w), (v, a, w')$, one has $w = w'$ (and the symmetric implication since the graph is symmetric).

Proposition 5.5 *Let S be a specular set and let $X \subset S$ be a finite symmetric bifix code. The coset automaton $\mathcal{C}_X$ is reversible. Moreover the subgroup described by $\mathcal{C}_X$ with respect to the class of the empty word is the group generated by X.*

Prime words with respect to a subgroup. Let H be a subgroup of the specular group G_θ and let S be a specular set on A relative to θ. The set of *prime* words in S with respect to H is the set of nonempty words in $H \cap S$ without a proper nonempty prefix in $H \cap S$. Note that the set of prime words with respect to H is a symmetric bifix code. One may verify that it is actually the unique bifix code X such that $X \subset S \cap H \subset X^*$.

The following statement is a generalization of Theorem 5.2 in [6] (Saturation Theorem). The proof can be found in [5].

Theorem 5.6. *Let S be a specular set. Any finite symmetric bifix code $X \subset S$ is the set of prime words in S with respect to the subgroup $\langle X \rangle$. Moreover $\langle X \rangle \cap S = X^* \cap S$.*

A converse of the Finite Index Basis Theorem. The following is a converse of Theorem 5.2. For the proof, see [5].

Theorem 5.7. *Let S be a recurrent and symmetric set of reduced words of factor complexity $p_n = n(\mathrm{Card}(A)-2)+2$. If $S \cap A^n$ is a monoidal basis of the subgroup $\langle A^n \rangle$ for all $n \geq 1$, then S is a specular set.*

References

1. Peter Baláži, Zuzana Masáková, and Edita Pelantová. Factor versus palindromic complexity of uniformly recurrent infinite words. *Theoret. Comput. Sci.*, 380(3):266–275, 2007.
2. Laurent Bartholdi. Growth of groups and wreath products. 2014. preprint.
3. Jean Berstel, Clelia De Felice, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and Sturmian words. *J. Algebra*, 369:146–202, 2012.
4. Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and Automata.* Cambridge University Press, 2009.
5. Valérie Berthé, , Clelia De Felice, Vincent Delecroix, Francesco Dolce, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Specular sets. 2015. https://arXiv.org/abs/1505.00707.
6. Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Acyclic, connected and tree sets. *Monatsh. Math.*, 176:521–550, 2015.
7. Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. The finite index basis property. *J. Pure Appl. Algebra*, 219:2521–2537, 2015.
8. Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Maximal bifix decoding. *Discrete Math.*, 338:725–742, 2015.
9. Valérie Berthé, Vincent Delecroix, Francesco Dolce, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Return words of linear involutions and fundamental groups. *Ergodic Th. Dyn. Syst.*, 2015. To appear, http://arxiv.org/abs/1405.3529.
10. Valérie Berthé and Michel Rigo. *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.* Cambridge Univ. Press, Cambridge, 2010.
11. Julien Cassaigne. Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc. Simon Stevin*, 4(1):67–88, 1997. Journées Montoises (Mons, 1994).
12. Pierre de la Harpe. *Topics in geometric group theory.* Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 2000.
13. Francesco Dolce and Dominique Perrin. Enumeration formulæ in neutral sets. 2015. Submitted, `http://arxiv.org/abs/1503.06081`.
14. Xavier Droubay, Jacques Justin, and Giuseppe Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci.*, 255(1-2):539–553, 2001.
15. Amy Glen, Jacques Justin, Steve Widmer, and Luca Q. Zamboni. Palindromic richness. *European J. Combin.*, 30(2):510–531, 2009.
16. Roger C. Lyndon and Paul E. Schupp. *Combinatorial Group Theory.* Classics in Mathematics. Springer-Verlag, 2001. Reprint of the 1977 edition.
17. Wilhelm Magnus, Abraham Karrass, and Donald Solitar. *Combinatorial group theory.* Dover Publications, Inc., Mineola, NY, second edition, 2004.
18. Edita Pelantová and Štěpán Starosta. Palindromic richness for languages invariant under more symmetries. *Theoret. Comput. Sci.*, 518:42–63, 2014.