Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 18–45

ON THE GROUP OF A RATIONAL MAXIMAL BIFIX CODE

JORGE ALMEIDA, ALFREDO COSTA, REVEKKA KYRIAKOGLOU AND DOMINIQUE PERRIN

ABSTRACT: We give necessary and sufficient conditions for the group of a rational maximal bifix code Z to be isomorphic with the F-group of $Z \cap F$, when F is recurrent and $Z \cap F$ is rational. The case where F is uniformly recurrent, which is known to imply the finiteness of $Z \cap F$, receives special attention. The proofs are done by exploring the connections with the structure of the free profinite monoid over the alphabet of F.

KEYWORDS: maximal bifix code, rational code, group code, syntactic monoid, F-group, Schützenberger group, uniformly recurrent set, free profinite monoid.
MATH. SUBJECT CLASSIFICATION (2010): 20M05, 20E18, 37B10, 68R15.

1. Introduction

Maximal bifix codes (and, more generally, maximal prefix codes and maximal suffix codes) play a central role in the theory of codes [BPR10]. In the past few years, special attention has been given to (bifix, prefix, suffix) codes which may not be maximal but are maximal within some language, usually a recurrent or uniformly recurrent language. This line of research has produced new and strong connections between bifix codes, subgroups of free groups and symbolic dynamical systems (see the survey [Per18] and the series of papers [BDFP$^+$12, BDFD$^+$15a, BDFD$^+$15b, BDFD$^+$15c, BDFD$^+$15d]).

If Z is a thin maximal bifix code and F is a recurrent set, then $X = Z \cap F$ is an F-maximal bifix code, that is, a maximal bifix code within F [BDFP$^+$12]. Moreover, X is finite if F is uniformly recurrent. One can informally speak of a process of relativization, or localization, going from maximal bifix codes to F-maximal bifix codes, when F is recurrent. Inevitably, the study of F-maximal bifix codes leads to a process of relativization of several previously

Received November 7, 2018.
J. Almeida and A. Costa acknowledge partial funding by CMUP (UID/MAT/ 00144/2013) and CMUC (UID/MAT/00324/2013), respectively, which are funded by FCT (Portugal) with national (MCTES) and European structural funds (FEDER) under the partnership agreement PT2020. The work of J. Almeida was carried out in part at Masaryk University, whose hospitality is gratefully acknowledged, with the support of the FCT sabbatical scholarship SFRH/BSAB/142872/2018.

known notions and results about maximal bifix codes. This paper is focused in one of these notions. The group of a rational code Z, denoted by $G(Z)$, is the Schützenberger group of the minimum ideal of the syntactic monoid $M(Z^*)$ of Z^*. This group is an important parameter in the study of rational codes [BPR10]. The relativization of this parameter consists in taking the intersection $X = Z \cap F$ and the Schützenberger group of the minimum $\mathcal{J}$-class that intersects the image of F in the syntactic monoid of X^*. This group, denoted by $G_F(X)$, is the *F-group of X*. How are $G(Z)$ and $G_F(X)$ related? They are not always isomorphic, even if Z is a group code and F is uniformly recurrent. In [BDFP$^+$12] it is shown that if Z is a group code and F is a Sturmian set, then $G(Z)$ and $G_F(X)$ are isomorphic. This result was extended to arbitrary uniformly recurrent tree sets in the manuscript [KP17]. That was done with a novel approach consisting in exploring and applying links between $G(Z)$, $G_F(X)$ and the Schützenberger (profinite) group $G(F)$ of the minimum $\mathcal{J}$-class $J(F)$ of the topological closure of F within the free profinite monoid generated by the alphabet of F, and by taking advantage, with the help of these links, of results on $G(F)$ established in [AC13] and [AC16].

Free profinite monoids have proved in the last few decades to be of major importance in the study of formal (rational) languages [Alm95, RS09]. Their elements, called pseudowords, can be seen as generalizations of words, but the algebraic structure of free profinite monoids is much richer than that of free monoids. The first author established a connection with symbolic dynamics [Alm05a] which led to research on the $\mathcal{J}$-classes of the form $J(F)$, when F is recurrent, and of their maximal subgroups, thereby elucidating structural aspects of free profinite monoids [Alm05a, CS11, AC13, AC16]. The approach followed in [KP17] indicates that it is worthwhile to extend to the theory of codes this connection between free profinite monoids and recurrent sets. In this paper we corroborate this, by further exploring the relationship between $G(Z)$, $G_F(X)$ and $G(F)$. We do it inspired by [KP17], but without depending on results first appearing in that manuscript.

Our main result (Theorem 4.8) gives necessary and sufficient conditions for the isomorphism $G(Z) \simeq G_F(X)$ when Z is a rational maximal bifix code, F is recurrent, and $Z \cap F$ is rational. When Z is a group code and F is uniformly recurrent, we recover the corresponding result from the manuscript [KP17], which in this paper is slightly improved (cf. Theorem 4.13). Moreover, we deduce the isomorphism $G(Z) \simeq G_F(X)$ when Z is a group code and F is a uniformly recurrent connected set (cf. Corollary 4.9). These results are framed under the new notion of *F-charged code*.

The paper is organized as follows. Following the introduction, we have a section of preliminaries. It is divided in several subsections, in order to encompass codes, recurrent sets, free profinite monoids and syntactic monoids, and connections between all of these. The section contains some preparatory results needed for our main contributions. In the third section, we further develop the extension to pseudowords, initiated in [KP17], of the key notion of parse of a word, considering now arbitrary rational codes. In particular, we study the continuity of the function that counts the number of parses, with respect to the discrete topology of the set of natural numbers. This facilitates the development of the fourth section, which presents the main results. In this section, preference was given to the more algebraic characterization of the syntactic monoid as the monoid of classes of words with the same contexts, and its variation in terms of pseudowords (cf. Lemma 2.22). In the manuscript [KP17] the characterization of the syntactic monoid as the transition monoid of the minimal automaton seems to be more notorious. That perspective is also explored in the fifth section of this paper, where we look at $G(Z)$ and $G_F(X)$ as permutation groups and establish conditions under which they are equivalent.

2. Preliminaries

2.1. Codes contained in factorial sets of words. Along this paper, all alphabets are finite, A^* denotes the free monoid generated by the alphabet A, the empty word is denoted by 1, and $A^+ = A^* \setminus \{1\}$ is the free semigroup generated by A.

Recall that a *code* of A^* is a subset X of A^+ such that the submonoid of A^* generated by X is freely generated by X.

Let X be a nonempty subset of A^+. If $X \cap XA^+ = \emptyset$, then X is a code, and it is said to be a *prefix code*. Dually, if $A^+X \cap X = \emptyset$, then X is a *suffix code*. A *bifix code* is a code that is simultaneously a prefix and a suffix code.

See [BPR10] for a systematic study of codes. For the purposes of this paper, the paper [BDFP$^+$12] is also a useful reference.

Let F be a nonempty subset of A^*. A prefix code X is an *F-maximal prefix code* if $X \subseteq F$ and X is not properly contained in any other prefix code contained in F. Replacing the word "prefix" by "suffix" or by "bifix", we obtain the notions of *F-maximal suffix code* and *F-maximal bifix code*, respectively.

A subset X contained in F is *right F-complete* if every element of F is a prefix of an element of X^*. The next proposition is an immediate application of the combination of Propositions 3.3.1 and 3.3.2 from [BDFP$^+$12], which in turn are extensions of results from [BPR10] established for the special case $F = A^*$. Recall that a subset F of A^* is *factorial* if it contains every word of A^* that is a factor of at least one element of F.

Proposition 2.1. *Suppose that F is a factorial subset of A^*, and let X be a prefix code contained in F. Then X is an F-maximal prefix code if and only if X is right F-complete.*

There is a dual definition of *left F-complete* subset of A^*, and a result for suffix codes that is dual to Proposition 2.1.

2.2. Recurrent and uniformly recurrent sets. We say that a factorial subset F of A^* is *recurrent* if $F \neq \{1\}$ (in the definition given in [BDFP$^+$12] one may have $F = \{1\}$) and for every $u, v \in F$ there is $w \in F$ such that $uwv \in F$. A recurrent set is said to be *uniformly recurrent* if, for every $u \in F$, there is a positive integer n such that u is a factor of every element of F with length at least n. A special case of a uniformly recurrent set is that of a *periodic set*, that is a set which is the set of factors of a language of the form u^*, with u a nonempty word. These notions appear in the field of symbolic dynamics, for which we give [LM95, Lot02, Fog02] as references. Indeed, a subset F of A^* is recurrent (respectively, uniformly recurrent) if and only if it is the language of finite words appearing in the elements of an irreducible symbolic dynamical system (respectively, of a minimal symbolic dynamical system) of $A^{\mathbb{Z}}$, and F is periodic when the corresponding symbolic dynamical system is periodic.

We briefly describe an important mechanism for producing uniformly recurrent sets. A *substitution* φ over a finite alphabet A is an endomorphism of A^*. If there is n such that, for all $a \in A$, every letter of A is a factor of $\varphi^n(a)$, then φ is said to be *primitive*. If φ is primitive and not the identity on the free monoid over a one-letter alphabet, then the set $F(\varphi)$ of factors of elements of $\{\varphi^k(a) \mid a \in A, k \geq 1\}$ is a uniformly recurrent subset of A^*.

Example 2.2. *Let $A = \{a, b\}$ and let φ be the substitution over A given by $\varphi(a) = ab$ and $\varphi(b) = a$. This substitution is primitive. It is called the* Fibonacci substitution*. The uniformly recurrent set $F(\varphi)$ is the* Fibonacci set*.*

Example 2.3. *A related example is the* Tribonacci substitution*, the substitution ψ over $A = \{a, b, c\}$, defined by $\psi(a) = ab$, $\psi(b) = ac$ and $\psi(c) = a$. The corresponding uniformly recurrent set $F(\psi)$ is the* Tribonacci set*.*

2.3. The extension graph. Consider a recurrent subset F of A^*. Given $w \in F$, let

$$\begin{aligned} L(w) &= \{a \in A \mid aw \in F\}, \\ R(w) &= \{a \in A \mid wa \in F\}, \\ E(w) &= \{(a, b) \in A \times A \mid awb \in F\}. \end{aligned}$$

The *extension graph* $G(w)$ is the bipartite undirected graph whose vertex set is the union of disjoint copies of $L(w)$ and $R(w)$, and whose edges are the pairs $(a, b) \in E(w)$, with incidence in $a \in L(w)$ and $b \in R(w)$. One says that F is a *tree set* if $G(w)$ is a tree for every $w \in F$. If $G(w)$ is connected for every $w \in F$, then one says that F is *connected*.

The class of uniformly recurrent tree sets contains the extensively studied class of *Arnoux-Rauzy sets*. The Arnoux-Rauzy sets over two letters, of which the Fibonacci set is an example, are the well known *Sturmian sets*. The Tribonacci set is also an example of an Arnoux-Rauzy set. See the survey [GJ09] and the research paper [BDFD$^+$15a] (note that in [BDFD$^+$15a] the Arnoux-Rauzy sets are called Sturmian).

Example 2.4. *Here is an example, taken from* [DP18]*, of a uniformly recurrent connected set which is not a tree set. Take the Tribonnaci set $F = F(\psi)$, as in Example 2.3. Consider the morphism α given by $\alpha(a) = \alpha(b) = a$ and $\alpha(c) = c$. Since identifying letters clearly transforms connected extensions graphs into connected extension graphs, $\alpha(F)$ is connected, but, as observed in* [DP18, Example 3.6]*, the extension graph of a^3 in $\alpha(F)$ is not a tree (it is the complete bipartite graph $K_{2,2}$).*

2.4. Parses. A *parse* of a word w with respect to a subset X of A^* is a triple (v, x, u) such that $w = vxu$ with $v \in A^* \setminus A^*X$, $x \in X^*$ and $u \in A^* \setminus XA^*$. The number of parses of w with respect to X is denoted by $\delta_X(w)$.

Remark 2.5. For every set X, the integer $\delta_X(w)$ is greater than or equal to the number of prefixes of w not in A^*X. Equality holds when X is a prefix code (cf. [BPR10, Proposition 6.1.6]).

Lemma 2.6 (cf. [BDFP$^+$12, Subsection 4.1])**.** *If X is a bifix code, then the inequality $\delta_X(v) \leq \delta_X(uvw)$ holds, for every $u, v, w \in A^*$.*

Consider a factorial subset F of A^*. The *F-degree* of X, denoted by $d_F(X)$, is the supremum of the set $\{\delta_X(w) \mid w \in F\}$. It may be infinite. The *degree* of X is $d_{A^*}(X)$, usually denoted by $d(X)$.

In this paper, we pay special attention to bifix codes with finite F-degree, where F is recurrent. A characterization of such codes is given in the next theorem, taken from [BDFP$^+$12]. For its statement, we need some more definitions. When $X, F \subseteq A^*$, one says that X is *F-thin* if there is a word in F that is not a factor of an element of X; a *thin set* is an A^*-thin set. A word $u \in A^*$ is an *internal factor* of a word $w \in A^*$ if $w \in A^+uA^+$.

Theorem 2.7 (cf. [BDFP$^+$12, Theorem 4.2.8])**.** *Let F be a recurrent set and let X be a bifix code contained in F. The F-degree of X is finite if and only if X is an F-thin and F-maximal bifix code. In this case, a word $w \in F$ is an internal factor of a word of X if and only if $\delta_X(w) < d_F(X)$.*

We turn our attention to rational bifix codes.

Proposition 2.8. *For any recurrent set F, if X is a rational bifix code contained in F, then X is F-thin.*

Proof: It suffices to show that there is a word in X which is not a proper factor of a word of X. Suppose that that is not the case. Then, there is a sequence $(u_n)_n$ of words of X, with strictly increasing lengths, such that, for every $n \geq 1$, one has $u_{n+1} = x_n u_n y_n$ for some words x_n, y_n. Since X is a rational prefix code, applying Proposition 3.2.9 from [BPR10], we conclude that there is n_0 such that $y_n = 1$ for all $n \geq n_0$. And since X is a bifix code, we get $u_n = u_{n_0}$ for all $n \geq n_0$, a contradiction. ∎

Remark 2.9. Combining Theorem 2.7 and Proposition 2.8 we conclude that a rational bifix code, contained in a recurrent set F, has finite F-degree if and only if it is an F-maximal bifix code. In particular, a rational bifix code has finite degree if and only if it is a maximal bifix code — in fact, every rational code is thin [BPR10, Proposition 2.5.20].

We next state some properties of the intersection of a (uniformly) recurrent set with a thin maximal bifix code (equivalently, a bifix code of finite degree, cf. Theorem 2.7).

Theorem 2.10 ([BDFP$^+$12, Theorems 4.2.11 and 4.4.3])**.** *Let Z be a maximal bifix code of A^* with finite degree and let F be a recurrent subset of A^*. The intersection $X = Z \cap F$ is an F-maximal bifix code. One has $d_F(X) \leq d(Z)$, with equality if Z is finite. If, moreover, F is uniformly recurrent, then X is finite.*

A *group code* of A^* is a subset Z of A^* such that Z^* is recognized by a finite *group automaton*, that is, a trim automaton whose initial state is the unique final state and such that the action of each letter of A on the finite set of states is a permutation. Notice that the group automaton defining a group code Z is complete, deterministic and reduced, and so it is the minimal automaton of Z^*. Several properties of group codes can be found in [BDFP$^+$12, Section 6]. For instance, a group code is a maximal bifix code with degree equal to the number of states of the group automaton defining it.

The next result is a property which is explained in detail in the proof of [BDFD$^+$15d, Theorem 5.10]. It is a consequence of the fact that uniformly recurrent tree sets satisfy the so called *finite index basis property*, cf. [BDFD$^+$15c, Theorem 4.4].

Theorem 2.11. *If Z is a group code of A^* and F is a uniformly recurrent tree set with alphabet A, then $d_F(Z \cap F) = d(Z)$.*

In Section 4 a generalization of Theorem 2.11 is deduced for arbitrary uniformly recurrent connected sets (cf. Corollary 4.9), and actually for a more general setting expressed in Theorem 4.8. Note that the finite index basis property applies only for uniformly recurrent tree sets (cf. [BDFD$^+$15c, Corollary 4.11]).

In contrast, we have the following example. It concerns a uniformly recurrent set which is not connected, but that belongs to the class of *eventually tree sets* (studied in [DP18]), meaning that the extension graph of every sufficiently long word is a tree.*

Example 2.12. *Consider the group automaton over the alphabet $A = \{a, b, c, d\}$ presented in Figure 1, with initial state* 1*.*

Let Z be the group code defined by this group automaton. Consider the set $Y = \{ab, ac, bc, cd, ca, da\}$. If $u \in Y^* \cup Y^*A$, then every path in the automaton which is labeled by u, and starts at state 3*, passes only through*

*In [DP18], and also in [Per18], the subshifts corresponding to (eventually) tree sets are called *(eventually) dendric subshifts.*

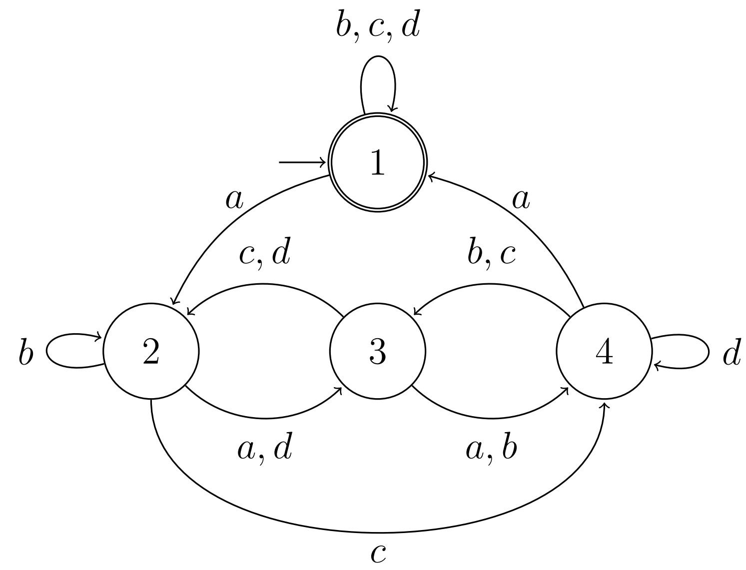

FIGURE 1. A group automaton of degree 4.

states belonging to $\{2,3,4\}$*. Hence, every element of* $Y^* \cup Y^*A$ *is an internal factor of an element of* Z*.*

Let φ *be the substitution over* A *given by*

$$\varphi(a) = ab, \quad \varphi(b) = cda, \quad \varphi(c) = cd \quad \text{and} \quad \varphi(d) = abc.$$

This substitution is borrowed from [BDFD+15a, Example 3.4]*. Let* F *be the uniformly recurrent subset of* A^* *defined by* φ*. If* $u \in F \setminus \{1\}$*, then* $G(u)$ *is a tree, but* $G(1)$ *is acyclic with two connected components, displayed in Figure 2. Notice that* $F \subseteq Y^* \cup Y^*A$*, and so every element of* F *is an*

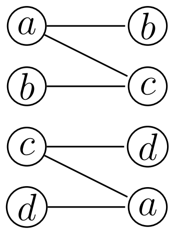

FIGURE 2. The extension graph $G(1)$, with $L(1)$ on the left column and $R(1)$ on the right column.

internal factor of an element of Z*. Hence, we have* $d_F(Z \cap F) < d(Z)$*.*

2.5. Restriction to rational languages. In this paper we are concerned with intersections of the form $X = Z \cap F$, in which Z is a maximal bifix code of A^* and F is a recurrent subset of A^*, for some finite alphabet A. Our results restrict to the case where both Z and X are rational. The rationality of Z implies that of X when the recurrent set F is in one of the following two situations: F is rational, or F is uniformly recurrent, and in the second case X is even finite, as seen in Theorem 2.10 (cf. Remark 2.9). From the viewpoint of symbolic dynamics, this corresponds to the two most studied classes of symbolic dynamical systems: sofic systems (corresponding to rational recurrent sets) and minimal systems (corresponding to uniformly recurrent sets). Out of this realm, it is possible to have Z rational and X not rational, as seen in the next example.

Example 2.13. *Consider the group code Z whose minimal automaton is represented in Figure 3, with initial state* 1*. Suppose that a, b, c and d*

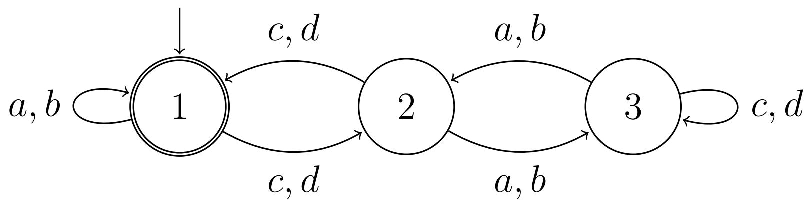

FIGURE 3. A group automaton of degree 3.

represent respectively the four parentheses "(", ")", "[" and "]", and that F is the recurrent set of factors of words of $\{a, b, c, d\}^*$ that are correctly parenthesized. The set F defines a Dyck shift *(a class of symbolic dynamical systems going back to* [Kri75]*). For every positive integer n, the word ca^nb^nd belongs to $X = Z \cap F$. If X were rational, then, by the Pumping Lemma, we would have $ca^nb^{n+k}d$ for some positive integers n and k. But such a word does not belong to F, whence X is not rational.*

2.6. The free profinite monoid. We introduce free profinite monoids, giving the introductory texts [Alm05b, Pin09] and the books [Alm95, RS09] as supporting references. Let A be a finite alphabet. When u and v are distinct elements of A^*, there is some finite monoid M and some homomorphism $\varphi\colon A^* \to M$ such that $\varphi(u) \neq \varphi(v)$. Denote by $r(u, v)$ the least possible value for the cardinal of M. The function $d\colon A^* \times A^* \to \mathbb{R}^+$ such that

$$d(u,v) = \begin{cases} 2^{-r(u,v)} & \text{if } u \neq v \\ 0 & \text{if } u = v \end{cases}$$

is a metric. It is in fact an ultrametric: $d(u,v) \leq \max\{d(u,w), d(w,v)\}$, for all $u,v,w \in A^*$. The metric space A^* thus defined admits a completion $\widehat{A^*}$, which is actually a compact metric space. The multiplication in A^* is uniformly continuous with respect to d, and so $\widehat{A^*}$ is a topological monoid whose multiplication is the unique continuous extension of the multiplication of A^*. In particular, A^* is a dense submonoid of $\widehat{A^*}$. Intuitively, the elements of $\widehat{A^*}$ can be viewed as generalizations of words over A. The elements of A^* are isolated in the topological space $\widehat{A^*}$. We say that the elements of $\widehat{A^*}$ are *pseudowords* over the alphabet A. The elements of $\widehat{A^*} \setminus A^*$ are the *infinite pseudowords*, while those of A^* are the *finite pseudowords*. The topological monoid $\widehat{A^*}$ is a special example of a profinite monoid. A *profinite monoid* is a compact[†] monoid M such that, if u and v are distinct elements of M, then $\varphi(u) \neq \varphi(v)$ for some continuous homomorphism $\varphi\colon M \to N$, where N is a finite monoid endowed with the discrete topology. Note that the finite monoids are profinite, if endowed with the discrete topology, as we do from hereon. It turns out that $\widehat{A^*}$ is the *free profinite monoid generated by* A, in the following sense: if $\varphi\colon A \to N$ is a mapping into a profinite monoid, then there is a unique extension of φ to a continuous homomorphism $\hat{\varphi}\colon \widehat{A^*} \to N$.

The notion of profinite monoid can be generalized to abstract topological algebras. In particular, we can consider *profinite semigroups* and *profinite groups*, and the latter will appear frequently along this paper. The *free profinite semigroup* generated by a finite alphabet A is denoted by $\widehat{A^+}$. The construction of $\widehat{A^+}$ is entirely similar to that of $\widehat{A^*}$. Moreover, the closed subsemigroup $\widehat{A^*} \setminus \{1\}$ can be identified with $\widehat{A^+}$. Frequently, results concerning $\widehat{A^+}$ have an immediate translation to $\widehat{A^*}$, and vice-versa. The reader should keep this in mind when checking references. Finally, we denote by $F_G(A)$ the free group generated by A, and by $\widehat{F_G}(A)$ the *free profinite group* generated by A, which has $F_G(A)$ as a dense subgroup. The cardinal of A is the *rank* of $\widehat{F_G}(A)$. These notions can be generalized to infinite alphabets, but some care is needed when doing that. Except for mentioning *en passant* an example, we shall not need to consider such generalizations.

Viewing $\widehat{F_G}(A)$ as a profinite monoid, we may consider the canonical projection from $\widehat{A^*}$ onto $\widehat{F_G}(A)$, which is the unique continuous homomorphism $p_G\colon \widehat{A^*} \to \widehat{F_G}(A)$ fixing the elements of A.

[†]We adopt the convention that being compact requires being Hausdorff.

In this paper we explore some connections between structural aspects of the free profinite monoid and the structure of the syntactic monoid (whose definition we recall in Subsection 2.9) of the monoid generated by a code. That is why we next recall some basics of the structural theory of monoids [CP61, Alm95, RS09]. Green's relations $\mathcal{J}$, $\mathcal{R}$ and $\mathcal{L}$ are defined by

$$u \mathrel{\mathcal{J}} v \Leftrightarrow MuM = MvM, \quad u \mathrel{\mathcal{R}} v \Leftrightarrow uM = vM, \quad u \mathrel{\mathcal{L}} v \Leftrightarrow Mu = Mv.$$

The other important Green's relations are $\mathcal{D} = \mathcal{R} \vee \mathcal{L}$ and $\mathcal{H} = \mathcal{R} \cap \mathcal{L}$. An element u of M is *regular* if $u \in uMu$. A $\mathcal{D}$-class is *regular* if it contains a regular element, equivalently, if all its elements are regular. The regular $\mathcal{D}$-classes are the ones that contain idempotents. In the monoid M, the $\mathcal{H}$-classes of idempotents are the maximal subgroups of M, with respect to the inclusion relation (a *subgroup* of M is a subsemigroup of M which is a group). The maximal subgroups contained in the same regular $\mathcal{D}$-class are isomorphic. If M is profinite, then $\mathcal{J} = \mathcal{D}$, and the maximal subgroups of the same regular $\mathcal{D}$-class are isomorphic as profinite groups. Furthermore, we have the following elementary property, isolated for future reference.

Proposition 2.14. *Consider a continuous homomorphism $\varphi\colon M \to N$ of profinite monoids. Let H and H' be maximal subgroups of M contained in the same $\mathcal{D}$-class. Then $\varphi(H)$ and $\varphi(H')$ are isomorphic profinite groups. Moreover, $\varphi(H)$ is a maximal subgroup of N if and only if $\varphi(H')$ is a maximal subgroup of N.*

Proof: The proposition is a direct consequence of Green's Lemma (cf. [RS09, Appendix A]), which in this setting affirms in particular the existence of elements $x, y \in M$ such that $h \in H \mapsto xhy \in H'$ is a continuous isomorphism, with inverse given by $h \in H' \mapsto x'hy' \in H$ for some $x', y' \in M$. ■

The abstract (profinite) group defining the isomorphism class of the maximal subgroups of a regular $\mathcal{D}$-class of a (profinite) monoid is the *Schützenberger group* of the $\mathcal{D}$-class. This notion can be extended to non-regular $\mathcal{D}$-classes, but we shall not need the generalization for this paper.

A profinite monoid M also satisfies the *stability property*, that states that $u \mathrel{\mathcal{J}} ux$ if and only if $u \mathrel{\mathcal{R}} ux$, and that $u \mathrel{\mathcal{J}} xu$ if and only if $u \mathrel{\mathcal{L}} xu$, for every $u, x \in M$. We shall occasionally use this property without reference.

We denote by $\overline{Y}$ the topological closure in $\widehat{A^*}$ of a subset Y of $\widehat{A^*}$. We have the following characterization of rational languages, which offers a glimpse of the usefulness of the free profinite monoid.

Theorem 2.15 (cf. [Alm95, Theorem 3.6.1]). *Let L be a language of A^*. Then $\overline{L}$ is open if and only if L is rational.*

For a proof of the following useful technical result, see [ACCZ17, Section 3]. It is an improved version of [AC09, Lemma 2.5]. It expresses in terms of sequences the property that the multiplication in $\widehat{A^*}$ is an open mapping.

Lemma 2.16. *Let $u, v, w \in \widehat{A^*}$ be such that $w = uv$. Suppose that $(w_n)_n$ is a sequence of elements of $\widehat{A^*}$ converging to w. Then there are factorizations $w_n = u_n v_n$ such that the sequences $(u_n)_n$ and $(v_n)_n$ respectively converge to u and v.*

If S is a compact semigroup, and $s \in S$, then the closed subsemigroup $\overline{s^+}$ of S generated by s contains a unique idempotent, denoted s^ω. It is the neutral element of the unique maximal subgroup K_s of the compact semigroup $\overline{s^+}$. Moreover, K_s is the minimum ideal of $\overline{s^+}$. The element $s \cdot s^\omega$ is denoted $s^{\omega+1}$, and its inverse in K_s is denoted by $s^{\omega-1}$. If S is profinite, then $s^\omega = \lim s^{n!}$.

2.7. The topological closure of a recurrent set. If F is a factorial subset of A^*, then $\overline{F}$ is itself factorial, that is, it contains its factors in $\widehat{A^*}$, and if F is recurrent, then $\overline{F}$ also satisfies the property that $u, v \in \overline{F}$ implies the existence of $w \in \overline{F}$ such that $uwv \in \overline{F}$ [AC09]. Using a standard compactness argument (see [CS11, Proposition 3.6]), one also sees that when F is a recurrent subset of A^*, there is a regular $\mathcal{J}$-class $J(F)$ contained in $\overline{F}$ such that $\overline{F}$ is the set of factors of elements of $J(F)$. It is the unique $\mathcal{J}$-class with these properties, and it is the minimum $\mathcal{J}$-class which has nonempty intersection with $\overline{F}$. In the uniformly recurrent case, $J(F)$ occupies a very special place in the structure of $\widehat{A^*}$, as seen next.

Theorem 2.17 ([Alm05a]). *An element u of $\widehat{A^*} \setminus A^*$ is such that all of its proper factors are elements of A^* if and only if $u \in J(F)$ for some uniformly recurrent set F.*

We denote by $G(F)$ the Schützenberger group of $J(F)$, when F is recurrent. It is shown in [AV06] that if F is periodic, then $G(F)$ is a free profinite group of rank 1, while in [CS11] it is proved that if F is a non-periodic rational recurrent subset of A^*, then $G(F)$ is a free profinite group of rank $\aleph_0$. Concerning uniformly recurrent sets, the identification of $G(F)$ has only been made for some special classes. For instance, if F is a uniformly recurrent tree set with alphabet A, then $G(F)$ is a free profinite group with rank $|A|$, and the restriction to a maximal subgroup of $J(F)$ of the canonical projection from $\widehat{A^*}$ onto $\widehat{F_G}(A)$ is a continuous isomorphism [AC16, Section 6]. More generally, we have the next weaker property for all uniformly recurrent connected sets, a property which we shall invoke later in this paper.

Theorem 2.18. *Let F be a uniformly recurrent connected subset of A^* with alphabet A. If H is a maximal subgroup of $J(F)$ and p_G is the canonical projection from $\widehat{A^*}$ onto $\widehat{F_G}(A)$, then $p_G(H) = \widehat{F_G}(A)$.*

In the next few paragraphs we give a proof of Theorem 2.18. Like the proof for the above mentioned stronger property for uniformly recurrent tree sets deduced in [AC16, Section 6], it uses the important concept of return word. If F is a recurrent set and $u \in F$, then the set of *first return words of F to u* is the prefix code $R_F(u)$ consisting of the words v such that $uv \in F \cap A^*u$ and u is not an internal factor of uv. The set F is uniformly recurrent if and only if $R_F(u)$ is finite for every $u \in F$. The following theorem from [BDFD$^+$15a] is one of the main ingredients in the proof of Theorem 2.18. The alphabet of a factorial set F of words is the set of letters belonging to F.

Theorem 2.19. *If F is a uniformly recurrent connected set with alphabet A, then $R_F(u)$ is a generating set of the free group over A, for every $u \in F$.*

The following tool, encapsulated in [AC16, Lemma 5.3], is the other main ingredient in the proof of Theorem 2.18.

Lemma 2.20. *Let F be a uniformly non-periodic recurrent subset of A^*. If H is a maximal subgroup of $J(F)$, then there is a sequence $(X_n)_{n\geq 1}$ of finite codes of A^* satisfying $H = \bigcap_{n\geq 1} \overline{X_n^+}$, with $X_1^+ \supseteq X_2^+ \supseteq X_3^+ \supseteq \cdots$, and such that, for each $n \geq 1$, there is $u_n \in F$ for which the set X_n is conjugated in $F_G(A)$ with $R_F(u_n)$.*

We are now ready to prove Theorem 2.18.

Proof of Theorem 2.18: If F is a periodic set, then $R_F(u)$ is a singleton for all words u in F with sufficiently large length, and so, accordingly to Theorem 2.19, one has that $A = \{a\}$ is a one-letter alphabet and $F = a^*$. Hence, in that case, H is the set $\widehat{A^*} \setminus A^*$, which projects via p_G onto $\widehat{F_G}(A)$.

Suppose that F is not a periodic set. Let $(X_n)_n$ be a sequence of codes as in Lemma 2.20. Then, by the continuity of the homomorphism p_G, we know that, for every $n \geq 1$, the set $p_G\big(\overline{X_n^+}\big)$ is the closed subgroup of $\widehat{F_G}(A)$ generated by X_n. Using the hypothesis that X_n is conjugated in $F_G(A)$ with $R_F(u_n)$, for some $u_n \in F$, and applying Theorem 2.19, we deduce that $p_G\big(\overline{X_n^+}\big) = \widehat{F_G}(A)$, for every $n \geq 1$. Since the sequence of sets $(\overline{X_n^+})_n$ forms a descending chain for the inclusion, using a standard compactness argument we conclude that the image of the intersection $\bigcap_{n\geq 1} \overline{X_n^+}$ by p_G is $\widehat{F_G}(A)$, that is, $p_G(H) = \widehat{F_G}(A)$. ■

2.8. Stable and unitary sets. A submonoid N of a monoid M is called *stable* if for all $u, v, w \in M$, whenever $u, vw, uv, w \in N$ we have $v \in N$. It is called *right unitary* if for every $u, v \in A^*$, $u, uv \in N$ implies $v \in N$. There is also the dual notion of *left unitary* code. It is well known (cf. [BPR10, Propositions 2.2.5 and 2.2.7]) that a submonoid of A^* is stable if and only if it is generated by a code, and it is right unitary (respectively, left unitary) if and only if it is generated by a prefix code (respectively, by a suffix code).

Proposition 2.21. *Let N be a recognizable submonoid of A^*, and consider its topological closure $\overline{N}$ in $\widehat{A^*}$. The following implications are true:*

(1) if N is stable, then $\overline{N}$ is stable;
(2) if N is right unitary, then $\overline{N}$ is right unitary;
(3) if N is left unitary, then $\overline{N}$ is left unitary.

Proof: The set $\overline{N}$ is clearly a submonoid of $\widehat{A^*}$.

Assume that N is stable. Let $u, v, w \in \widehat{A^*}$ be such that $uv, w, u, vw \in \overline{N}$. Let $(u_n), (v_n)$ and (w_n) be sequences of words converging to u, v and w, respectively, with $u_n, w_n \in N$. Since, by Theorem 2.15, the set $\overline{N}$ is open in $\widehat{A^*}$, we have $u_n v_n, v_n w_n \in \overline{N}$ for all large enough n, and thus also $u_n v_n, v_n w_n \in N$ for all large enough n (indeed, $N = \overline{N} \cap A^*$, as the elements of A^* are isolated in $\widehat{A^*}$). Because N is stable, this implies that $v_n \in N$ for all large enough n, which in turn implies that $v \in \overline{N}$. Thus $\overline{N}$ is stable.

The proofs of the second and third implications are similar. ■

2.9. The syntactic monoid. Fix a finite alphabet A. For a language L of A^*, and for $u \in A^*$, let

$$C_L(u) = \{(x, y) \in A^* \times A^* : xuy \in L\}$$

be the context of u in L. Let $M(L)$ be the syntactic monoid of L and let $\eta_L\colon A^* \to M(L)$ be the corresponding syntactic homomorphism. Recall that $\eta_L(u) = \eta_L(v)$ if and only if $C_L(u) = C_L(v)$.

Sometimes, one considers the *syntactic order* in $M(L)$, which is defined by $\eta_L(u) \leq \eta_L(v)$ if $C_L(u) \subseteq C_L(v)$. This notion goes back to Schützenberger [Sch56], and was rediscovered in [Pin95].[‡] The syntactic order is compatible with multiplication, in the sense that $\eta_L(u) \leq \eta_L(v)$ implies $\eta_L(wu) \leq \eta_L(wv)$ and $\eta_L(uw) \leq \eta_L(vw)$ for every $w \in A^*$. This makes $M(L)$ an *ordered monoid*, a monoid endowed with a partial order compatible with multiplication (see [Pin95] for details). We remark that in a finite ordered monoid the restriction of the partial order to a subgroup is always the identity (cf. [Pin96, Lemma 6.4]).

In case L is a rational language, that is, in case $M(L)$ is finite, we may also consider the unique continuous homomorphism $\hat{\eta}_L$ from $\widehat{A^*}$ to $M(L)$ extending η_L. This leads to the consideration of the set

$$C_{\overline{L}}(u) = \{(x,y) \in \widehat{A^*} : xuy \in \overline{L}\}.$$

The following lemma, proved in [Cos06], will be used several times and without reference.

Lemma 2.22. *Suppose L is a rational language. Then $\hat{\eta}_L(u) \leq \hat{\eta}_L(v)$ if and only if $C_{\overline{L}}(u) \subseteq C_{\overline{L}}(v)$.*

In what follows we consider deterministic trim automata, not necessarily complete. Recall that the syntactic monoid of L is (isomorphic to) the transition monoid of the minimal deterministic automaton $\mathcal{M}_L$ of L. For this reason, when L is rational, if p and q are states of $\mathcal{M}_L$, and if u is an element of $\widehat{A^*}$, we can write $p \cdot u = q$ with the meaning that $p \cdot \hat{\eta}_L(u) = q$. In such a setting, it is straightforward to see that if i is the initial state of $\mathcal{M}_L$ and t is one of its final states, then, for every $u \in \widehat{A^*}$, one has $i \cdot \hat{\eta}_L(u) = t$ if and only if $u \in \overline{L}$. In the proof of the following proposition, we use this fact, as well as the fact that if X is a prefix code, then the initial state i_X of $\mathcal{M}_{X^*}$ is the unique final state of $\mathcal{M}_{X^*}$.

Proposition 2.23. *Let F be a recurrent subset of A^*. Let the subset X of F be a rational F-maximal bifix code of A^*. If e is an idempotent belonging to $\overline{F}$, then $e \in \overline{X^*}$.*

[‡]The reader is warned that in [Pin95], and some other later papers, it is adopted the order which is the reverse of ours and of Schützenberger's, and also of the more recent paper [ACKP15].

Proof: Recall that X is F-thin, by Proposition 2.8. Let $(w_n)_n$ be a sequence of elements of F converging to e. It is shown in [BDFP$^+$12, Theorem 4.2.2] that an F-thin and F-maximal bifix code is left F-complete, whence there is $v_n \in A^*$ such that $v_n w_n \in X^*$. Therefore, if v is an accumulation point of the sequence $(v_n)_n$, we have $ve \in \overline{X^*}$, and so in $\mathcal{M}_{X^*}$ we have $i_X \cdot ve = i_X$. As e is idempotent, it follows that $i_X \cdot e = i_X$, establishing that $e \in \overline{X^*}$. ∎

Consider a rational code X of A^*. Let $J(X)$ denote[§] the minimum ideal of $M(X^*)$, and let $G(X)$ be the Schützenberger group of $J(X)$. One says that $G(X)$ is the *group of* X. This notion is introduced in [BPR10] for the larger class of the so called *very thin codes*, but we restrict to the simpler setting of rational codes, as our main results concern them. The reader should have in mind this when checking the references. Additionally, consider a recurrent subset F of A^*. We shall denote by $J_F(X)$ the regular $\mathcal{J}$-class containing $\hat{\eta}_{X^*}(J(F))$, and by $G_F(X)$ the Schützenberger group of $J_F(X)$. One says that $G_F(X)$ is the *F-group of* X. Using a simple continuity argument one sees that $J_F(X)$ is the minimum $\mathcal{J}$-class of $M(X^*)$ that has nonempty intersection with $\eta_{X^*}(F)$. For that reason, we may say that $J_F(X)$ is the *F-minimum $\mathcal{J}$-class* of $M(X^*)$.

Recall that in a monoid of (partial) transformations, the images of the transformations in the same $\mathcal{J}$-class have the same cardinal, the common cardinal being the *rank* of the $\mathcal{J}$-class. Viewing the syntactic monoid $M(L)$ as a partial transformation monoid acting in the minimal automaton $\mathcal{M}_L$, the rank of a word u in $\mathcal{M}_L$ is the rank of $\eta_L(u)$. The F-degree of a rational F-maximal bifix code is closely related with the rank of the elements of $J_F(X)$, as seen next (for which the reader should have in mind Remark 2.9).

Proposition 2.24 ([BDFP$^+$12, cf. Proposition 7.1.3 and Lemma 7.1.4])**.** *Let F be a recurrent subset and let X be a rational F-maximal bifix code. Take $u \in F$. Then u has rank $d_F(X)$ in the minimal automaton of X^* if and only if $\eta_{X^*}(u) \in J_F(X)$. If $\delta_X(u) = d_F(X)$, then u has rank $d_F(X)$.*

The converse of the last implication in Proposition 2.24 fails: if Z is a group code of A^* of degree d, then every word u of A^* has rank d in the minimal automaton of Z^*.

[§]Notice that X is never recurrent, and so there is no risk of confusion with the notation $J(F)$ when F is a recurrent set.

It is also well known that in a transformation monoid, a subgroup contained in a $\mathcal{J}$-class of rank n embeds in S_n. We shall revisit this fact in Section 5.8. Meanwhile, we register the following corollary of Proposition 2.24.

Corollary 2.25. *Let F be a recurrent subset and let X be a rational F-maximal bifix code. Then, we have $G_F(X) \leq S_{d_F(X)}$. In particular, if Z is a rational maximal bifix code then $G(Z) \leq S_{d(Z)}$.*

3. Parses of a pseudoword

The generalization to pseudowords of the notion of parse was introduced in [KP17]. In this section, we show that for rational sets with finite F-degree, the number of parses is continuous, if we endow $\mathbb{N}$ with the discrete topology. This property and a direct consequence of it are used in Section 4.

3.1. The number of parses of a pseudoword is continuous. Let X be a subset of A^*. A parse of a pseudoword w with respect to X is a triple (v, x, u) such that $w = vxu$ with $v \in \widehat{A^*} \setminus \overline{A^*X}$, $x \in \overline{X^*}$ and $u \in \widehat{A^*} \setminus \overline{XA^*}$. If $w \in A^*$, then this is precisely the notion of parse we already gave for elements of A^*, because for every language L of A^* we have $\overline{L} \cap A^* = L$. Let $\delta_X(w)$ be the number of parses of w, whenever $w \in \widehat{A^*}$.

We begin with a couple of preparatory technical lemmas.

Lemma 3.1. *Let X be a rational subset of A^*, and let $w \in \widehat{A^*}$. Suppose that v is a prefix of w such that $v \in \widehat{A^*} \setminus \overline{A^*X}$ and such that $v = vt$ for some $t \in \widehat{A^*}$ with $t \neq 1$. Consider a positive integer ℓ and let $(w_n)_n$ be a sequence of elements of A^* converging to w. There is n_0 such that if $n > n_0$ then $\delta_X(w_n) > \ell$.*

Proof: Let $u \in \widehat{A^*}$ be such that $w = vu$. Since $w = vt^\ell u$, and in view of Lemma 2.16, we can consider factorizations

$$w_n = v_n t_{n,1} t_{n,2} \cdots t_{n,\ell} u_n$$

such that the sequences $(v_n)_n$, $(t_{n,i})_n$ and $(u_n)_n$ respectively converge to v, t and u, where i is any element of $\{1, \ldots, \ell\}$. As $t \neq 1$, for all large enough n and every $i \in \{1, \ldots, \ell\}$, we have $t_{n,i} \neq 1$. For each $j \in \{0, 1, \ldots, \ell\}$, let $v_{n,j} = v_n t_{n,1} t_{n,2} \cdots t_{n,j}$, with $v_{n,0} = v_n$. Notice that $\lim v_{n,j} = vt^j = v$. Since X is rational, the set $\overline{A^* \setminus A^*X} = \widehat{A^*} \setminus \overline{A^*X}$ is an open neighborhood of v, and so $v_{n,j} \in A^* \setminus A^*X$ for all large enough n and every $j \in \{0, 1, \ldots, \ell\}$. Because $t_{n,j} \neq 1$ for every $j \in \{1, \ldots, \ell\}$, the set $\{v_{n,0}, v_{n,1}, \ldots, v_{n,\ell}\}$ has $\ell + 1$

elements, all of them in $A^* \setminus A^*X$, and all of them prefixes of w_n, for all large enough n. This implies that $\delta_X(w_n) \geq \ell + 1$, for all large enough n. ∎

Lemma 3.2. *Consider a rational subset X of A^*. Let $w \in \widehat{A^*}$. Suppose that $(w_n)_n$ is a sequence of elements of A^* converging to w. For every $d \in \mathbb{N}$, if the set $\{\delta_X(w_n) : n \in \mathbb{N}\}$ is bounded and the set $\{n \in \mathbb{N} : \delta_X(w_n) \geq d\}$ is infinite, then the inequality $\delta_X(w) \geq d$ holds. Conversely, if $\delta_X(w) \geq d$, then $\delta_X(w_n) \geq d$ for all large enough n. Moreover, if the integer d satisfies $\delta_X(w) = d$, then $\delta_X(w_n) = d$ for all large enough n.*

Proof: Suppose that the set $\{\delta_X(w_n) : n \in \mathbb{N}\}$ is bounded and that the set $\{n \in \mathbb{N} : \delta_X(w_n) \geq d\}$ is infinite. Clearly, by taking subsequences, we are reduced to consider the case where this set is precisely $\mathbb{N}$. For each n, and with respect to X, consider d distinct parses $p_{n,i} = (\alpha_{n,i}, \beta_{n,i}, \gamma_{n,i})$ of w_n, where $i \in \{1, \ldots, d\}$. By compactness, there is a strictly increasing sequence of integers $(n_k)_k$ such that the sequence

$$(p_{n_k,1}, p_{n_k,2}, \ldots, p_{n_k,d})_k \tag{3.1}$$

converges in $(\widehat{A^*} \times \widehat{A^*} \times \widehat{A^*})^d$ to a d-tuple $(p_1, p_2, \ldots, p_d)$. For each $i \in \{1, \ldots, d\}$, let $p_i = (\alpha_i, \beta_i, \gamma_i)$. Notice that, since the sets $\overline{A^* \setminus A^*X}$, $\overline{X^*}$, and $\overline{A^* \setminus XA^*}$ are closed, the triple p_i is a parse of w.

Let i, j be distinct elements of $\{1, \ldots, d\}$. For each n, there is $t_n \in A^*$ such that $\alpha_{n,j} = \alpha_{n,i} t_n$ or such that $\alpha_{n,i} = \alpha_{n,j} t_n$.

Suppose that $t_{n_k} \neq 1$ for infinitely many k, and let t be an accumulation point of $(t_{n_k})_k$. Then $t \neq 1$, and at least one of the equalities $\alpha_j = \alpha_i t$ or $\alpha_i = \alpha_j t$ holds. By Lemma 3.1, and since $\{\delta_X(w_n) : n \in \mathbb{N}\}$ is bounded, we must have $\alpha_j \neq \alpha_i$, whence $p_i \neq p_j$.

On the other hand, if $t_{n_k} = 1$, that is, if $\alpha_{n_k,i} = \alpha_{n_k,j}$, then, because $p_{n_k,i} \neq p_{n_k,j}$, we must have $\gamma_{n_k,i} \neq \gamma_{n_k,j}$. Therefore, if $t_{n_k} \neq 1$ for only finitely many k, then for infinitely many k we have either $\gamma_{n_k,i} \in A^+\gamma_{n_k,j}$ or $\gamma_{n_k,j} \in A^+\gamma_{n_k,i}$. By compactness, this implies the existence of some $s \in \widehat{A^*} \setminus \{1\}$ such that $\gamma_i = s\gamma_j$ or $\gamma_j = s\gamma_i$. Applying the dual of Lemma 3.1, we deduce that $\gamma_i \neq \gamma_j$, thus $p_i \neq p_j$.

All cases considered, we conclude that $p_i \neq p_j$ whenever $i \neq j$. Therefore, we have $\delta_X(w) \geq d$.

Conversely, suppose that $\delta_X(w) \geq d$. Consider a set

$$\bigl\{(v_i, x_i, u_i) : i \in \{1, \ldots, d\}\bigr\}$$

of d parses of w with respect to X. By Lemma 2.16, we can consider factorizations

$$w_n = v_{n,i}x_{n,i}u_{n,i}$$

such that $\lim_{n\to\infty}(v_{n,i}, x_{n,i}, u_{n,i}) = (v_i, x_i, u_i)$. In particular, for all large enough n, the set

$$\big\{(v_{n,i}, x_{n,i}, u_{n,i}) : i \in \{1,\ldots,d\}\big\} \tag{3.2}$$

has d elements. Since the sets $\overline{A^*\setminus A^*X}$, $\overline{X^*}$ and $\overline{A^*\setminus A^*X}$ are open, for all large enough n we have $v_{n,i} \in A^*\setminus A^*X$, $x_{n,i} \in X^*$ and $u_{n,i} \in A^*\setminus XA^*$. Therefore, the elements of (3.2) are distinct parses of w_n, thus $\delta_X(w_n) \geq d$, for all large enough n.

Suppose moreover that $\delta_X(w) = d$. Then $\delta_X(w_n) \geq d$ for all large enough n. If the set $\{n \in \mathbb{N} : \delta_X(w_n) \geq d+1\}$ were infinite, then by what was already proved, we would have $\delta_X(w) \geq d+1$. Therefore, we must have $\delta_X(w_n) = d$ for all large enough n. ∎

We are now ready for the main result of this section.

Proposition 3.3. *Consider a factorial set F of A^*. Let X be a rational subset of F with finite F-degree d. Then $\delta_X(w) \leq d$ for every $w \in \overline{F}$, and the mapping $\delta_X\colon \overline{F} \to \{1,\ldots,d\}$ thus defined is continuous when we endow $\{1,\ldots,d\}$ with the discrete topology.*

Proof: Let $w \in \overline{F}$. Suppose that $(w_n)_n$ is a sequence of elements of F converging to w. We first claim that $\delta_X(w) = \delta_X(w_n)$ for all sufficiently large n. For each $i \in \mathbb{N}$, let

$$W_i = \{n \in \mathbb{N} : \delta_X(w_n) = i\}.$$

Since X has finite F-degree d, the set

$$I = \{i \in \mathbb{N} : W_i \text{ is infinite}\}$$

is nonempty and its maximum M is at most d. By Lemma 3.2, we know that $\delta_X(w) \geq M$. If we had $\delta_X(w) \geq M+1$ then, again by Lemma 3.2, the set W_{M+1} would be infinite, contradicting the maximality of M. Therefore, we have $\delta_X(w) = M$. By Lemma 3.2, this implies that $\delta_X(w_n) = M$ for all large enough n, proving the claim.

Let $k \in \{1,\ldots,d\}$. Consider a sequence $(u_n)_n$ of elements $\widehat{A^*}$ converging to u and such that $\delta_X(u_n) = k$ for every n. By what was proved in the previous paragraph, for each n we can find $v_n \in A^*$ such that $d(u_n, v_n) < \frac{1}{n}$ and $\delta_X(v_n) = k$. As

$$d(v_n, u) \leq \max\{d(v_n, u_n), d(u_n, u)\} \xrightarrow[n\to\infty]{} 0,$$

we have $\lim v_n = u$, and so $\delta_X(u) = k$, also by the case considered in the first paragraph. Therefore, $\delta_X^{-1}(k)$ is closed. This shows that δ_X is continuous if we endow $\{1, \ldots, d\}$ with the discrete topology. ■

Remark 3.4. Let X be a subset of A^+, and let $w \in \widehat{A^*}$. Denote by $\tilde{\delta}_X(w)$ the number of prefixes of w not in $\overline{A^*X}$. In the statements of Lemmas 3.1 and 3.2, and of Proposition 3.3, if we replace all occurrences of δ_X by $\tilde{\delta}_X$, then we still have true statements, as is easily seen after a straightforward adaptation of the respective proofs. In particular, in view of Remark 2.5, if X is a rational prefix code of finite F-degree, then $\delta_X(w) = \tilde{\delta}_X(w)$ for every $w \in \widehat{A^*}$. We shall not use these facts.

The following lemma is an easy consequence of Proposition 3.3 that will be used later on.

Lemma 3.5. *Consider a recurrent subset F of A^*, Let X be a rational bifix code of A^* with finite F-degree. For every $u \in J(F)$, we have $\delta_X(u) = d_F(X)$.*

Proof: Let $u \in J(F)$ and $v \in F$ be such that $\delta_X(v) = d_F(X)$. There is a sequence $(u_n)_n$ of elements of F converging to u such that v is a factor of u_n, for every n. From Lemma 2.6 and the maximality of $\delta_X(v)$, we deduce that $\delta_X(v) = \delta_X(u_n)$. Applying Proposition 3.3, we conclude that $\delta_X(u) = d_F(X)$. ■

3.2. The number of parses of $\mathcal{H}$-equivalent pseudowords. We conclude this section with some observations that are not applied in the rest of the paper, but that the reader may find interesting.

Let $v \in \widehat{A^*}$. Say that v is *left-cancelable* if the following property holds: $v = vt$ implies $t = 1$. Because $\widehat{A^*}$ is equidivisible (proved in [AC09], see [AC17]), the pseudoword v is left-cancelable if and only if, for every $x, y \in \widehat{A^*}$, the equality $vx = vy$ implies $x = y$. Therefore, if v is left-cancelable and $u \in v \cdot \widehat{A^*}$, then there is a unique $w \in \widehat{A^*}$ such that $u = vw$. We denote such w by $v^{-1}u$. There is a dual notion of *right-cancelable* pseudoword v, for which, whenever $u \in \widehat{A^*} \cdot v$, we use the notation uv^{-1} for the unique element w of $\widehat{A^*}$ such that $u = wv$.

Proposition 3.6. *Consider a factorial subset F of A^*. Let X be a rational subset of A^* with finite F-degree. For every $u \in \overline{F}$, the number $\delta_X(u)$ is equal to the number of pairs (s,p) of pseudowords such that, for some pseudoword x, the triple (s,x,p) is a parse of u with respect to X. When (s,p) is such a pair, s is left-cancelable and p is right-cancelable.*

Proof: Suppose that (s,x,p) and (s,y,p) are parses of u with respect to X. By Lemma 3.1, if $s = st$ for some $t \neq 1$, then there are words in F whose image under δ_X is arbitrarily large, contradicting the assumption that X has finite F-degree. Hence, s is left-cancelable. Dually, p is right-cancelable. As $sxp = syp$, it follows that $x = y$. Hence, a parse of an element of $\overline{F}$ with respect to X is determined by the first and third component. ∎

The following result establishes a relationship between δ_X and η_{X^*}.

Proposition 3.7. *Consider a factorial subset F of A^*. Let X be a rational subset of A^* with finite F-degree. Suppose that the pseudowords u and v of $\widehat{A^*}$ are such that $u \in \overline{F}$, $u = euf$ and $v = evf$, with e, f idempotents and $e \mathrel{\mathcal{R}} u \mathrel{\mathcal{L}} f$. If $\hat{\eta}_{X^*}(u) \leq \hat{\eta}_{X^*}(v)$ holds, then we have $\delta_X(u) \leq \delta_X(v)$.*

Proof: Let (s,w,p) be a parse of u with respect to X. Then s is a prefix of e and p is a suffix of f. By Proposition 3.6, s is left-cancelable, and p is right-cancelable. Therefore, we may consider the pseudowords $s^{-1}e$ and ep^{-1}. Moreover, as $s(s^{-1}e)u(fp^{-1})p = euf = swp$, we have $w = (s^{-1}e)u(fp^{-1})$. As w belongs to $\overline{X^*}$, and $\hat{\eta}_{X^*}(u) \leq \hat{\eta}_{X^*}(v)$, it follows that the pseudoword $(s^{-1}e)v(fp^{-1})$ also belongs to $\overline{X^*}$. Hence, $(s,(s^{-1}e)v(fp^{-1}),p)$ is a parse of v with respect to X. By Proposition 3.6, we conclude that $\delta_X(u) \leq \delta_X(v)$. ∎

Corollary 3.8. *Consider a factorial subset F of A^*. Let X be a rational subset of A^* having finite F-degree. Suppose that u and v are $\mathcal{H}$-equivalent with elements of $\overline{F}$. If $\hat{\eta}_{X^*}(u) = \hat{\eta}_{X^*}(v)$ holds, then we have $\delta_X(u) = \delta_X(v)$.*

Proof: If u and v are not regular, then $u = v$ by [RS01, Corollary 13.2]. Assuming that u and v are $\mathcal{H}$-equivalent regular pseudowords, we can consider idempotents e and f such that u and v are $\mathcal{R}$-equivalent to e and $\mathcal{L}$-equivalent to f. We may then apply Proposition 3.7 to establish the result. ∎

4. Main results

For a subset Y of $\widehat{A^*}$, say that a pseudoword u is *forbidden in* Y if u is not a factor of an element of Y. For the next proposition, recall that, if Z is a rational and maximal bifix code, then $d(Z)$ is finite (cf. Remark 2.9).

Proposition 4.1. *Let Z be a rational maximal bifix code of A^*. Suppose that F is a recurrent subset of A^* and that the intersection $X = Z \cap F$ is rational. The equality $d_F(X) = d(Z)$ holds if and only if the elements of $J(F)$ are forbidden in $\overline{Z}$. Moreover, if $d_F(X) = d(Z)$ then $\hat{\eta}_{Z^*}(J(F)) \subseteq J(Z)$.*

Proof: Suppose that $d_F(X) = d(Z)$. Let $u \in J(F)$. Suppose that u is a factor of an element z of $\overline{Z}$. Take $w \in F$. Then there are $a, b \in A$ such that awb is a factor of u, and thus of z. Therefore, since $\overline{A^*awbA^*}$ is open in $\widehat{A^*}$, the word awb is a factor of an element of Z. This implies that $\delta_Z(w) < d(Z)$, by Theorem 2.7. Since one clearly has $\delta_X(w) = \delta_Z(w)$, it follows that $d_F(X) < d(Z)$, contradicting that $d_F(X) = d(Z)$. Therefore, if $d_F(X) = d(Z)$, then every element of $J(F)$ is forbidden in $\overline{Z}$.

Conversely, suppose that the elements of $J(F)$ are forbidden in $\overline{Z}$. Let $u \in J(F)$. Consider a sequence $(u_n)_n$ of elements of F converging to u. For all sufficiently large n, we know that u_n is not an internal factor of some element of Z. Hence, by Theorem 2.7, we have $\delta_Z(u_n) = d(Z)$, for all sufficiently large n. Notice that, as $u_n \in F$, we may write $\delta_Z(u_n) = \delta_X(u_n)$. Applying Proposition 3.3, we conclude that $\delta_X(u) = d(Z)$. On the other hand, by Lemma 3.5, we have $\delta_X(u) = d_F(X)$. This establishes the converse implication of the equivalence in the statement of the proposition.

Finally, since $\delta_Z(u_n) = d(Z)$, we have $\hat{\eta}_{Z^*}(u_n) \in J(Z)$ by Proposition 2.24, thus $\hat{\eta}_{Z^*}(u) \in J(Z)$. ∎

The notion of forbidden pseudoword in a set was introduced having in mind the following key proposition.

Proposition 4.2. *Let Z and F be subsets of A^*, with F factorial, and let $X = Z \cap F$. Suppose, moreover, that Z^* and X^* are rational. Let e, f be idempotents of $\widehat{A^*}$, and let $u \in \overline{F}$ and $v \in \widehat{A^*}$ be such that $u = euf$ and $v = evf$. The following implications hold:*

(1) $\hat{\eta}_{X^*}(u) \leq \hat{\eta}_{X^*}(v) \Rightarrow \hat{\eta}_{Z^*}(u) \leq \hat{\eta}_{Z^*}(v)$, if e and f are forbidden in $\overline{Z}$;
(2) $\hat{\eta}_{Z^*}(v) \leq \hat{\eta}_{Z^*}(u) \Rightarrow \hat{\eta}_{X^*}(v) \leq \hat{\eta}_{X^*}(u)$, if e and f are forbidden in $\overline{X}$.

Proof: Suppose that $\hat{\eta}_{X^*}(u) \leq \hat{\eta}_{X^*}(v)$. Consider pseudowords $\alpha, \beta \in \widehat{A^*}$ such that $\alpha u \beta \in \overline{Z^*}$. Because $u = euf$, and in view of Lemma 2.16, we may consider sequences (α_n), (β_n), (e_n), (f_n), (u_n) and (v_n), of words in A^*, with $e_n u_n f_n \in F$, respectively converging to α, β, e, f, u and v. It follows that $\lim \alpha_n e_n u_n f_n \beta_n = \alpha u \beta \in \overline{Z^*}$. Because Z^* is rational, we may as well assume that $\alpha_n e_n u_n f_n \beta_n \in Z^*$ for all n. Since, by hypothesis, the pseudowords e and

f are not factors of elements of $\overline{Z}$, there is n_0 such that for all $n > n_0$ there are factorizations $e_n = e_{n,1}e_{n,2}$ and $f_n = f_{n,1}f_{n,2}$ for which the words $\alpha_n e_{n,1}$, $e_{n,2}u_n f_{n,1}$, and $f_{n,2}\beta_n$ all belong to Z^*. Hence, as $e_{n,2}u_n f_{n,1} \in F$, we have $e_{n,2}u_n f_{n,1} \in X^*$, when $n > n_0$. Let (e_1, e_2, f_1, f_2) be an accumulation point of the tuple $(e_{n,1}, e_{n,2}, f_{n,1}, f_{n,2})$. We then have $e_2 u f_1 \in \overline{X^*}$ and $\alpha e_1, f_2\beta \in \overline{Z^*}$. Applying the hypothesis $\hat{\eta}_{X^*}(u) \leq \hat{\eta}_{X^*}(v)$, we obtain $e_2 v f_1 \in \overline{X^*}$. Therefore, we have

$$\alpha v\beta = \alpha e_1 \cdot e_2 v f_1 \cdot f_2\beta \in \overline{Z^*} \cdot \overline{X^*} \cdot \overline{Z^*} \subseteq \overline{Z^*},$$

thereby establishing that $\hat{\eta}_{Z^*}(u) \leq \hat{\eta}_{Z^*}(v)$.

Conversely, suppose that $\hat{\eta}_{Z^*}(v) \leq \hat{\eta}_{Z^*}(u)$. Let $\alpha, \beta \in \widehat{A^*}$ be such that $\alpha v\beta \in \overline{X^*}$. We may consider sequences (α_n), (β_n), (e_n), (f_n), (u_n) and (v_n), of words in A^*, with $e_n u_n f_n \in F$, respectively converging to α, β, e, f, u and v. We have $\lim \alpha_n e_n v_n f_n \beta_n = \alpha v\beta \in \overline{X^*}$. Since X^* is rational, we may as well assume that $\alpha_n e_n v_n f_n \beta_n \in X^*$ for all n. As e, f are forbidden in $\overline{X}$, we conclude that, for some n_0, and every $n > n_0$, there are factorizations $e_n = e_{n,1}e_{n,2}$ and $f_n = f_{n,1}f_{n,2}$ such that the words $\alpha_n e_{n,1}$, $e_{n,2}v_n f_{n,1}$, and $f_{n,2}\beta_n$ all belong to X^*. Let (e_1, e_2, f_1, f_2) be an accumulation point of the tuple $(e_{n,1}, e_{n,2}, f_{n,1}, f_{n,2})$. We then have $e_2 v f_1 \in \overline{X^*}$ and $\alpha e_1, f_2\beta \in \overline{X^*}$. Applying the hypothesis $\hat{\eta}_{Z^*}(v) \leq \hat{\eta}_{Z^*}(u)$, we obtain $e_2 u f_1 \in \overline{Z^*}$. But, since $e_2 u f_1 \in \overline{F}$ and Z^* is rational, we actually have $e_2 u f_1 \in \overline{X^*}$. We conclude that $\alpha u\beta = \alpha e_1 \cdot e_2 u f_1 \cdot f_2\beta \in \overline{X^*}$. ∎

Corollary 4.3. *Under the assumptions of Proposition 4.2, if H is a maximal subgroup of $J(F)$ whose elements are forbidden in $\overline{X}$, then the group $\hat{\eta}_{X^*}(H)$ is a homomorphic image of the group $\hat{\eta}_{Z^*}(H)$. If, moreover, the elements of H are forbidden in $\overline{Z}$, then $\hat{\eta}_{X^*}(H)$ and $\hat{\eta}_{Z^*}(H)$ are isomorphic.*

Proof: By Proposition 4.2(2), if the elements of H are forbidden in $\overline{X}$, then the correspondence

$$\hat{\eta}_{Z^*}(u) \mapsto \hat{\eta}_{X^*}(u) \quad (u \in H),$$

is a well-defined homomorphism from $\hat{\eta}_{Z^*}(H)$ onto $\hat{\eta}_{X^*}(H)$. By Proposition 4.2(1), this homomorphism is an isomorphism if the elements of H are forbidden in $\overline{Z}$. ∎

Under additional assumptions, we next obtain a closer relationship between the homomorphisms $\hat{\eta}_{X^*}$ and $\hat{\eta}_{Z^*}$ appearing in Proposition 4.2.

Proposition 4.4. *Let F be a factorial subset of A^*. Suppose that Z is a rational prefix code of A^* and that $X = Z \cap F$ is a rational F-maximal prefix code. Consider idempotents e, f that are forbidden in $\overline{Z}$. Let $u \in \overline{F}$ and $v \in \widehat{A^*}$ be such that $u = euf$ and $v = evf$. If we have $\hat{\eta}_{X^*}(u) \leq \hat{\eta}_{X^*}(v)$, then the equality $\hat{\eta}_{Z^*}(u) = \hat{\eta}_{Z^*}(v)$ holds.*

Proof: Suppose that the inequality $\hat{\eta}_{X^*}(u) \leq \hat{\eta}_{X^*}(v)$ holds. By Proposition 4.2(1), we have $\hat{\eta}_{Z^*}(u) \leq \hat{\eta}_{Z^*}(v)$. We proceed to prove that $\hat{\eta}_{Z^*}(v) \leq \hat{\eta}_{Z^*}(u)$. Let $\alpha, \beta \in \widehat{A^*}$ be such that $\alpha v \beta \in \overline{Z^*}$. Since e and f are forbidden in $\overline{Z}$ and by Lemma 2.16, there are factorizations $e = e_1 e_2$ and $f = f_1 f_2$ such that the pseudowords αe_1, $e_2 v f_1$ and $f_2 \beta$ belong to $\overline{Z^*}$. Because f is idempotent and $v = vf$, we have $f = (f f_1) f_2$ and $e_2 v f_1 = e_2 v (f f_1)$, and so we are reduced to the case where $f_1 = f f_1$, which we suppose from hereon to hold. Notice that the pseudoword $e_2 u f_1$ is a factor of u, and thus it belongs to $\overline{F}$. Let $(w_n)_n$ be a sequence of elements of F converging to $e_2 u f_1$. Thanks to Lemma 2.16, there is a factorization $w_n = e_{2,n} u_n f_{1,n}$ such that $\lim e_{2,n} = e$, $\lim u_n = u$ and $\lim f_{1,n} = f_1$. By Proposition 2.1, every element of F is a prefix of an element of X^*. On the other hand, $f_1 = f f_1$ is not a factor of an element of $\overline{X}$, as f itself is not. Hence, taking subsequences, we may suppose that there is a factorization $f_{1,n} = t_n s_n$ such that $e_{2,n} u_n t_n$ is an element of X^* and s_n is a proper prefix of an element of X. Let t and s be respectively an accumulation point of the sequence $(t_n)_n$ and of the sequence $(s_n)_n$. Note that

$$e_2 u t \in \overline{X^*}. \tag{4.1}$$

Applying the hypothesis $\hat{\eta}_{X^*}(u) \leq \hat{\eta}_{X^*}(v)$, we get

$$e_2 v t \in \overline{X^*}. \tag{4.2}$$

On the other hand, by the definition of e_2 and f_1, we have

$$e_2 v t \cdot s = e_2 v f_1 \in \overline{Z^*}. \tag{4.3}$$

Since $\overline{Z^*}$ is right unitary (cf. Proposition 2.21) it follows from (4.2) and (4.3) that $s \in \overline{Z^*}$. This means that $s_n \in Z^*$ for sufficiently large n. Because s_n is a proper prefix of an element of $X \subseteq Z$ and Z is a prefix code, we deduce that $s_n = 1$ for sufficiently large n, thus $s = 1$ and $t = f_1$. Therefore, in view of (4.1), we conclude that $\alpha u \beta = \alpha e_1 \cdot e_2 u f_1 \cdot f_2 \beta \in \overline{Z^*}$, thereby establishing that $\hat{\eta}_{Z^*}(v) \leq \hat{\eta}_{Z^*}(u)$. ∎

We now apply the preceding tools to deduce relationships between the maximal subgroups of $J(F)$, $J(Z)$ and $J_F(Z \cap F)$, for suitable Z and F.

Theorem 4.5. *Let F be a factorial subset of A^*. Suppose that Z is a rational prefix code of A^* and that $X = Z \cap F$ is a rational F-maximal prefix code. Let H be a maximal subgroup of $\widehat{A^*}$ contained in $\overline{F}$. Suppose also that the elements of H are forbidden in $\overline{Z}$. Consider the maximal subgroup H_X of $M(X^*)$ containing $\hat{\eta}_{X^*}(H)$ and the maximal subgroup H_Z of $M(Z^*)$ containing $\hat{\eta}_{Z^*}(H)$. There is an injective homomorphism $\alpha\colon H_X \to H_Z$ such that the diagram*

$$\begin{tikzcd} H \arrow[r, "\hat{\eta}_{Z^*}"] \arrow[d, "\hat{\eta}_{X^*}"'] & H_Z \\ H_X \arrow[ur, "\alpha"'] & \end{tikzcd} \tag{4.4}$$

commutes.

Proof: Denote by e the idempotent of H. Let $x \in e \cdot \widehat{A^*} \cdot e$ be such that $\hat{\eta}_{X^*}(x) \in H_X$. Then $\hat{\eta}_{X^*}(e) = \hat{\eta}_{X^*}(x^\omega)$ holds. Applying Proposition 4.4(1), we conclude that $\hat{\eta}_{Z^*}(e) = \hat{\eta}_{Z^*}(x^\omega)$. Since $x = exe$, it follows that $\hat{\eta}_{Z^*}(x) \in H_Z$.

Suppose that $\hat{\eta}_{X^*}(x) = \hat{\eta}_{X^*}(y) \in H_X$, with x, y both in $e \cdot \widehat{A^*} \cdot e$. As argued in the previous paragraph, we know that $\hat{\eta}_{Z^*}(x)$ and $\hat{\eta}_{Z^*}(y)$ belong to H_Z. On the other hand, again by Proposition 4.4(1), from $\hat{\eta}_{X^*}(e) = \hat{\eta}_{X^*}(yx^{\omega-1})$ we deduce $\hat{\eta}_{Z^*}(e) = \hat{\eta}_{Z^*}(yx^{\omega-1})$. Hence, the equality $\hat{\eta}_{Z^*}(x) = \hat{\eta}_{Z^*}(y)$ holds.

Notice that H_X is the image by $\hat{\eta}_{X^*}$ of the closed submonoid $M = \hat{\eta}_{X^*}^{-1}(H_X) \cap e \cdot \widehat{A^*} \cdot e$. In view of the previous two paragraphs, we conclude that there is a unique homomorphism of groups $\alpha\colon H_X \to H_Z$ such that the diagram

$$\begin{tikzcd} M \arrow[r, "\hat{\eta}_{Z^*}"] \arrow[d, "\hat{\eta}_{X^*}"'] & H_Z \\ H_X \arrow[ur, "\alpha"'] & \end{tikzcd} \tag{4.5}$$

commutes. Suppose that $x \in M$ is such that $\hat{\eta}_{Z^*}(x) = \hat{\eta}_{Z^*}(e)$. Applying Proposition 4.2(2), we obtain $\hat{\eta}_{X^*}(x) \leq \hat{\eta}_{X^*}(e)$. Since $\hat{\eta}_{X^*}(x)$ and $\hat{\eta}_{X^*}(e)$ belong to the same maximal subgroup of $M(X^*)$, it follows that $\hat{\eta}_{X^*}(x) = \hat{\eta}_{X^*}(e)$. Therefore, the homomorphism α is injective. It is a homomorphism for which clearly Diagram (4.4) commutes, since Diagram (4.5) commutes and $H \subseteq M$. ■

Since we want to compare maximal subgroups of the form $G(Z)$ and $G_F(X)$, we are led by Theorem 4.5 to the following couple of definitions. Let F be a recurrent subset of A^*. Say that a rational code Z of A^* is *F-charged* if $\hat{\eta}_{Z^*}$ maps all (equivalently, some) maximal subgroups of $J(F)$ onto maximal subgroups of $J(Z)$ (cf. Proposition 2.14). We also say that a rational code X contained in F is *weakly F-charged* if $\hat{\eta}_{X^*}$ maps all (equivalently, some) maximal subgroups of $J(F)$ onto maximal subgroups of $J_F(X)$. The next proposition gives a vast class of examples of F-charged codes.

Proposition 4.6. *Let Z be a group code of A^*. If F is a uniformly recurrent connected set with alphabet A, then Z is F-charged.*

Proof: Let p_G be the canonical projection from $\widehat{A^*}$ onto $\widehat{F_G}(A)$. Since $M(Z^*)$ is a finite group, we may consider the unique continuous homomorphism $\bar{\eta}_{Z^*}$ from the free profinite group $\widehat{F_G}(A)$ onto $M(Z^*)$ such that $\bar{\eta}_{Z^*} \circ p_G = \hat{\eta}_{Z^*}$. Thanks to Theorem 2.18, we have

$$\hat{\eta}_{Z^*}(H) = \bar{\eta}_{Z^*}(p_G(H)) = \bar{\eta}_{Z^*}(\widehat{F_G}(A)) = M(Z^*) = G(Z),$$

establishing the proposition. ■

With the definitions of F-charged and weakly F-charged code on hand, we state the following corollary of Theorem 4.5.

Corollary 4.7. *Let F be a factorial subset of A^*. Suppose that Z is a rational prefix code of A^* and that $X = Z \cap F$ is a rational F-maximal prefix code. Suppose also that the elements of $J(F)$ are forbidden in $\overline{Z}$. The following conditions are equivalent:*

(1) Z is F-charged;
(2) $G_F(X) \simeq G(Z)$ and X is weakly F-charged;
(3) $|G_F(X)| = |G(Z)|$ and X is weakly F-charged.

Proof: We retain the notation used in Theorem 4.5. If we have the equality $\hat{\eta}_{Z^*}(H) = H_Z$, then the homomorphism α in Diagram (4.4) is onto. As Theorem 4.5 also asserts that α is injective, we conclude that (1)⇒(2) holds.

The implication (2)⇒(3) is trivial.

Finally, suppose that $|G_F(X)| = |G(Z)|$, that is, that $|H_Z| = |H_X|$. Then the injective homomorphism α is onto, because H_Z is finite. Therefore, if $|H_Z| = |H_X|$ and $\hat{\eta}_{X^*}(H) = H_X$, we must have $\hat{\eta}_{Z^*}(H) = H_Z$ by the commutativity of Diagram (4.4), thus (3)⇒(1). ■

We are now ready to show our main result, recovering and generalizing the case, treated in the manuscript [KP17], for group codes and uniformly recurrent sets.

Theorem 4.8. *Let F be a recurrent subset of A^*. Suppose that Z is a rational bifix code of finite degree. Let $X = Z \cap F$ and suppose that X is rational. The following conditions are equivalent:*

(1) Z is F-charged;
(2) $d_F(X) = d(Z)$, $G_F(X) \simeq G(Z)$ and X is weakly F-charged;
(3) $d_F(X) = d(Z)$, $|G_F(X)| = |G(Z)|$ and X is weakly F-charged.

Proof: Let H be a maximal subgroup of $J(F)$. Denote by H_X the maximal subgroup of $M(X^*)$ containing $\hat{\eta}_{X^*}(H)$ and by H_Z the maximal subgroup of $M(Z^*)$ containing $\hat{\eta}_{Z^*}(H)$.

The implication (2)⇒(3) its trivial. If the equality $d_F(X) = d(Z)$ is satisfied, then, by Proposition 4.1, the elements of $J(F)$ are forbidden in $\overline{Z}$ and their image under $\hat{\eta}_{Z^*}$ belongs to $J(Z)$, and so, by Corollary 4.7, we have (3)⇒(1). Also by Corollary 4.7 and Proposition 4.1, to prove the implication (1)⇒(2) it suffices to show that $\hat{\eta}_{Z^*}(H) = H_Z$ implies that $d_F(X) = d(Z)$.

Suppose that $\hat{\eta}_{Z^*}(H) = H_Z$. Let $u \in H$ and let e be the idempotent in H. Suppose that u is a factor of an element of $\overline{Z}$. Let $\alpha, \beta \in \widehat{A^*}$ be such that $\alpha u \beta \in \overline{Z}$. Since $u = eu$, we may as well assume that $\alpha = \alpha e$. Consider the minimal automaton $\mathcal{M}_{Z^*}$ of Z^*, and let i be its initial state, which is also the unique final state. Let $p = i \cdot \alpha$. Since $\mathcal{M}_{Z^*}$ is trim, there is $v \in A^*$ such that $p \cdot v = i$. Notice that $p = p \cdot e$, because $\alpha = \alpha e$. On the other hand, i is fixed by e, by Proposition 2.23. It follows that $p \cdot eve = i$. Because $\hat{\eta}_{Z^*}(H)$ is a maximal subgroup of the minimum ideal $J(Z)$, whose idempotent is $\hat{\eta}_{Z^*}(e)$, there is $w \in H$ such that $\hat{\eta}_{Z^*}(eve) = \hat{\eta}_{Z^*}(w)$. Notice that $i \cdot \alpha w = p \cdot w = p \cdot eve = i$, whence $\alpha w \in \overline{Z^+}$. On the other hand, as u and w both belong to the group H, we may consider $t \in H$ such that $u = wt$. Note that $\alpha u \beta = \alpha w t \beta \in \overline{Z^+ A^+}$. By our hypothesis that $\alpha u \beta$ belongs to $\overline{Z}$, and due to the rationality of Z, we conclude that $Z \cap Z^+ A^+ \neq \emptyset$, contradicting the hypothesis that Z is a bifix code. As the absurd originated from assuming that u is a factor of an element of $\overline{Z}$, in view of Proposition 4.1 this concludes the proof that $d_F(X) = d(Z)$ if $\hat{\eta}_{Z^*}(H)$ is a maximal subgroup of $J(Z)$. ■

Recall that, in the setting of Theorem 4.8, if F is uniformly recurrent, then X is finite (cf. Theorem 2.10).

Corollary 4.9. *Consider a group code Z of A^* and let F be a uniformly recurrent connected set with alphabet A. Take the intersection $X = Z \cap F$. We have $d_F(X) = d(Z)$ and $G_F(X) \simeq G(Z)$.*

Proof: It suffices to invoke Proposition 4.6 and apply Theorem 4.8. ■

Example 4.10. *Let $A = \{a, b\}$ and consider the group code $Z = A^2$, whose syntactic group is the cyclic group of order 2. Let F be the Fibonacci set. Note that Z is F-charged, by Proposition 4.6. Therefore, the bifix code $X = F \cap Z = \{aa, ab, ba\}$ is weakly F-charged by Theorem 4.8. The minimal automaton and the egg-box diagram of the syntactic monoid of X^* are shown in Figure 4. Since $d_F(X) = 2$, the F-minimum $\mathcal{J}$-class of $M(X^*)$*

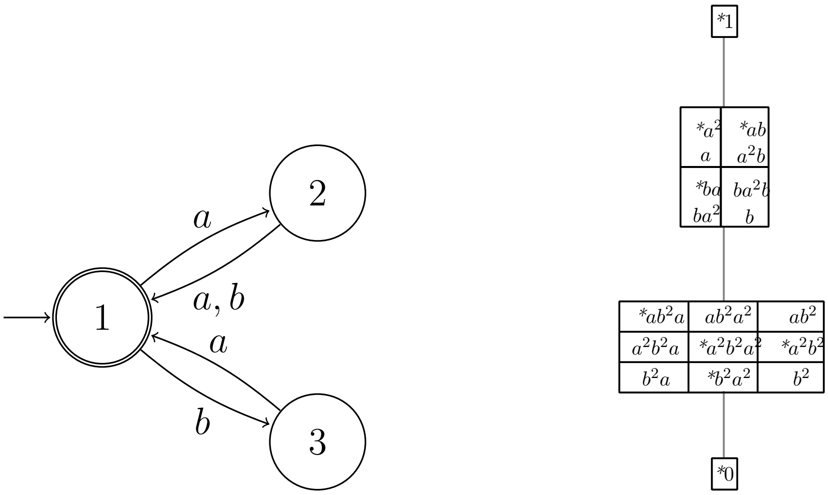

Figure 4. Example 4.10: minimal automaton of X^* and egg-box diagram of its transition monoid.

has rank 2, and so $J_F(X)$ is the $\mathcal{J}$-class of $\eta_{X^*}(a)$. This can be seen also using Corollary 4.9. Indeed, the group $G_F(X)$ is the cyclic group of order 2, as expected by Corollary 4.9, and so $J_F(X)$ must be the unique $\mathcal{J}$-class of $M(X^*)$ with a maximal subgroup of order 2.

The special case of Corollary 4.9 for uniformly recurrent tree sets is established in [KP17] as a consequence of the specialization of Theorem 4.8 for the setting of uniformly recurrent sets and group codes. The special case of Corollary 4.9 in which Z is a group code and F is a Sturmian set was first proved in [BDFP+12].

In the next example we see that both conclusions of Corollary 4.9 fail if we only require F to be uniformly recurrent, even if Z is still a group code.

Example 4.11. *Consider the group code Z and the uniformly recurrent set F of Example 2.12, and let $X = Z \cap F$. We saw in Example 2.12 that $d_F(X) < d(Z) = 4$. Therefore, by Corollary 2.25, we have $G(Z) \leq S_4$ and $G_F(X) \leq S_3$. Actually, a direct computation shows that $G(Z) \simeq S_4$.*

The following is an example of application of Theorem 4.8 to the setting of recurrent sets which are not uniformly recurrent and of maximal bifix codes which are not group codes. It is based on [BDFP$^+$12, Example 4.2.15].

Example 4.12. *Let $A = \{a, b\}$ and $Z = \{aa, ab, ba\} \cup b^2(a^+b)^*b$. The language Z is a maximal bifix code of degree 3. It is not a group code. This can be seen by applying Theorem 2.11, because if F is the Fibonacci set, then $Z \cap F = \{aa, ab, ba\}$ has F-degree 2.*

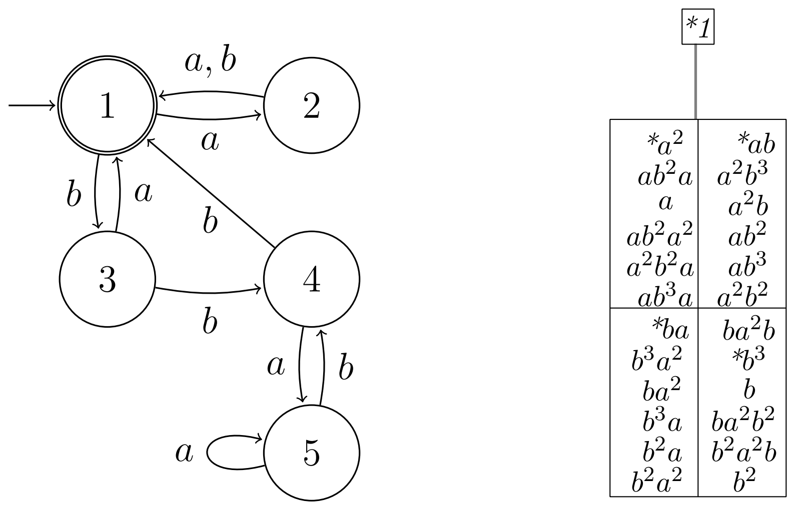

FIGURE 5. Example 4.12: minimal automaton of Z^* and egg-box diagram of its transition monoid.

The monoid $M(Z^*)$ is the disjoint union of its unit with its minimum ideal: see Figure 5. Let E be the recurrent set formed by the words labeling finite paths of the labeled graph $\mathcal{G}$ in Figure 6 (this is the set of factors of the so called even subshift, *see* [LM95]*). The set $X = Z \cap E$ is given by $X = \{aa, ab, ba, bbb\} \cup b^2a^+b^2$. Since $a \in E$ and E is recurrent, there is $u \in J(E)$ such that $u \in a \cdot \widehat{A^*} \cdot a$. Let $v = u^\omega ab^2a^2u^\omega$. Because the action of a in $\mathcal{G}$ consists in just fixing 1, we have $1 = 1 \cdot ab^2a^2 = 1 \cdot u$ in $\mathcal{G}$. Hence, u^ω and v belong to $J(E)$. Denote by H the maximal subgroup of $J(E)$ containing u^ω. Notice that $v \in H$. Similarly, for $w = u^\omega au^\omega$ we have $w \in H$. The maximal subgroup K of $M(Z^*)$ containing $\hat{\eta}_{Z^*}(a)$ is isomorphic*

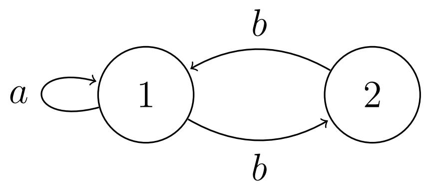

FIGURE 6. Presentation of the even subshift.

to S_3 and it is generated by $\{\hat{\eta}_{Z^*}(a), \hat{\eta}_{Z^*}(ab^2a^2)\}$. Clearly, one has $\hat{\eta}_{Z^*}(v) = \hat{\eta}_{Z^*}(ab^2a^2)$ and $\hat{\eta}_{Z^*}(w) = \hat{\eta}_{Z^*}(a)$. We conclude that $\hat{\eta}_{Z^*}(H) = K$, thus Z is E-charged. Applying Theorem 4.8, we deduce that $d_E(X) = 3$ and that $G_E(X)$ is isomorphic to S_3; see Figure 7 on page 31. The computations of Figures 5 and 7 were carried out using GAP [GAP13, DLM, DM08].

In the case where F is uniformly recurrent, we have the following variation of Theorem 4.8. A finite semigroup is said to be *nil-simple* if all its idempotents lie in the minimum ideal.

Theorem 4.13. *Let F be a uniformly recurrent subset of A^*. Suppose that Z is a rational bifix code of finite degree such that $\eta_{Z^*}(A^+)$ is nil-simple, and let $X = Z \cap F$. The following conditions are equivalent:*

(1) Z is F-charged;
(2) $G_F(X) \simeq G(Z)$ and X is weakly F-charged;
(3) $|G_F(X)| = |G(Z)|$ and X is weakly F-charged.

As examples of rational bifix codes Z such that $\eta_{Z^*}(A^+)$ is nil-simple, we have the finite maximal bifix codes (cf. [BPR10, Theorem 11.5.2]) and the group codes. There are rational maximal bifix codes such that $\eta_{Z^*}(A^+)$ is not nil-simple (cf. [BPR10, Example 11.5.3]).

Proof of Theorem 4.13: Implication (1)⇒(2) holds by Theorem 4.8, and implication (2)⇒(3) is trivial. Suppose that (3) holds. Take a maximal subgroup H of $J(F)$. Let $H_X = \hat{\eta}_{X^*}(H)$ and denote by H_Z the maximal subgroup of $M(Z^*)$ containing $\hat{\eta}_{Z^*}(H)$. Note that $H_Z \subseteq J(Z)$, because the semigroup $\eta_{Z^*}(A^+) = \hat{\eta}_{Z^*}(\widehat{A^+})$ is nil-simple. Since X is finite (cf. Theorem 2.10), the elements of H are forbidden in $\overline{X} = X$. As seen in the proof of Corollary 4.3, it follows from Proposition 4.2(2) that there is a homomorphism $\beta\colon H_Z \to H_X$

$*1$

b

$*b^4a$ b^3a^2 b^4a^2 b^3a b^2a b^2a^2	$b^2a^2b^2$ $*b^3$ b^4 $b^2a^2b^3$ $b^3a^2b^2$ b^2	b^4a^2b $*b^3ab$ b^4ab b^3a^2b b^2a^2b b^2ab
$*a^2$ ab^2a a ab^2a^2 a^2b^2a ab^3a	$*ab^4$ a^2b^3 a^2b^4 ab^2 ab^3 a^2b^2	ab ab^2a^2b a^2b ab^2ab ab^3ab a^2b^2ab
ba ba^2b^2a ba^2 bab^3a bab^2a bab^2a^2	ba^2b^4 $*bab^3$ bab^4 ba^2b^2 ba^2b^3 bab^2	ba^2b $*bab^3ab$ bab ba^2b^2ab bab^2a^2b bab^2ab

$*a^2ba$ aba	$*abab$ a^2bab	a^2bab^3 $abab^3$	$abab^2a$ a^2bab^2a	$abab^2a^2b$ $abab^3ab$	$abab^2a^2$ $abab^3a$	$abab^2$ a^2bab^2	$abab^2ab$ a^2bab^2ab
$*baba$ ba^2ba	ba^2bab $babab$	$*babab^3$ ba^2bab^3	ba^2bab^2a $babab^2a$	$*babab^3ab$ $babab^2a^2b$	$babab^3a$ $babab^2a^2$	ba^2bab^2 $babab^2$	ba^2bab^2ab $babab^2ab$
b^3a^2ba b^3aba	$*b^3abab$ b^3a^2bab	$b^3a^2bab^3$ b^3abab^3	$*b^3abab^2a$ $b^3a^2bab^2a$	$b^3abab^2a^2b$ b^3abab^3ab	$b^3abab^2a^2$ b^3abab^3a	b^3abab^2 $b^3a^2bab^2$	b^3abab^2ab $b^3a^2bab^2ab$
ba^2b^2aba bab^3aba	$*bab^3abab$ ba^2b^2abab	$ba^2b^2abab^3$ bab^3abab^3	bab^3abab^2a $ba^2b^2abab^2a$	$ba^2b^2abab^3ab$ $ba^2b^2abab^2a^2b$	$ba^2b^2abab^3a$ $ba^2b^2abab^2a^2$	bab^3abab^2 $ba^2b^2abab^2$	bab^3abab^2ab $ba^2b^2abab^2ab$
ab^2aba ab^2a^2ba	ab^2a^2bab ab^2abab	$*ab^2abab^3$ $ab^2a^2bab^3$	$*ab^2a^2bab^2a$ ab^2abab^2a	ab^2abab^3ab $ab^2abab^2a^2b$	ab^2abab^3a $ab^2abab^2a^2$	$ab^2a^2bab^2$ ab^2abab^2	$ab^2a^2bab^2ab$ ab^2abab^2ab
b^2aba b^2a^2ba	b^2a^2bab b^2abab	b^2abab^3 $b^2a^2bab^3$	$b^2a^2bab^2a$ b^2abab^2a	b^2abab^3ab $b^2abab^2a^2b$	$*b^2abab^3a$ $b^2abab^2a^2$	$b^2a^2bab^2$ $*b^2abab^2$	$b^2a^2bab^2ab$ $*b^2abab^2ab$
a^2b^2aba ab^3aba	ab^3abab a^2b^2abab	$a^2b^2abab^3$ ab^3abab^3	ab^3abab^2a $a^2b^2abab^2a$	$a^2b^2abab^3ab$ $a^2b^2abab^2a^2b$	$*a^2b^2abab^3a$ $a^2b^2abab^2a^2$	$*ab^3abab^2$ $a^2b^2abab^2$	ab^3abab^2ab $a^2b^2abab^2ab$
bab^2aba bab^2a^2ba	bab^2a^2bab bab^2abab	bab^2abab^3 $bab^2a^2bab^3$	$bab^2a^2bab^2a$ bab^2abab^2a	bab^2abab^3ab $bab^2abab^2a^2b$	bab^2abab^3a $bab^2abab^2a^2$	$bab^2a^2bab^2$ $*bab^2abab^2$	$bab^2a^2bab^2ab$ $*bab^2abab^2ab$

$*abab^2aba$	$abab^2a^2ba$	$abab^2abab$	$abab^2abab^2$	$abab^2abab^2a$	$abab^2abab^2ab$
a^2bab^2aba	$*abab^3aba$	$*a^2bab^2abab$	$a^2bab^2abab^2$	$a^2bab^2abab^2a$	$a^2bab^2abab^2ab$
$babab^2aba$	$*babab^2a^2ba$	$babab^2abab$	$*babab^2abab^2$	$babab^2abab^2a$	$*babab^2abab^2ab$
b^2abab^2aba	$b^2abab^2a^2ba$	$*b^2abab^2abab$	$b^2abab^2abab^2$	$*b^2abab^2abab^2a$	$b^2abab^2abab^2ab$
bab^2abab^2aba	$bab^2abab^2a^2ba$	$*bab^2abab^2abab$	$bab^2abab^2abab^2$	$bab^2abab^2abab^2a$	$bab^2abab^2abab^2ab$
ab^2abab^2aba	$ab^2abab^2a^2ba$	ab^2abab^2abab	$*ab^2abab^2abab^2$	$*ab^2abab^2abab^2a$	$ab^2abab^2abab^2ab$

$*0$

FIGURE 7. Example 4.12: egg-box diagram of $M(X^*)$, a monoid with 221 elements.

such that the diagram

$$\begin{tikzcd} H \arrow[r, "\hat\eta_{Z^*}"] \arrow[d, "\hat\eta_{X^*}"'] & H_Z \arrow[dl, "\beta"] \\ H_X & \end{tikzcd}$$

commutes. Notice that β is onto, as $H_X = \hat\eta_{X^*}(H)$. Therefore, β is an isomorphism, because $|H_X| = |H_Z|$. It follows that $H_Z = \hat\eta_{Z^*}(H)$. This establishes (3)⇒(1), concluding the proof of the theorem. ■

5. F-groups as permutation groups

Let the pair (X, G) denote a faithful right action of a group G on a set X. We also say that (X, G) is a *permutation group* because this action induces an injective homomorphism from G into the symmetric group of permutations on X acting on the right. The cardinal of X is the *degree* of the permutation group (X, G). Two permutation groups (X, G) and (Y, H) are said to be *equivalent* if there is a pair (f, φ) formed by a bijection $f\colon X \to Y$ and an isomorphism $\varphi\colon G \to H$ such that $f(x \cdot g) = f(x) \cdot \varphi(g)$, for every $x \in X$ and $g \in G$. We then say that $(f, \varphi)\colon (X, G) \to (Y, H)$ is an equivalence.

In this section, we show that, loosely speaking, if the conditions of Theorem 4.8 are satisfied, then, in a natural manner, the maximal subgroups of $J_F(X)$ and $J(Z)$ are equivalent permutation groups.

The following lemma will be useful to describe the bijection on the sets where the groups act. Recall that i_X is the initial state and unique final state of the minimal automaton $\mathcal{M}_{X^*}$ of X^*, whenever X is a prefix code.

Lemma 5.1. *Let F be a recurrent subset of A^*. Suppose that Z is a rational bifix code of finite degree, and that the intersection $X = Z \cap F$ is also rational. Consider an $\mathcal{L}$-class K of $J(F)$. For every $u, v \in K$, we have*

$$i_X \cdot u = i_X \cdot v \Leftrightarrow i_Z \cdot u = i_Z \cdot v. \tag{5.1}$$

Proof: Because $J(F)$ is regular, we may take $u' \in J(F)$ such that $u = uu'u$. And since $u \mathrel{\mathcal{L}} v$, there is $t \in J(F)$ such that $v = tu$.

Suppose that $i_Z \cdot u = i_Z \cdot v$. Then we have

$$i_Z \cdot uu' = i_Z \cdot vu' = i_Z \cdot tuu'. \tag{5.2}$$

Note that uu' is idempotent. Therefore, we have $uu' \in \overline{Z^*}$ by Proposition 2.23. Hence, (5.2) simplifies to $i_Z = i_Z \cdot tuu'$, whence $tuu' \in \overline{Z^*}$. Because $uu' \in \overline{Z^*}$, applying Proposition 2.21 we get $t \in \overline{Z^*}$. And since $t \in \overline{F}$ and X

is rational, we conclude that $t \in \overline{X^*}$. Therefore, we have $i_X = i_X \cdot t$, whence $i_X \cdot u = i_X \cdot tu = i_X \cdot v$.

Conversely, suppose that $i_X \cdot u = i_X \cdot v$. Then we have

$$i_X \cdot uu' = i_X \cdot vu' = i_X \cdot tuu'. \tag{5.3}$$

Since uu' is an idempotent factor of $u \in J(F)$ and $\overline{F}$ is factorial, we know that $uu' \in \overline{X^*}$ by Proposition 2.23. Hence, (5.3) simplifies to $i_X = i_X \cdot tuu'$. Therefore, we have $uu', tuu' \in \overline{X^*}$, thus $t \in \overline{X^*}$ by Proposition 2.21. This implies $i_Z = i_Z \cdot t$, whence $i_Z \cdot u = i_Z \cdot tu = i_Z \cdot v$. ■

Consider the set Q of states of the minimal automaton $\mathcal{M}_L$ of L, where L is a rational language. For each element s of the syntactic monoid $M(L)$, let

$$Q \cdot s = \{q \cdot s \mid q \in Q\}.$$

In a transformation monoid, $\mathcal{L}$-equivalent elements have the same image. In particular, in the cases of interest for us, when K is an $\mathcal{L}$-class or a subgroup of $M(L)$, the set

$$Q \cdot K = \{q \cdot g \mid q \in Q \text{ and } g \in K\},$$

is such that $Q \cdot K = Q \cdot g$ for all $g \in K$. Moreover, if K is a subgroup, then K acts faithfully as a permutation group on $Q \cdot K$. As it is well known, if K and K' are maximal subgroups of $M(L)$ contained in the same $\mathcal{J}$-class, then $(Q \cdot K, K)$ and $(Q \cdot K', K')$ are equivalent permutation groups.

Suppose that X is a rational code. We denote by Q_X the set of states of the minimal automaton of $\mathcal{M}_{X^*}$. When $u \in \widehat{A^*}$, then $Q_X \cdot u$ denotes $Q_X \cdot \hat{\eta}_{X^*}(u)$, and if K is an $\mathcal{L}$-class or a subgroup of $\widehat{A^*}$, then $Q_X \cdot K$ denotes $Q_X \cdot \hat{\eta}_{X^*}(K)$, so that in particular we have $Q_X \cdot K = Q_X \cdot u$ whenever $u \in K$. Finally, we denote by $i_X \cdot K$ the set $\{i_X \cdot u \mid u \in K\}$.

Lemma 5.2. *Let F be a recurrent subset of A^*, and let X be a rational bifix code contained in F and with finite F-degree. If K is an $\mathcal{L}$-class of $J(F)$, then the equality $Q_X \cdot K = i_X \cdot K$ holds.*

Our proof of Lemma 5.2 depends on the following property. In its statement, by "a prefix of X" we mean "a prefix of some element of X".

Lemma 5.3 ([BDFP+12, Lemma 7.1.4])**.** *Let F be a recurrent subset of A^*, and let X be a bifix code contained in F and with finite F-degree d. If w is a word of F such that $\delta_X(w) = d$, then, for every $q \in Q_X \cdot w$ there is a unique proper prefix v of X which is a suffix of w, and such that $q = i_X \cdot v$.*

Proof of Lemma 5.2: Take $q \in Q_X$ and $u \in K$. Consider also an element u' of $J(F)$ such that $u = uu'u$. By Lemma 3.5, we have $\delta_X(u) = d_F(X)$. Consider a sequence $(u_n)_n$ of elements of F converging in $\widehat{A^*}$ to u. Taking subsequences, we may as well suppose that $\hat{\eta}_{X^*}(u_n) = \hat{\eta}_{X^*}(u)$ and $\delta_X(u_n) = \delta_X(u) = d_F(X)$ for all n, respectively because of the continuity of $\hat{\eta}_{X^*}$ and of δ_X (cf. Proposition 3.3). In particular, we have $q \cdot u_n = q \cdot u$ for all n. Moreover, by Lemma 5.3, for each n there is a unique proper prefix v_n of X such that v_n is a suffix of u_n and $q \cdot u_n = i_X \cdot v_n$. Let v be an accumulation point of the sequence $(v_n)_n$. Then, v is a suffix of u and $q \cdot u = i_X \cdot v$. Multiplying by $u'u$, we get $q \cdot u = i_X \cdot vu'u$. Since v is a suffix of u, we know that $vu'u$ is a suffix of $uu'u = u$, and so $vu'u$ and u are $\mathcal{L}$-equivalent. This shows that $q \cdot u \in i_X \cdot K$, thus establishing the equality $Q_X \cdot K = i_X \cdot K$. ■

We are ready for the main result of this section. Note that, in view of Theorem 4.8, the group H_X in the next statement is indeed a maximal subgroup of $J_F(X)$.

Theorem 5.4. *Consider a recurrent subset F of A^*. Suppose that Z is an F-charged rational bifix code of finite degree d. Let $X = Z \cap F$ and suppose that X is rational. Let H be a maximal subgroup of $J(F)$. Consider the maximal subgroup $H_Z = \hat{\eta}_{Z^*}(H)$ of $J(Z)$ and the maximal subgroup $H_X = \hat{\eta}_{X^*}(H)$ of $J_F(X)$. Denoting by $\mathcal{L}(H)$ the $\mathcal{L}$-class of $J(F)$ containing H, take the correspondence f from $Q_X \cdot H_X$ to $Q_Z \cdot H_Z$ defined by*

$$f(i_X \cdot u) = i_Z \cdot u, \quad \forall u \in \mathcal{L}(H),$$

together with the unique isomorphism $\alpha \colon H_X \to H_Z$ such that the diagram

$$\begin{tikzcd} H \arrow[r, "\hat{\eta}_{Z^*}"] \arrow[d, "\hat{\eta}_{X^*}"'] & H_Z \\ H_X \arrow[ur, "\alpha"'] & \end{tikzcd}$$

commutes. Then $(f, \alpha) \colon (Q_X \cdot H_X, H_X) \to (Q_Z \cdot H_Z, H_Z)$ is an equivalence between permutation groups of degree d.

Proof: Since $Q_X \cdot H = Q_X \cdot \mathcal{L}(H)$ and $Q_Z \cdot H = Q_Z \cdot \mathcal{L}(H)$, the correspondence f is well defined and injective by the combination of Lemmas 5.1 and 5.2. According to Proposition 2.24, we have $|Q_X \cdot H| = d_F(X)$ and $|Q_Z \cdot H| = d(Z)$. Applying Theorem 4.8, we get that $|Q_X \cdot H| = |Q_Z \cdot H| = d$, establishing in particular that f is a bijection. The existence and uniqueness of the isomorphism α also follows from Theorem 4.8 and its proof, based on Theorem 4.5. Finally, let $q \in Q_X \cdot H$ and $g \in H_X$. Then, there are $u \in \mathcal{L}(H)$ and $v \in H$ such that $q = i_X \cdot u$ and $g = \hat{\eta}_{X^*}(v)$, the former because of Lemma 5.2. Noting that we also have $uv \in \mathcal{L}(H)$, we then have

$$f(q) \cdot \alpha(g) = (i_Z \cdot u) \cdot \hat{\eta}_{Z^*}(v) = i_Z \cdot uv = f(i_X \cdot uv) = f(q \cdot g),$$

thus concluding the proof. ∎

As an example of application of Theorem 5.4, we next deduce a version of Corollary 4.9 for permutation groups. Recall that a permutation group (Y, G) is *transitive* when, for all $x, y \in Y$, there is $g \in G$ such that $x \cdot g = y$. In the following statement, we consider the natural action of a maximal subgroup of $M(X^*)$ on the minimal automaton of X^*.

Corollary 5.5. *Consider a group code Z of A^* and let F be a uniformly recurrent connected set with alphabet A. Take the intersection $X = Z \cap F$. The permutation groups $G(Z)$ and $G_F(X)$ are isomorphic transitive permutation groups of degree $d(Z)$.*

Proof: The group $G(Z)$ acts transitively on the minimal automaton of Z^* as a permutation group of degree $d(Z)$, because the minimal automaton of Z^* is trim and $G(Z)$ is the syntactic monoid of Z^*. Since transitivity is preserved by equivalence of permutation groups, the result follows immediately from Theorem 5.4, together with Proposition 4.6 and Theorem 4.8. ∎

In general, for an arbitrary F-maximal bifix code X with F uniformly recurrent, the natural action of $G_F(X)$ may not be transitive (cf. [BDFP$^+$12, Example 7.2.1]).

We end this section with a description of F-groups as permutation groups defined by suitable sets of first return words, under general conditions that in particular apply to the setting of Theorem 5.4.

Proposition 5.6. *Consider a recurrent subset F of A^*. Let X be a weakly F-charged code of A^*. Let K be a maximal subgroup of $J_F(X)$. Suppose that $u \in F$ is such that $\eta_{X^*}(u)$ is $\mathcal{L}$-equivalent to the elements of K. For each $v \in R_F(u)$, let π_v be the permutation of $Q_X \cdot K$ resulting from the restriction to $Q_X \cdot K$ of the action of v in the minimal automaton $\mathcal{M}_{X^*}$ of X^*. Then the group of permutations of $Q_X \cdot K$ generated by $\{\pi_v \mid v \in R_F(u)\}$ is the group of permutations resulting from the restriction to $Q_X \cdot K$ of the action in $\mathcal{M}_{X^*}$ of the elements of K.*

Proof: We claim that there is some maximal subgroup H of $J(F)$ such that u is a suffix of the elements of H. Take an idempotent e of $J(F)$. Since $u \in F$, there are $x, y \in \widehat{A^*}$ with $e = xuy$. From $e = (exuy)^3$ we get that both $yexu$ and $(yexu)^2$ are $\mathcal{J}$-equivalent to e, and so $yexu$ belongs to a maximal subgroup H of $J(F)$, establishing the claim.

By the hypothesis that X is weakly F-charged, the image $H_X = \hat{\eta}_{X^*}(H)$ is a maximal subgroup of $J_F(X)$. Observe also that K and H_X are maximal subgroups contained in the same $\mathcal{L}$-class, namely the one that contains $\eta_{X^*}(u)$, and that therefore the permutations of $Q_X \cdot H_X$ induced by elements of H_X are precisely the permutations of $Q_X \cdot K = Q_X \cdot H_X$ induced by elements of K.

Let $v \in R_F(u)$. As u is a suffix of v and of the elements of H, and since $\eta_{X^*}(u) \in J_F(X)$, $uv \in F$ and $H_X \subseteq J(F)$, we know that $\eta_{X^*}(uv)$ and $\eta_{X^*}(u)$ both belong to the $\mathcal{L}$-class of $M(X^*)$ containing H_X. In particular, we have $Q_X \cdot uv = Q_X \cdot u = Q_X \cdot H_X$. Therefore, $(Q_X \cdot H_X) \cdot v = Q_X \cdot H_X$ holds. This means that v acts as a permutation on the finite set $Q_X \cdot H_X$. Consider the idempotent f in H. Then we have $\hat{\eta}_{X^*}(f) \mathrel{\mathcal{L}} \eta_{X^*}(u)$, whence $\hat{\eta}_{X^*}(fv) \mathrel{\mathcal{L}} \eta_{X^*}(uv)$. But we saw that we also have $\eta_{X^*}(u) \mathrel{\mathcal{L}} \eta_{X^*}(uv)$, and so $\hat{\eta}_{X^*}(f) \mathrel{\mathcal{L}} \hat{\eta}_{X^*}(fv)$. Since $M(X^*)$ is stable, the latter implies that $\hat{\eta}_{X^*}(fv)$ belongs to the maximal subgroup containing $\hat{\eta}_{X^*}(f)$, that is, to H_X. Clearly, the action of v in $Q_X \cdot H_X$ is the same as the action of fv in $Q_X \cdot H_X$, and so the permutation π_v of $Q_X \cdot H_X$ induced by v is a permutation induced by an element of H_X, namely $\hat{\eta}_{X^*}(fv)$.

Conversely, let $h \in H_X$. Then $h = \hat{\eta}_{X^*}(w)$ for some $w \in H$. Because we also have $w^2 \in H$ and the elements of H have u as a suffix, we know that $uw \in J(F)$. Hence $uw \in \overline{F}$, and we may consider a sequence of words of F converging to uw. Applying Lemma 2.16 together with the continuity of $\hat{\eta}_{X^*}$, we conclude that there is some finite word w' such that $h = \hat{\eta}_{X^*}(w) = \eta_{X^*}(w')$, $uw' \in F$ and $w' \in A^*u$. In particular, $w' \in R_F(u)^+$, and h belongs to the subsemigroup of $M(X^*)$ generated by $\eta_{X^*}(R_F(u))$. Therefore, the permutation on $Q_X \cdot H_X$ induced by h belongs to the group of permutations of $Q_X \cdot H_X$ generated by $\{\pi_v \mid v \in R_F(u)\}$. ■

Example 5.7. *Let us go back to Example 4.10. In this example, we have* $R_F(a) = \{a, ba\}$. *The image of* a *and* ba *on the minimal automaton of* X^*, *represented in Figure 4, is the set* $\{1, 2\}$*, and the set of corresponding permutations, respectively the transposition* $(1, 2)$ *and the identity, generates the group or order* 2*, as predicted by Proposition 5.6.*

Remark 5.8. In Proposition 5.6, while $\{\pi_v \mid v \in R_F(u)\}$ generates a group isomorphic to a maximal subgroup of $J_F(X)$, one may have distinct elements of $R_F(u)$ being mapped into distinct $\mathcal{H}$-classes of $J_F(X)$: see Example 5.7, where $\eta_{X^*}(a)$ and $\eta_{X^*}(ba)$ are not $\mathcal{H}$-equivalent, as one can see in the egg-box diagram of Figure 4.

We end with an example concerning an F-charged group code such that F is a non-connected uniformly recurrent set.

Example 5.9. *Let* $A = \{a, b\}$*. Consider the* Prouhet-Thue-Morse substitution $\tau \colon A^* \to A^*$*, defined by* $\tau(a) = ab$ *and* $\tau(b) = ba$*, and let* $F = F(\tau)$*. The set* F *is not connected: the extension graph of* aba *with respect to* F *has two connected components, since* $A^*abaA^* \cap F = \{aabab, babaa\}$*.*

Let h *be the homomorphism* $h : a \mapsto (123), b \mapsto (345)$ *from* A^* *onto the alternating group* A_5*. It is shown in* [AC13] *that its unique continuous extension* $\hat{h} \colon \widehat{A^*} \to A_5$ *maps each maximal subgroup of* $J(F)$ *onto* A_5 *(cf.* [AC13, proof of Remark 7.8]*).*

Let Z *be the group code generating the submonoid of* A^* *stabilizing* 1 *via the action induced by* h*. This action describes the group automaton of* Z^**, with vertex set* $\{1, 2, 3, 4, 5\}$ *and initial set* 1*, and so* h *is precisely the syntactic homomorphism* η_{Z^*}*. Therefore,* Z *is* F*-charged. It follows from Theorem 4.8 that the* F*-maximal bifix code* $X = Z \cap F$ *is weakly* F*-charged. The code* X *is represented in Figure 8. We represent in Figure 8 only the nodes corresponding to* right special words*, that is, vertices with two sons. The image of* $\tau^4(b)$ *in* $\mathcal{M}_{X^*}$ *is* $\{1, 3, 4, 9, 10\}$*. Since* $\tau^4(b)$ *has rank* 5 *in* $\mathcal{M}_{X^*}$ *and* $d_F(X) = d(Z) = 5$ *by Theorem 4.8, we know that the image of* $\tau^4(b)$ *in* $\mathcal{M}_{X^*}$ *belongs to the* F*-minimum* $\mathcal{J}$*-class* $J_F(X)$*. The action of* $\tau^4(b)$ *on its image is shown in Figure 9. The return words to* $\tau^4(b)$ *are* $\tau^4(b), \tau^3(a)$ *and* $\tau^5(ab)$*. The permutations on the image of* $\tau^4(b)$ *are the three cycles of length* 5 *indicated in Figure 9. They generate the group* $G_F(X) = A_5 = G(Z)$*, in agreement with Theorem 4.8 and Proposition 5.6.*

References

[AC09] J. Almeida and A. Costa, *Infinite-vertex free profinite semigroupoids and symbolic dynamics*, J. Pure Appl. Algebra **213** (2009), 605–631.

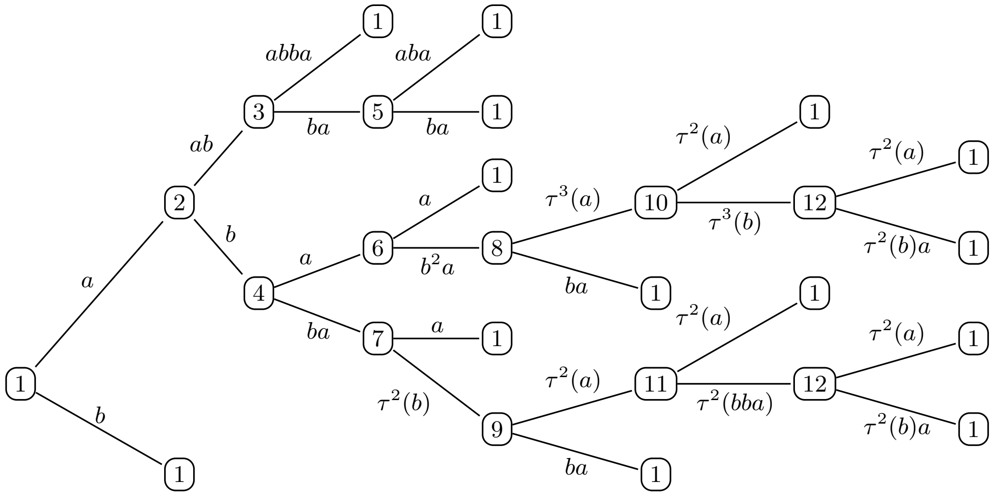

FIGURE 8. The bifix code X from Example 5.9.

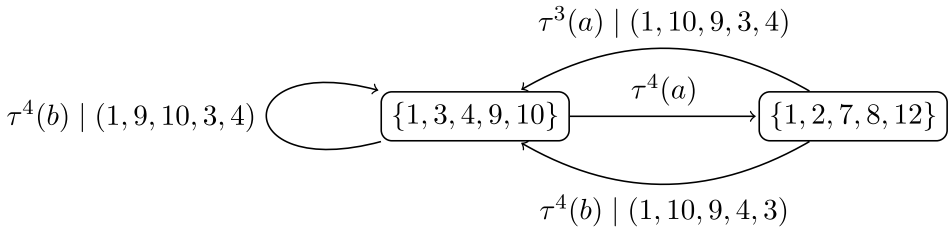

FIGURE 9. The action on the image of $\tau^4(b)$.

[AC13] ______, *Presentations of Schützenberger groups of minimal subshifts*, Israel J. Math. **196** (2013), 1–31.

[AC16] ______, *A geometric interpretation of the Schützenberger group of a minimal subshift*, Ark. Mat. **54** (2016), 243–275.

[AC17] ______, *Equidivisible pseudovarieties of semigroups*, Publ. Math. Debrecen **90** (2017), 435–453.

[ACCZ17] J. Almeida, A. Costa, J. C. Costa, and M. Zeitoun, *The linear nature of pseudowords*, Tech. report, 2017, arXiv:1702.08083 [cs.FL], submitted.

[ACKP15] J. Almeida, A. Cano, O. Klíma, and J.-É. Pin, *On fixed points of the lower set operator*, Int. J. Algebra Comput. **25** (2015), 259–292.

[Alm95] J. Almeida, *Finite semigroups and universal algebra*, World Scientific, Singapore, 1995, English translation.

[Alm05a] ______, *Profinite groups associated with weakly primitive substitutions*, Fundamentalnaya i Prikladnaya Matematika (Fundamental and Applied Mathematics) **11** (2005), 13–48, In Russian. English version in J. Math. Sciences **144**, No. 2 (2007) 3881–3903.

[Alm05b] ______, *Profinite semigroups and applications*, Structural theory of automata, semigroups and universal algebra (New York) (V. B. Kudryavtsev and I. G. Rosenberg, eds.), Springer, 2005, pp. 1–45.

[AV06] J. Almeida and M. V. Volkov, *Subword complexity of profinite words and subgroups of free profinite semigroups*, Int. J. Algebra Comput. **16** (2006), 221–258.

[BDFD+15a] V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, and G. Rindone, *Acyclic, connected and tree sets*, Monatsh. Math. **176** (2015), 521–550.

[BDFD+15b] ———, *Bifix codes and interval exchanges*, J. Pure Appl. Algebra **219** (2015), 2781–2798.

[BDFD+15c] ———, *The finite index basis property*, J. Pure Appl. Algebra **219** (2015), 2521–2537.

[BDFD+15d] ———, *Maximal bifix decoding*, Discrete Math. **338** (2015), no. 5, 725–742.

[BDFP+12] J. Berstel, C. De Felice, D. Perrin, Ch. Reutenauer, and G. Rindone, *Bifix codes and Sturmian words*, J. Algebra **369** (2012), 146–202.

[BPR10] J. Berstel, D. Perrin, and Ch. Reutenauer, *Codes and automata*, Encyclopedia of Mathematics and its Applications, vol. 129, Cambridge University Press, Cambridge, 2010.

[Cos06] A. Costa, *Conjugacy invariants of subshifts: an approach from profinite semigroup theory*, Int. J. Algebra Comput. **16** (2006), 629–655.

[CP61] A. H. Clifford and G. B. Preston, *The algebraic theory of semigroups*, vol. I, Amer. Math. Soc., Providence, R.I., 1961.

[CS11] A. Costa and B. Steinberg, *Profinite groups associated to sofic shifts are free*, Proc. London Math. Soc. **102** (2011), 341–369.

[DLM] M. Delgado, S. Linton, and J. Morais, *Automata: A GAP package on finite automata*, http://www.gap-system.org/Packages/automata.html.

[DM08] M. Delgado and J. Morais, *SgpViz: A GAP package to visualize finite semigroups*, 2008, http://www.gap-system.org/Packages/sgpviz.html.

[DP18] F. Dolce and D. Perrin, *Eventually dendric subshifts*, Tech. report, 2018, arXiv:1807.05124 [math.CO].

[Fog02] N. Pytheas Fogg, *Substitutions in dynamics, arithmetics and combinatorics*, Lect. Notes in Math., vol. 1794, Springer-Verlag, Berlin, 2002.

[GAP13] The GAP Group, *GAP – Groups, Algorithms, and Programming, Version 4.6*, 2013, (`http://www.gap-system.org`).

[GJ09] A. Glen and J. Justin, *Episturmian words: a survey*, Theor. Inform. Appl. **43** (2009), 403–442.

[KP17] R. Kyriakoglou and D. Perrin, *Profinite semigroups*, Tech. report, 2017, arXiv:1703.10088 [math.GR].

[Kri75] W. Krieger, *On the uniqueness of the equilibrium state*, Math. Systems Theory **8** (1974/75), 97–104.

[LM95] D. Lind and B. Marcus, *An introduction to symbolic dynamics and coding*, Cambridge University Press, Cambridge, 1995.

[Lot02] M. Lothaire, *Algebraic combinatorics on words*, Cambridge University Press, Cambridge, UK, 2002.

[Per18] D. Perrin, *Groups, languages and dendric shifts*, Developments in Language Theory. DLT 2018. (Hoshi M. and Seki S., eds.), Lect. Notes in Comput. Sci., vol. 11088, Springer, Cham, 2018.

[Pin95] J.-E. Pin, *A variety theorem without complementation*, Russian Math. (Iz. VUZ) **39** (1995), 80–90.

[Pin96] ———, *Polynomial closure of groups and open sets of the hall topology*, Theor. Comp. Sci. **169** (1996), 185–200.

[Pin09] ______, *Profinite methods in automata theory*, vol. 26th International Symposium on Theoretical Aspects of Computer Science (STACS 2009), Internationales Begegnungs- Und Forschungszentrum für Informatik (IBFI), 2009, pp. 31–50.

[RS01] J. Rhodes and B. Steinberg, *Profinite semigroups, varieties, expansions and the structure of relatively free profinite semigroups*, Int. J. Algebra Comput. **11** (2001), 627–672.

[RS09] ______, *The q-theory of finite semigroups*, Springer Monographs in Mathematics, Springer, 2009.

[Sch56] M.-P. Schützenberger, *Une théorie algébrique du codage*, Séminaire Dubreil-Pisot, année 1955-56, Exposé No. 15, 27 février 1956, 24 pages, Inst. H. Poincaré, Paris, 1956.

Jorge Almeida
CMUP, Departamento de Matemática, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal.
E-mail address: `jalmeida@fc.up.pt`

Alfredo Costa
CMUC, Department of Mathematics, University of Coimbra, Apartado 3008, EC Santa Cruz, 3001-501 Coimbra, Portugal.
E-mail address: `amgc@mat.uc.pt`

Revekka Kyriakoglou
LIGM, Université Paris-Est
E-mail address: `krevekka@hotmail.gr`

Dominique Perrin
LIGM, Université Paris-Est
E-mail address: `dominique.perrin@esiee.fr`