

ON A COMPLETE SET OF OPERATIONS FOR FACTORIZING CODES*

CLELIA DE FELICE¹

Abstract. It is known that the class of *factorizing codes*, *i.e.*, codes satisfying the *factorization conjecture* formulated by Schützenberger, is closed under two operations: the classical *composition* of codes and *substitution* of codes. A natural question which arises is whether a finite set \mathcal{O} of operations exists such that each factorizing code can be obtained by using the operations in \mathcal{O} and starting with prefix or suffix codes. \mathcal{O} is named here a *complete set* of operations (for factorizing codes). We show that composition and substitution are not enough in order to obtain a complete set. Indeed, we exhibit a factorizing code over a two-letter alphabet $A = \{a, b\}$, precisely a 3-code, which cannot be obtained by decomposition or substitution.

Mathematics Subject Classification. 94A45, 68Q45, 20K01.

1. INTRODUCTION

In this paper we follow the algebraic approach initiated by Schützenberger in [40], and *codes* are defined as the *bases* of the free submonoids of A^* . Thus, a subset C of A^* is a code if each word in A^* has at most one factorization into words of C [3, 29, 40]. In spite of their simple definition, the structure of the codes is still unknown and no systematic method for constructing them exists. Conjectures have been given in order to make this structure clearer. In this direction, Schützenberger formulated the conjecture known as the *factorization conjecture*, that given a finite maximal code C , there would be finite subsets P, S of A^* such that $\underline{C} - 1 = \underline{P}(\underline{A} - 1)\underline{S}$, with \underline{X} denoting the characteristic polynomial

Keywords and phrases. Variable length codes, formal languages, factorizations of cyclic groups.

* Partially supported by MIUR Project “Linguaggi Formali e Automi: Metodi, Modelli e Applicazioni” (2003) and by 60% Project “Linguaggi Formali e Codici: Problemi classici e Modelli innovativi” (University of Salerno, 2003).

¹ Dipartimento di Informatica e Applicazioni, Università di Salerno, 84081 Baronissi (SA), Italy; defelice@dia.unisa.it

© EDP Sciences 2005

of X [3, 5, 10]. A finite *maximal code* is a maximal object in the class of finite codes for the order of set inclusion. Any code C which satisfies the above equality is finite, maximal and is called a *factorizing code*. For example, finite biprefix maximal codes are factorizing [3]. The factorization conjecture is one of the most difficult problems which is still open in the theory of codes and the partial results which are known about this conjecture are all reported in Section 2.2. The most impressive result is given by Reutenauer which proves that the above-mentioned equality holds for a finite maximal code C with P, S polynomials with integer coefficients [37, 38].

Another early Schützenberger conjecture questioned whether each finite maximal code could be obtained by means of a simple operation, called *composition* of codes, starting with prefix or suffix codes. This conjecture was proved to be false by Césari in [13].

Inspired by the above-mentioned problems, we investigate the existence of a finite set \mathcal{O} of operations such that each factorizing code can be obtained by using operations in \mathcal{O} and starting with prefix or suffix codes. Here, \mathcal{O} is named a *complete set* of operations (for factorizing codes). Obviously, the class of factorizing codes must be closed under each operation τ in \mathcal{O} (*i.e.*, the result of the application of τ to a pair of factorizing codes must again be a factorizing code). The class of factorizing codes was showed to be closed under composition and under *substitution*, another operation which was initially considered for finite prefix maximal codes in [3] and subsequently defined for factorizing codes in [2]. This operation is based on the idea which frequently recurs in the literature on codes, of changing a word w with a set of words constructed starting from w [6, 12, 30, 41]. Precisely, given factorizing codes $C' = P'(A - 1)S + 1$, $C'' = P''(A - 1)S + 1$ and $w \in C'$, $C = (P' + wP'')(A - 1)S + 1$ is again a factorizing code which is called a *substitution* of C' and C'' by means of w [1, 2]. We can also consider a dual version of this operation working on $C' = P(A - 1)S' + 1$, $C'' = P(A - 1)S'' + 1$ and $w \in C'$. By analogy with the operation of composition, the result of a finite number of applications of substitution or of its dual version can once again be called “substitution”.

The question whether composition and substitution are a complete set of operations for factorizing codes was asked in [17, 18]. In spite of results proved in the same papers, here we show how composition and substitution are not powerful enough to generate all factorizing codes by giving an example of a factorizing code which cannot be obtained by means of these two operations starting with simpler codes (Props. 3.5 and 3.6). The above-mentioned example is a 3-code, an m -code being a finite maximal code such that each word in it has at most m occurrences of b 's, $m \in \mathbf{N}$. Nevertheless, we strongly conjecture that factorizing codes C such that $a \in C$ can be recursively constructed by using only composition and substitution and starting with prefix/suffix codes. In order to find a complete set, we end this paper with examples which can be used to rule out some new operations as possible candidates.

This paper is organized as follows. Section 2 contains all the basic definitions and known results on codes and factorizations of cyclic groups. Precisely, after

the basics on codes and polynomials given in Section 2.1, we report the factorization conjecture and its partial known results in Section 2.2. Factorizations of cyclic groups are considered in Sections 2.3, 2.4 and the two operations of composition and substitution, including all the necessary known results, are defined in Section 2.5. Section 3.1 collects a technical lemma and known results regarding 3-codes which are subsequently used in Section 3.2 to exhibit a factorizing code which cannot be constructed by means of substitution and composition and starting with prefix/suffix codes (Props. 3.5 and 3.6). In Section 4 we gathered a collection of examples showing how some natural attempts of defining new operations are unsuccessful.

2. BASICS

Basics on codes and polynomials are gathered in Section 2.1 whereas Section 2.2 is devoted to the factorization conjecture. Furthermore, since the construction of the factorizing codes is related to special factorizations of cyclic groups and to the solutions of some equations, we will also report all the basic notions on factorizations which we will need later on (Sects. 2.3 and 2.4). We end this section with a discussion on the two special operations on codes mentioned in Section 1 (Sect. 2.5).

2.1. CODES AND POLYNOMIALS

Let A^* be the *free monoid* generated by a finite alphabet A and let $A^+ = A^* \setminus 1$ where 1 is the empty word. For a word $w \in A^*$ and a letter $a \in A$, we denote by $|w|$ the *length* of w and by $|w|_a$ the number of the occurrences of a 's in w . When $|w|_a = r$, we will say that w has r a 's. The same notation $|X|$, when referred to $X \subseteq A^*$, means the cardinality of X . The *reversal* of a word $w = a_1 \dots a_n$, $a_i \in A$, is the word $w^\sim = a_n \dots a_1$ and we set $X^\sim = \{w^\sim \mid w \in X\}$. A word $x \in A^*$ is a *factor* of $w \in A^*$ if $u_1, u_2 \in A^*$ exist such that $w = u_1 x u_2$ and x is a *proper factor* of w if $u_1 u_2 \neq 1$. Furthermore, x is a *prefix* (resp. *suffix*) of $w \in A^*$ if $u_1 = 1$ (resp. $u_2 = 1$); x is a *proper prefix* (resp. *proper suffix*) of $w \in A^*$ if $u_2 \neq 1 = u_1$ (resp. $u_1 \neq 1 = u_2$).

A *code* C is a subset of A^* such that, for all $c_1, \dots, c_h, c'_1, \dots, c'_k \in C$, we have:

$$c_1 \cdots c_h = c'_1 \cdots c'_k \Rightarrow h = k; \quad \forall i \in \{1, \dots, h\} \quad c_i = c'_i.$$

A set $C \subseteq A^+$ such that $C \cap CA^+ = \emptyset$ is a *prefix code*. C is a *suffix code* if C^\sim is a prefix code and C is a *biprefix code* when C is both a suffix and a prefix code. A code C is a *maximal code* over A if for each code C' over A such that $C \subseteq C'$ we have $C = C'$. As one of Schützenberger's basic theorems shows, a finite code C is maximal if and only if C is *complete*, that is $C^* \cap A^* w A^* \neq \emptyset$, for all $w \in A^*$ [3]. If C', C are codes with C being maximal and $C' \subseteq C$ then C is called a *completion* of C' . If, in addition, C is finite, then C is a *finite completion* of C' .

Denote $\mathbf{Z}\langle A \rangle$ (respectively $\mathbf{N}\langle A \rangle$) the semiring of the *polynomials* with non-commutative variables in A and integer (respectively nonnegative integer) coefficients. A finite subset X of A^* will be identified with its *characteristic polynomial*: $\underline{X} = \sum_{x \in X} x$. Henceforth we will use a capital letter to refer to a set and to its characteristic polynomial. For a polynomial P and a word $w \in A^*$, (P, w) denotes the coefficient of w in P and we set $\text{supp}(P) = \{w \in A^* \mid (P, w) \neq 0\}$. When we write $P \geq 0$, we mean $P \in \mathbf{N}\langle A \rangle$.

2.2. FACTORIZING CODES AND THE FACTORIZATION CONJECTURE

Conjecture 2.1, given in a weaker form in [32], is among the most difficult and still open problems in the theory of codes. This conjecture was formulated by Schützenberger and, as far as we know, it does not appear explicitly in any of his papers. It was been quoted as the *factorization conjecture* in [31] for the first time and then also reported in [3, 5, 10].

Conjecture 2.1 (Schützenberger). *Given a finite maximal code C , there are finite subsets P, S of A^* such that:*

$$C - 1 = P(A - 1)S. \quad (1)$$

Each pair (P, S) , with P, S finite subsets of A^* and such that equation (1) holds, will be called a *factorizing pair* for C , and their number, for a given code C , is also referred to as the number of *factorizations* of C . Each code C verifying the previous conjecture is finite, maximal and is called a *factorizing code*. Finite maximal prefix codes are the simplest examples of factorizing codes. Indeed, C is a finite maximal prefix code if and only if $C = P(A - 1) + 1$ for a finite subset P of A^* [3]. In the previous relation, P is the set of the proper prefixes of the words in C . More interesting constructions of factorizing codes can be found in [7] and the result which is closest to a solution of the conjecture, partially reported in Theorem 2.1, was obtained by Reutenauer [5, 37, 38]. He proved that if we allow that $P, S \in \mathbf{Z}\langle A \rangle$, then (1) holds for each finite maximal code C .

Theorem 2.1 [38]. *Let C be such that $C \in \mathbf{N}\langle A \rangle$, $(C, 1) = 0$ and let P, S be such that $P, S \in \mathbf{Z}\langle A \rangle$, $C = P(A - 1)S + 1$. Then, C is a finite maximal code. Furthermore, if $P, S \in \mathbf{N}\langle A \rangle$, then P, S have coefficients 0, 1. Conversely, let C be a finite maximal code. Then, there exist $P, S \in \mathbf{Z}\langle A \rangle$ such that $C = P(A - 1)S + 1$.*

Other results concern the proof of the conjecture for special classes of finite maximal codes. More precisely, given $m \in \mathbf{N}$, an m -code C is a finite maximal code over $\{a, b\}$ such that each word in C has at most m occurrences of the letter b . If m is less than or equal to three then C is factorizing [15, 21, 33]. Furthermore, C is also factorizing if $b^m \in C$ and m is a prime number or $m = 4$ [42].

If C', C are codes with C being factorizing and $C' \subseteq C$ then C is called a *factorizing completion* of C' .

2.3. FACTORIZATIONS OF CYCLIC GROUPS

A relationship exists between factorizing codes and factorizations of cyclic groups that has not yet been thoroughly investigated. We discussed factorizing codes in Section 2.2 and we now consider factorizations of cyclic groups. One of the main references for this argument is [23].

As usual, we realize the cyclic group of order n , $n \geq 2$, as the factor group \mathbf{Z}_n of the integers modulo n . Furthermore, for all positive integers h, n the notation $h|n$ means that h is a divisor of n , whereas, for $H \subseteq \mathbf{N}$, we denote $\min H$ (resp. $\max H$) the smallest (resp. greatest) element in H .

Definition 2.1. A pair (R, T) of subsets of \mathbf{N} is a factorization of \mathbf{Z}_n if $R \oplus T = \mathbf{Z}_n$, that is, for each $z \in \{0, \dots, n-1\}$ there exists a unique pair (r, t) , with $r \in R$ and $t \in T$, such that $r + t = z \pmod{n}$.

We explicitly observe that the classical hypotheses $R, T \subseteq \{0, \dots, n-1\}$ and $0 \in R \cap T$ are not taken into account in this paper, as we are considering factorizations of \mathbf{Z}_n in relation to codes.

The particular class of factorizations involved in this paper was constructed by Hajós. In [24], Hajós gave a method for the construction of a class of factorizations of an abelian group $(G, +)$. This method was slightly corrected later by Sands in [39] and we call these factorizations *Hajós factorizations*. In the definition given by Hajós the operation \circ is also introduced: for subsets $S = \{s_1, \dots, s_q\}$, T of \mathbf{Z}_n , $S \circ T$ denotes the family of subsets of \mathbf{Z}_n having the form $\{s_i + t_i \mid i \in \{1, \dots, q\}\}$, where $T' = \{t_1, \dots, t_q\}$ is any multiset of elements of T having the same cardinality as S .

We often translate the definitions in a polynomial form. Therefore, for a polynomial in $\mathbf{N}\langle a \rangle$, the notation $a^H = \sum_{n \in \mathbf{N}} (H, n) a^n$ will be used with $H \in \mathbf{N}\langle 1 \rangle$, *i.e.*, with H being a finite multiset of nonnegative integers. Computation rules are also defined: $a^{M+L} = a^M a^L$, $a^{M \cup L} = a^M + a^L$, $a^{M \circ L} = a^M \circ a^L$, $a^0 = 0$, $a^1 = 1$. Thus, if $H_1, H_2, \dots, H_k \in \mathbf{N}\langle 1 \rangle$, the expression $a^{H_1} b a^{H_2} \dots a^{H_k}$ is a notation for the product of the formal power series $a^{H_1}, b, a^{H_2}, \dots, a^{H_k}$. For instance, $a^{\{2,3\}} b a^{\{1,5\}} = a^2 b a + a^2 b a^5 + a^3 b a + a^3 b a^5$.

As in [16], we have reported below the original definition given by Hajós in the case $G = \mathbf{Z}_n$.

Definition 2.2. Let R, T be subsets of \mathbf{N} . (R, T) is a Hajós factorization of \mathbf{Z}_n if and only if there exists a chain of divisors of n :

$$k_0 = 1 \mid k_1 \mid k_2 \mid \dots \mid k_s = n, \quad (2)$$

such that:

$$a^R \in \left(\left(\left(\frac{a-1}{a-1} \circ \frac{a^{k_1}-1}{a-1} \right) \cdot \frac{a^{k_2}-1}{a^{k_1}-1} \right) \circ \dots \circ \frac{a^n-1}{a^{k_{s-1}}-1} \right), \quad (3)$$

$$a^T \in \left(\left(\left(\frac{a-1}{a-1} \cdot \frac{a^{k_1}-1}{a-1} \right) \circ \frac{a^{k_2}-1}{a^{k_1}-1} \right) \cdot \dots \circ \frac{a^n-1}{a^{k_{s-1}}-1} \right). \quad (4)$$

Furthermore, we have $R, T \subseteq \{0, \dots, n-1\}$.

The simplest examples of Hajós factorizations are given by Krasner factorizations (I, J) , which were discovered in a completely different context and defined in [25] as follows. Let us consider again the chain $k_0 = 1 \mid k_1 \mid k_2 \mid \dots \mid k_s = n$ of divisors of n given in equation (2) and the subsets I, J of \mathbf{N} defined by equations (5) below:

$$a^I = \prod_{j \text{ even}, 1 \leq j \leq s} \frac{(a^{k_j}-1)}{(a^{k_{j-1}}-1)}, \quad a^J = \prod_{j \text{ odd}, 1 \leq j \leq s} \frac{(a^{k_j}-1)}{(a^{k_{j-1}}-1)}. \quad (5)$$

In [25], Krasner and Ranulac proved that a pair (I, J) of subsets of \mathbf{N} satisfies equations (5) if and only if for any $z \in \{0, \dots, n-1\}$ there exists a unique (i, j) , with $i \in I, j \in J$ and $i + j = z$, i.e., $a^I a^J = (a^n - 1)/(a - 1)$. (I, J) is called a *Krasner factorization*.

2.4. RECURSIVE CONSTRUCTIONS OF HAJÓS FACTORIZATIONS

There are at least two recursive constructions of the Hajós factorizations, depending on whether we look at the first term $k_1 \neq 1$ or at the last term $k_s = n$ in the chain of divisors in equation (2). The first was obtained thanks to a characterization of the Hajós factorizations given in Theorem 2.2 which makes some equations between polynomials in $\mathbf{N}\langle a \rangle$ intervene, the second is reported in Proposition 2.1.

Theorem 2.2 [16]. *Let (R, T) be subsets of $\{0, \dots, n-1\}$. The following conditions are equivalent:*

- 1) (R, T) is a Hajós factorization of \mathbf{Z}_n .
- 2) There exists a Krasner factorization (I, J) of \mathbf{Z}_n such that $(I, T), (R, J)$ are (Hajós) factorizations of \mathbf{Z}_n .
- 3) There exist $L, M \subseteq \mathbf{N}$ and a Krasner factorization (I, J) of \mathbf{Z}_n such that:

$$a^R = a^I(1 + a^M(a-1)), \quad a^T = a^J(1 + a^L(a-1)). \quad (6)$$

Furthermore, 2) \Leftrightarrow 3) also holds for $R, T \subseteq \mathbf{N}$.

Two observations related to this result are worthy of mention. First, Theorem 2.2 points out that for each Hajós factorization (R, T) , we can associate a

Krasner factorization (I, J) with (R, T) , called a *Krasner companion factorization* of (R, T) in [27]. Given a Hajós factorization (R, T) of \mathbf{Z}_n , defined starting with the chain of divisors of n in equation (2), a Krasner companion factorization (I, J) is naturally associated with (R, T) : in order to get (I, J) we have to erase from equation (3) polynomials $P_j = (a^{k_j} - 1)/(a^{k_{j-1}} - 1)$ with j odd, and from equation (4) polynomials P_j with j even [16]. (I, J) will be called the Krasner companion factorization of (R, T) *with respect to the chain of divisors of n* given in equation (2). It is easy to prove that all the Krasner companion factorizations of a given Hajós factorization (R, T) are exactly the Krasner companion factorizations of (R, T) obtained in this way (*i.e.*, (I, J) is a Krasner companion factorization of (R, T) if and only if there exists a chain \mathcal{C} of divisors of n defining (R, T) such that (I, J) is the Krasner companion factorizations of (R, T) with respect to \mathcal{C}) [19].

Secondly, looking at Definition 2.2, we see that for a Hajós factorization (R, T) of \mathbf{Z}_n , we have $R, T \subseteq \{0, \dots, n-1\}$. On the other hand, the equivalence between conditions 2) and 3) in Theorem 2.2 has been stated under the more general hypothesis that R, T are arbitrary subsets of \mathbf{N} (not necessarily with $\max R < n$, $\max T < n$). In order to maintain this general framework in the next part of this paper, for $R, T \subseteq \mathbf{N}$, we will say that (R, T) is a Hajós factorization of \mathbf{Z}_n if $(R_{(n)}, T_{(n)})$ satisfies the conditions contained in Definition 2.2 where, for a subset X of \mathbf{N} and $n \in \mathbf{N}$, we denote $X_{(n)} = \{x' \mid 0 \leq x' \leq n-1, \exists x \in X, x = x' \pmod{n}\}$. This is equivalent, as Lemma 2.1 shows, to defining Hajós factorizations of \mathbf{Z}_n as those pairs satisfying equations (6).

Lemma 2.1 [18]. *Let (I, J) be a Krasner factorization of \mathbf{Z}_n . Let R, R', M be subsets of \mathbf{N} such that $a^R = a^I(1 + a^M(a-1))$ and $a^{R'} = a^{R_{(n)}}$. Then, $M' \subseteq \mathbf{N}$ exists such that $a^{R'} = a^I(1 + a^{M'}(a-1))$ and $I + \max M' + 1 \subseteq \{0, \dots, n-1\}$. Furthermore, if we set $R' = \{r_1, \dots, r_q\}$, $R = \{r_1 + \lambda_1 n, \dots, r_q + \lambda_q n\}$, for $\lambda_1, \dots, \lambda_q \geq 0$, and if we set $a^H = a^{r_1 + \{0, n, \dots, (\lambda_1 - 1)n\}} + \dots + a^{r_q + \{0, n, \dots, (\lambda_q - 1)n\}}$ then we have a disjoint union $M = M' \cup M''$ with $M'' \subseteq \mathbf{N}$, $a^{M''} = a^J a^H$ and $a^R = a^{R'} + a^I(a-1)a^{M''}$.*

The recursive construction of the solutions of equations (6), given in [14] and partially illustrated in Proposition 3.2, allowed us to obtain the recursive construction of the Hajós factorizations given in [16]. Furthermore, equations (6) are also related to the structure of factorizing codes, as Proposition 3.1 shows.

Observing the definition of the Hajós factorizations we can obtain another recursive construction of them with ease. Proposition 2.1, given in [27] as a direct result and proved here for the sake of completeness, illustrates this recursive construction.

Proposition 2.1 [27]. *Let $R, T \subseteq \{0, \dots, n-1\}$ and suppose that (R, T) is a Hajós factorization of \mathbf{Z}_n with respect to the chain $k_0 = 1 \mid k_1 \mid k_2 \mid \dots \mid k_s = n$ of divisors of n . Then either $(R, T) = (R_1, T_1)$ or $(R, T) = (T_1, R_1)$, where (R_1, T_1) satisfies one of the two conditions that follow.*

- 1) *There exists $t \in \{0, \dots, n-1\}$ such that $R_1 = \{0, \dots, n-1\}$ and $T_1 = \{t\}$. Furthermore, $s = 1$.*

- 2) $R_1 = R^{(1)} + \{0, 1, \dots, g-1\}h$, $T_1 = T^{(1)} \circ \{0, 1, \dots, g-1\}h$, $(R^{(1)}, T^{(1)})$ being a Hajós factorization of \mathbf{Z}_h , $g, h \in \mathbf{N}$, $n = gh$, $R^{(1)}, T^{(1)} \subseteq \{0, \dots, h-1\}$. The chain of divisors defining $(R^{(1)}, T^{(1)})$ is $k_0 = 1 \mid k_1 \mid k_2 \mid \dots \mid k_{s-1} = h$.

Proof. Let $R, T \subseteq \{0, \dots, n-1\}$ and suppose that (R, T) is a Hajós factorization of \mathbf{Z}_n with respect to the chain $k_0 = 1 \mid k_1 \mid k_2 \mid \dots \mid k_s = n$ of divisors of n . Thus, looking at Definition 2.2, we see that if $s = 1$ then obviously either $(R, T) = (R_1, T_1)$ or $(R, T) = (T_1, R_1)$, where (R_1, T_1) satisfies condition 1) in the statement. Otherwise, (R, T) is given by equations (3), (4) with $s > 1$. Now, if we eliminate the last term in the expressions given in equations (3) and (4), we get a Hajós factorization $(R^{(1)}, T^{(1)})$ of $\mathbf{Z}_{k_{s-1}}$ with respect to the chain $k_0 = 1 \mid k_1 \mid k_2 \mid \dots \mid k_{s-1}$ of divisors of $k_{s-1} = h$. Furthermore, from Definition 2.2, we have that $R^{(1)}, T^{(1)} \subseteq \{0, \dots, h-1\}$. Thus, it is easy to see that either $(R, T) = (R_1, T_1)$ or $(R, T) = (T_1, R_1)$, where (R_1, T_1) satisfies condition 2) in the statement. \square

2.5. COMPOSITION AND SUBSTITUTION

As we have already said, a central problem in the theory of codes is the description of their structure. We can do so by using operations which allow us to construct codes starting with “simpler” ones, thus reducing the above-mentioned problem to the description of the structure of the “simplest” elements in the family. Two of these operations will be discussed in this section: *composition* and *substitution*.

Let $C' \subseteq B^*$, $D \subseteq A^*$ be two codes such that each letter $b \in B$ is a factor of at least one word in C' and with $|B| = |D|$. Let $\varphi : B^* \rightarrow D^*$ be an isomorphism that extends a bijection from B onto D . Then, the set $C = \varphi(C') = C' \diamond D$ is a code over A and we say that C is obtained by *composition* of C' and D [3]. Furthermore, if C is a maximal code over A then C', D are also maximal codes (respectively over B and A) [3]. A code $C \subseteq A^+$ is *indecomposable* if, whenever $C = C' \diamond D$ then either $D = A$ or $C' = B$, otherwise C *decomposes* or is *decomposable* (over D). We have that a code $C \subseteq A^*$ decomposes over a code D if and only if $C \subseteq D^*$ and each word in D is a factor of at least one word in C [3].

Composition is an interesting operation which can lighten the structure of codes. For instance, in [35], the authors proved that each two-word code has a finite completion since this code can be obtained by composition of prefix/suffix codes, an *n-word code* being a code with exactly n elements. The same result cannot be extended to three-word codes, as showed in [22], and we still do not know whether each three-word code has a finite completion. More specialized statements have been obtained by taking into account the close notion of a *maximal free submonoid*. Indeed, it is known that, if C is a maximal code, then C is indecomposable if and only if C^* is a maximal submonoid of A^* (i.e., $C^* \subseteq D^* \subseteq A^*$, with D being a code, implies either $C = D$ or $D = A$). This partial order between free submonoids can be related to division of finite trees and primeness properties when C is a maximal prefix code [11, 36]. Following this approach, in [36] the authors proved that we

can decide whether a free submonoid C^* is maximal when C is a rational maximal prefix code. Furthermore, a polynomial time algorithm has been given for deciding the maximality of C^* when C is a finite maximal prefix code in [11] and when C is a finite maximal code in [28]. The latter algorithm is a consequence of results given in the same paper [28] on locally complete sets. Informally, the notion of a locally complete set arises in connection with the attempt to define maximality and completeness of a code with respect to a set which has more constraints than the free monoid (as also done in [20,34] with different viewpoints). Regarding the problem examined in this paper, it is known that the class of factorizing codes is closed under composition and specific relations exist between factorizing pairs for C' and D and a factorizing pair for C [7].

The same class of factorizing codes is also closed under another operation which was first considered for finite prefix maximal codes in [3] and subsequently defined for factorizing codes in [1,2]. Let us briefly recall this with some additional remarks and results given in [17,18] which will also be used later on.

In [1,2], the author proved that, given factorizing codes $C' = P'(A-1)S+1$, $C'' = P''(A-1)S+1$ and $w \in C'$, $C = (P' + wP'')(A-1)S+1$ is again a factorizing code which is called a *substitution* of C' and C'' by means of w . C is the characteristic polynomial of $(C' \setminus w) \cup wC''$. Furthermore, the result C^\sim of the obvious dual operation working on C'^\sim, C''^\sim and w^\sim will be once again a factorizing code since the class of factorizing codes is closed under the operation \sim . More generally, the result of a finite number of applications of such an operation, or of the dual version of it will once again be called “substitution”. No specific relationships apparently exist between composition and substitution. However, notice that if C' is a factorizing code, then the substitution C of C' and $C'' = C'$ by means of $w \in C'$ also decomposes over C' .

As we already said in Section 1, an early Schützenberger question asked whether each finite maximal code could be obtained by means of composition of codes, starting from prefix or suffix codes. The first negative answer to this question was given by Césari in [13] and subsequently simpler codes were constructed by Boë and Vincent in [6] and [41] respectively.

A natural question which arises is to ask whether each factorizing code (or finite maximal code) can be obtained by substitution of prefix and suffix codes. This question was explicitly posed in [17,18] with results that led us to believe that we would obtain a positive answer.

In [17], for instance, the author showed that for all $w \in A^*$ and for all factorizing codes $C = P(A-1)(1+w)+1$, C can be obtained by substitution starting from prefix or suffix codes and, as a consequence stated in the same paper, all the counterexamples to the above-mentioned first Schützenberger conjecture can also be obtained in this way. It has also been proved that each finite maximal code C such that $C = P(a+b-1)S+1$ with $P, S \in \mathbf{Z}\langle\{a, b\}\rangle$ and $P \in \mathbf{Z}\langle a \rangle$ or $S \in \mathbf{Z}\langle a \rangle$ can be obtained by substitution of prefix and suffix codes [17,18]. As a consequence, all 1- and 2-codes can be obtained by substitution of prefix and suffix codes. Finally, another result in this direction is recalled in Proposition 2.2. Despite these

encouraging results, a negative answer to the above-mentioned question will be given in Section 3.2.

Remark 2.1. In order to make the statement of Proposition 2.2 clear, we observe that for each factorizing code C such that $C \cap a^* = a^n$, if (P, S) is a factorizing pair for C , then the pair (I, J) defined by $P \cap a^* = a^I$, $S \cap a^* = a^J$ is a Krasner factorization of \mathbf{Z}_n . As a result, (I, J) satisfies the conditions reported in the statement of Proposition 2.1.

Proposition 2.2 [18]. *Let C be a factorizing code with $a^n \in C$, $n > 1$. Suppose that for each $a^i w a^j \in C$, with $w \in (A \setminus a)A^*(A \setminus a) \cup (A \setminus a)$, we have $i, j < n$. Then, for each factorizing pair (P, S) for $C = P(A-1)S + 1$, there exist h with $h \in \mathbf{N}$, $h|n$, $h < n$ and there exist factorizing codes $C^{(k)}$ with $a^h \in C^{(k)}$, $k \in \{0, \dots, (n/h) - 1\}$, such that we have:*

- either $C^{(k)} = P^{(k)}(A-1)S + 1$, $P = \sum_{k=0}^{(n/h)-1} a^{kh} P^{(k)}$ and
 $C = \sum_{k=0}^{(n/h)-1} a^{kh} (C^{(k)} - 1) + 1$;
- or $C^{(k)} = P(A-1)S^{(k)} + 1$, $S = \sum_{k=0}^{(n/h)-1} S^{(k)} a^{kh}$ and
 $C = \sum_{k=0}^{(n/h)-1} (C^{(k)} - 1) a^{kh} + 1$;

the first or the second condition being satisfied depending on whether $P \cap a^ = a^{I_1 + \{0, \dots, (n/h) - 1\}h}$ or $S \cap a^* = a^{J_1 + \{0, \dots, (n/h) - 1\}h}$.*

Remark 2.2 [18]. Let n, h be positive integers with $h|n$, $n/h \geq 2$, let $C^{(k)} = P^{(k)}(A-1)S + 1$ (resp. $C^{(k)} = P(A-1)S^{(k)} + 1$) be a factorizing code with $P^{(k)}, S \in \mathbf{N}\langle A \rangle$ (resp. $P, S^{(k)} \in \mathbf{N}\langle A \rangle$) and $a^h \in C^{(k)}$, for $k \in \{0, \dots, (n/h) - 1\}$. Thus, $C = \sum_{k=0}^{(n/h)-1} a^{kh} (C^{(k)} - 1) + 1$ (resp. $C = \sum_{k=0}^{(n/h)-1} (C^{(k)} - 1) a^{kh} + 1$) is obtained by substitution of codes $C^{(k)}$ and so C is a factorizing code.

Finally, as observed in [17, 18], substitution could also be defined starting with two finite maximal codes $C' = P'(A-1)S + 1$, $C'' = P''(A-1)S + 1$ and $w \in C'$, where $P', P'', S \in \mathbf{Z}\langle A \rangle$, giving as a result the characteristic polynomial $C = P'(A-1)S + wP''(A-1)S + 1 = C' + wC'' - w$ of the finite maximal code $(C' \setminus w) \cup wC''$.

In the next part of this paper we will consider a two-letter alphabet $A = \{a, b\}$. We will also use the following notation: for $P \in \mathbf{Z}\langle A \rangle$ and $g \in \mathbf{N}$, we denote P_g polynomials such that for all $w \in \text{supp}(P_g)$ we have $|w|_b = g$ and $P = P_0 + \dots + P_h$.

3. THE POWER OF COMPOSITION AND SUBSTITUTION

In this section we exhibit a 3-code C which is indecomposable and cannot be obtained by substitution with other codes (Props. 3.5 and 3.6). This result shows that composition and substitution do not suffice to obtain each factorizing code.

Let us briefly outline the contents of this section. Section 3.1 is devoted to results on 3-codes which are subsequently referred to. Most of these results are already known. We begin with the description of the structure of the above-mentioned class of codes. Subsequently, we point out that the investigation of

the behaviour of a 3–code C with respect to the operations of composition and substitution can be made by looking at the structure of the factorizing pairs of C . This is clearly evident for the substitution operation and shown in [8, 9] for the composition operation (see Prop. 3.3). In Section 3.2, we present the above-mentioned example of a 3–code C .

As a preliminary step in the proof of Proposition 3.6, we show that C is a factorizing code having only one factorization (Lem. 3.3). This result allows us to conclude that C has degree 1. We recall that the *degree* $d \geq 1$ of a code C can be defined in terms of *interpretations* of words [3]: the triple (r, w, l) is an interpretation of $u \in A^*$ (with respect to C) if $u = rwl$, with $w \in C^*$, r (respectively l) a proper suffix (respectively prefix) of a word in C . Two interpretations $(r, w, l), (r', w', l')$ of u are *adjacent* if there exist $w_1, w_2, w'_1, w'_2 \in C^*$ such that $w = w_1w_2$, $w' = w'_1w'_2$, $rw_1 = r'w'_1$, $w_2l = w'_2l'$ otherwise $(r, w, l), (r', w', l')$ are *disjoint*. If $\delta_C(u)$ denotes the maximal number of pairwise disjoint interpretations of u , then the *degree* of C is $d = \min \{\delta_C(u) \mid u \in C^* \text{ and } A^*uA^* \cap C = \emptyset\}$. A code C is *synchronous* if C has degree 1, otherwise C is called *asynchronous*.

In this framework, an important result was proved in [38]: let C be a finite maximal code with degree d . Then, there exist $X, Y, Z \in \mathbf{Z}\langle A \rangle$ such that $C - 1 = X(d(A - 1) + (A - 1)Z(A - 1))Y$. Using the same argument as in [4], we prove that this result has the following byproduct: for each $n \in \mathbf{N}$, $a \in A$, such that $a^n \in C$, n is a multiple of d . Indeed, if we substitute the value 0 to all letters $b \neq a$ in the above equation, we get $1 + a + \dots + a^{n-1} = x(a)(d + (a - 1)z(a))y(a)$. Then, when we set $a = 1$ we get $n = dx(1)y(1)$. The following open problem restricts a question posed in [17, 18] to factorizing (synchronous) codes C such that $a \in A \cap C$.

Problem 3.1. *Let C be a factorizing code such that $a \in C$. Can C be obtained by a finite number of applications of composition and substitution, starting with a prefix or suffix code?*

3.1. STRUCTURE AND PROPERTIES OF 3–CODES

Proposition 3.1 describes the structure of 3–codes.

Proposition 3.1 [14, 15]. *Let (P, S) be a factorizing pair for a 3–code $C = P(A - 1)S + 1$. Then, there exists a Krasner factorization (I, J) of \mathbf{Z}_n such that (P, S) satisfies one of the four conditions which follows.*

- a) $P = a^I + \sum_{i \in I} a^i b a^{L_i}$, $S = a^J + \sum_{j \in J'} a^{M_j} b a^j$, with $J', L_i, M_j \subseteq \mathbf{N}$ satisfying the conditions reported below.
 - For all $i \in I$, $a^{T_i} = a^{L_i}(a - 1)a^J + a^J \geq 0$;
 - $J' \subseteq \cup_{i \in I} T_i$;
 - for all $j \in J'$, $a^{M_j}(a - 1)a^I + a^{I_j} = a^{R'_j} \geq 0$, where $I_j = \{i \in I \mid j \in T_i\}$ (thus, we also have $a^{M_j}(a - 1)a^I + a^I = a^{R_j} \geq 0$);
 - for all $j \in J' \setminus J$, for all $i \in I$, $a^{M_j}(a - 1)a^{L_i} + a^{M_j} \geq 0$.
- a') $P = a^I + \sum_{i \in I'} a^i b a^{L_i}$, $S = a^J + \sum_{j \in J} a^{M_j} b a^j$, with $I', L_i, M_j \subseteq \mathbf{N}$ satisfying the conditions reported below.

- for all $j \in J$, $a^{R_j} = a^{M_j}(a-1)a^I + a^I \geq 0$;
 - $I' \subseteq \cup_{j \in J} R_j$;
 - for all $i \in I'$, $a^{L_i}(a-1)a^J + a^{J_i} = a^{T'_i} \geq 0$, where $J_i = \{j \in J \mid i \in R_j\}$ (thus, we also have $a^{L_i}(a-1)a^J + a^J = a^{T_i} \geq 0$);
 - for all $i \in I' \setminus I$, for all $j \in J$, $a^{M_j}(a-1)a^{L_i} + a^{M_j} \geq 0$.
- b) $P = a^I + \sum_{i \in I} a^i b a^{L_i} + \sum_{i \in I, l_i \in L_i} a^i b a^{l_i} b a^{L_i, l_i}$, $S = a^J$, with $L_i, L_{i, l_i} \subseteq \mathbf{N}$, $a^{T_i} = a^{L_i}(a-1)a^J + a^J \geq 0$, $a^{T_{i, l_i}} = a^{L_{i, l_i}}(a-1)a^J + a^J \geq 0$, for all $i \in I$, $l_i \in L_i$.
- b') $P = a^I$, $S = a^J + \sum_{j \in J} a^{M_j} b a^j + \sum_{j \in J, m_j \in M_j} a^{M_{j, m_j}} b a^{m_j} b a^j$, with $M_j, M_{j, m_j} \subseteq \mathbf{N}$, $a^{R_j} = a^{M_j}(a-1)a^I + a^I \geq 0$, $a^{R_{j, m_j}} = a^{M_{j, m_j}}(a-1)a^I + a^I \geq 0$, for all $j \in J$, $m_j \in M_j$.

Remark 3.1. Let C be an n -code with $n \leq 2$. Proposition 3.1 also describes the structure of the factorizing pairs (P, S) of C : (P, S) satisfies condition a) with $J' = \emptyset$ or a') with $I' = \emptyset$, whereas, for 1-codes, we have both $I' = J' = \emptyset$.

Remark 3.2. In [27], the author gave a construction of an infinite family of pairs (R, T) of Hajós factorizations of \mathbf{Z}_n such that, for the corresponding pair (M, L) satisfying $a^R = a^M(a-1)a^I + a^I \geq 0$, $a^T = a^L(a-1)a^J + a^J \geq 0$, we have $a^M(a-1)a^L + a^L < 0$, $a^M(a-1)a^L + a^M < 0$. The Hajós factorization $(R, T) = (\{0, 4, 8, 12, 16, 20\}, \{0, 3, 6, 21\})$ of \mathbf{Z}_{24} is an element of this family and the corresponding pair (M, L) is $(\{1, 9, 11, 13\}, \{2, 3\})$. This example will be taken into account at the time we consider the modulo operation for factorizing codes.

As we have already said in Section 2.4, a recursive construction of all the sets mentioned in Proposition 3.1 can be found in [14]. Below we report a part of this recursive construction since this statement will be used in the proof of Lemma 3.6.

Proposition 3.2 [14]. *Let (I, J) be a Krasner factorization of \mathbf{Z}_n and suppose that k exists with $k \in \mathbf{N}$, $k|n$, $k = \min I \setminus 0$ and $I = kJ_1$. Thus, for each M such that $M \subseteq \mathbf{N}$ and $a^M(a-1)a^I + a^I \geq 0$, there exists $t \geq 0$, $L \subseteq \mathbf{N}$ such that $M = \{0, \dots, t-1\} \cup (t+kL + \{0, \dots, k-1\})$, $\min L > 0$ and $a^L(a-1)a^{J_1} + a^{J_1} \geq 0$.*

Obviously, in the statement of the proposition above, for $t = 0$ we set $\{0, \dots, t-1\} = \emptyset$. As we have already said, n -codes with $n \leq 3$ have been investigated with respect to the degree and the operation of composition in [8, 9]. Decomposable codes in this family have been classified thanks to Proposition 3.3.

Proposition 3.3 [8, 9]. *Let C, C' be two finite maximal codes such that $C' \neq A$. If $C-1 = P(C'-1)S$ with $P, S \subseteq A^*$ and P or $S \neq 1$ then C decomposes over C' . Conversely, if C is an n -code, $1 \leq n \leq 3$, decomposable over the code $C' \subseteq A^+$, then $C-1 = P(C'-1)S$ with $P, S \subseteq A^*$ and P or $S \neq 1$.*

Remark 3.3. If C is a 3-code which is decomposable over the code $C' \subseteq A^+$, then C' is maximal and obviously C' is an n -code, $1 \leq n \leq 3$. So, C' is factorizing.

In [8], the authors stated a characterization of factorizing codes having more than one factorizing pair. Thanks to this result, in [8, 9], they showed that n -codes C with $n \leq 3$ have at most two factorizations and they gave a characterization of

the number of the factorizations of C in terms of the degree of C . In particular, for 3-codes the following result holds.

Proposition 3.4 [9]. *Any 3-code C has one or two factorizations. C has degree 1 if and only if C has one factorization.*

Finally, we prove a result concerning the substitution operation.

Lemma 3.1. *Let C be a 3-code having only one factorization P, S with $P \cap a^*ba^* \neq \emptyset$, $S \cap a^*ba^* \neq \emptyset$. Suppose that C can be obtained by substitution with two factorizing codes C' and C'' by means of $w \in C'$, i.e. $C - 1 = (C' - 1) + (C'' - 1)w$ or $C - 1 = (C' - 1) + w(C'' - 1)$. Thus, we have $w \in a^* \cup a^*ba^*$.*

Proof. Let C be a 3-code which satisfies the hypotheses contained in the statement. Looking at the definition of the substitution operation we know that P or S is a member of a factorizing pair for C'' (and for C'). Then C'' is a p -code with $p \geq 2$. On the other hand, we know that $w C'' \subseteq C$ or $C'' w \subseteq C$. Thus, C being a 3-code, we have $w \in a^* \cup a^*ba^*$. \square

3.2. A COUNTEREXAMPLE

In this section we will show that the expression below:

$$C = \left(a^{\{0,2,4\}} + a^{\{0,2,4\}} ba^{\{0,7,9,11\}} \right) (a + b - 1) \left(a^{\{0,1,6,7\}} + a^{\{0,1,2,3,4,5,6,7,8,9,10,11,12\}} ba^{19} \right) + 1$$

defines a 3-code C which is indecomposable and which cannot be obtained by using the substitution operation (Props. 3.5 and 3.6). Let us briefly outline the proof of this result.

An easy computation in Lemma 3.2 shows that C is a polynomial with nonnegative coefficients and $(C, 1) = 0$. So, C is a 3-code, thanks to Theorem 2.1. Then we prove that C has only one factorization (Lem. 3.3). C is showed to be indecomposable in Prop. 3.5. Therefore, we prove that we cannot find $w \in a^*$ such that C can be obtained by substitution by using w (Lem. 3.4). Finally, we prove that we cannot find $w \in a^*ba^*$ such that C can be obtained by substitution by using w (Lemmata 3.5 and 3.6). Thus, in view of Lemma 3.1, C cannot be obtained by using the substitution operation (Prop. 3.6).

Lemma 3.2. *C is a 3-code.*

Proof. In view of Theorem 2.1, in order to prove that C is a 3-code it suffices to show that C is a polynomial with nonnegative coefficients and $(C, 1) = 0$. This is equivalent to proving that C_i is a polynomial with nonnegative coefficients, for $i \in \{0, 1, 2, 3\}$, and $1 \notin C_0$.

Set $M = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$, it is easy to see that we have:

$$\begin{aligned}
C_0 &= a^{\{0,2,4\}}(a-1)a^{\{0,1,6,7\}} + 1 = a^{12}, \\
C_1 &= a^{\{0,2,4\}}ba^{\{0,1,6,7\}} + a^{\{0,2,4\}}ba^{\{0,7,9,11\}}(a-1)a^{\{0,1,6,7\}} \\
&\quad + a^{\{0,2,4\}}(a-1)a^Mba^{19} \\
&= a^{\{0,2,4\}}ba^{\{1,2,8\}} + a^{\{13,15,17\}}ba^{19}, \\
C_2 &= a^{\{0,2,4\}}ba^{\{0,7,9,11\}}ba^{\{0,1,6,7\}} + a^{\{0,2,4\}}ba^Mba^{19} \\
&\quad + a^{\{0,2,4\}}ba^{\{0,7,9,11\}}(a-1)a^Mba^{19} \\
&= a^{\{0,2,4\}}ba^{\{0,7,9,11\}}ba^{\{0,1,6,7\}} + a^{\{0,2,4\}}ba^{\{0,7,9,11\}}a^{13}ba^{19} \\
&\quad + a^{\{0,2,4\}}ba^{\{1,2,3,4,5,6,8,10,12\}}ba^{19}, \\
C_3 &= a^{\{0,2,4\}}ba^{\{0,7,9,11\}}ba^Mba^{19},
\end{aligned}$$

i.e., C is a polynomial with nonnegative coefficients and $(C, 1) = 0$. \square

Remark 3.4. C is a factorizing completion for $C_1 + a^{12} = a^{\{0,2,4\}}ba^{\{1,2,8\}} + a^{\{13,15,17\}}ba^{19} + a^{12}$ and we will prove that C is indecomposable and cannot be obtained by using the substitution operation (Props. 3.5 and 3.6). Nevertheless $C_1 + a^{12}$ has a factorizing completion obtained by composition and substitution, starting with prefix/suffix codes. Indeed, $C_1 + a^{12}$ decomposes over a suffix code and over $C'_1 + a^{12}$, where $C'_1 = a^{\{0,2,4\}}ba^{\{1,2,8\}} + a^{\{1,3,5\}}ba^7$ (see also [26]). Now, it is sufficient to find a factorizing completion C' (obtained by composition and substitution, starting with prefix/suffix codes) for $C'_1 + a^{12}$: an easy generalization of a result from [35] guarantees that $C_1 + a^{12}$ would have a factorizing completion D which decomposes over C' (and over a maximal suffix code). Now, C' given below is a factorizing completion for $C'_1 + a^{12}$:

$$C' = \left(a^{\{0,2,4\}} + a^{\{0,2,4\}}b \right) (a+b-1) \left(a^{\{0,1,6,7\}} + ba^7 \right) + 1.$$

Furthermore, we can see that C' is a substitution of $C^{(0)} = (a^{\{0,2,4\}} + a^{\{0,2,4\}}b)(a+b-1)a^{\{0,1\}} + 1 = a^{\{0,2,4\}}(1+b)(a+b-1)a^{\{0,1\}} + 1$ and $C^{(1)} = (a^{\{0,2,4\}} + a^{\{0,2,4\}}b)(a+b-1)(a^{\{0,1\}} + ba) + 1 = a^{\{0,2,4\}}(1+b)(a+b-1)(a^{\{0,1\}} + ba) + 1$ by means of a^6 (Prop. 2.2). Finally, $C^{(0)}$ and $C^{(1)}$ can be obtained by composition starting with a suffix code. (Indeed, in view of Prop. 3.3, $C^{(0)}$ (resp. $C^{(1)}$) decomposes over $D_1 = (1+b)(a+b-1)a^{\{0,1\}} + 1$ (resp. $D'_1 = (1+b)(a+b-1)(a^{\{0,1\}} + ba) + 1$) and, in turn, by using once again Prop. 3.3, D_1 (resp. D'_1) decomposes over the suffix code $D_0 = (a+b-1)a^{\{0,1\}} + 1$ (resp. $D'_0 = (a+b-1)(a^{\{0,1\}} + ba) + 1$.)

Lemma 3.3 and Proposition 3.5 give additional information on C .

Lemma 3.3. C has only one factorization.

Proof. By contradiction, suppose that C has two factorizations $C - 1 = P(A - 1)S = P'(A - 1)S'$ with $P = a^{\{0,2,4\}} + a^{\{0,2,4\}}ba^{\{0,7,9,11\}}$, $S = a^{\{0,1,6,7\}} + a^Mba^{19}$, $M = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$, and $P' \neq P$ or $S' \neq S$.

Thus (proof of Lem. 3.2): $C_0 = a^{12}$, $C_1 = a^{\{0,2,4\}}ba^{\{1,2,8\}} + a^{\{13,15,17\}}ba^{19}$, $C_2 = a^{\{0,2,4\}}ba^{\{0,7,9,11\}}ba^{\{0,1,6,7\}} + a^{\{0,2,4\}}ba^{\{0,7,9,11\}}a^{13}ba^{19} + a^{\{0,2,4\}}ba^{\{1,2,3,4,5,6,8,10,12\}}ba^{19}$, $C_3 = a^{\{0,2,4\}}ba^{\{0,7,9,11\}}ba^Mba^{19}$.

Hence, (P', S') satisfies one of the four cases reported in Proposition 3.1. Therefore, if case b) occurs for (P', S') there exists J such that (I, J) is a Krasner factorization of \mathbf{Z}_{12} , $C_1 = a^Iba^J + \sum_{i \in I} a^i ba^{L_i}(a - 1)a^J$ and so $I = \{i \mid a^i ba^j \in C_1\}$, if case $b')$ occurs for (P', S') there exists J such that (I, J) is a Krasner factorization of \mathbf{Z}_{12} , $C_1 = a^Iba^J + \sum_{j \in J} a^{M_j}(a - 1)a^Iba^j$ and so $J = \{j \mid a^i ba^j \in C_1\}$. This is a contradiction since in the first case we have $17 \in I$ and in the second case we have $19 \in J$, so (I, J) cannot be a Krasner factorization of \mathbf{Z}_{12} .

Thus, case a) or case $a')$ occurs for (P', S') . The relation $C_3 = a^{\{0,2,4\}}ba^{\{0,7,9,11\}}ba^Mba^{19}$ yields $P' = a^I + a^{\{0,2,4\}}ba^{\{0,7,9,11\}}$, $S' = a^J + a^Mba^{19}$, (I, J) being a Krasner factorization of \mathbf{Z}_{12} . We have that case $a')$ cannot occur for (P', S') since otherwise, $19 \in J$ contradicts that (I, J) is a Krasner factorization of \mathbf{Z}_{12} . We also have $\{0, 2, 4\} \subseteq I$. Now, looking at the chains of divisors of 12 and at the corresponding Krasner pairs given by equation (5), we observe that under the hypothesis $\{0, 2, 4\} \subseteq I$, only the four cases that follow can occur: $(I, J) = (\{0, \dots, 11\}, \{0\})$, $(I, J) = (\{0, 1, 2, 3, 4, 5\}, \{0, 6\})$, $(I, J) = (\{0, 2, 4, 6, 8, 10\}, \{0, 1\})$, $(I, J) = (\{0, 2, 4\}, \{0, 1, 6, 7\})$. On the other hand, looking at condition a) in Proposition 3.1, we must have $a^{\{0,7,9,11\}}(a - 1)a^J + a^J \geq 0$. A direct computation shows that this relation is not satisfied for $J = \{0\}$, $J = \{0, 6\}$ and $J = \{0, 1\}$. So, $J = \{0, 1, 6, 7\}$, $I = \{0, 2, 4\}$ and consequently $P' = P$, $S' = S$. \square

Proposition 3.5. C is indecomposable.

Proof. By contradiction, suppose that C is decomposable over the code $C' \subseteq A^+$, $C' \neq A$. Then, in view of Proposition 3.3, we have $C - 1 = P(C' - 1)S$ with $P, S \subseteq A^*$ and P or $S \neq 1$. Furthermore, as we have observed in Remark 3.3, C' is an n -code, $1 \leq n \leq 3$ and C' is factorizing. Thus we have $C - 1 = PP'(A - 1)S'S$, where (P', S') is a factorizing pair for C' . In addition, since $C' \neq A$, P' or $S' \neq 1$. Then, thanks to Lemma 3.3, we have:

$$PP' = a^{\{0,2,4\}} + a^{\{0,2,4\}}ba^{\{0,7,9,11\}}, \quad (7)$$

$$S'S = a^{\{0,1,6,7\}} + a^Mba^{19}, \quad (8)$$

where $M = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. As a preliminary observation, we notice that, since $1 \in S'S$, $1 \in PP'$, then $1 \in S' \cap S \cap P \cap P'$. We also know that for m, I, J such that $a^m \in C'$, $a^I = P' \cap a^*$, $a^J = S' \cap a^*$, (I, J) is a Krasner factorization of \mathbf{Z}_m (see Rem. 2.1). Furthermore, since $C \subseteq (C')^*$, we have $m \mid 12$.

Now, one of the cases reported below occurs for S' :

- 1) either $S' \subseteq a^* \cup a^*ba^*$, $S' \not\subseteq a^*$;
- 2) or $S' \subseteq a^*$.

Suppose that case 1) holds. Since $S', S'S \subseteq a^* \cup a^*ba^*$ and $S' \not\subseteq a^*$, we must have $S \subseteq a^*$. Furthermore, if $a^h, a^k \in S$ with $h < k$, looking at equation (8), we have $S'S \cap a^*ba^* = a^Mba^{19}$ and so, for each $a^i ba^j \in S'$, we have $19 = j+h < j+k = 19$, which is a contradiction. Consequently, we have $|S| = 1$ which, together with $1 \in S$, yields $S = 1$.

So, $P \neq 1$ and $S' = a^{\{0,1,6,7\}} + a^Mba^{19}$. As we have already said, for m, I, J such that $a^m \in C'$, $a^I = P' \cap a^*$, $a^J = a^{\{0,1,6,7\}} = S' \cap a^*$, (I, J) is a Krasner factorization of Z_m . Furthermore, as we have already observed, $m|12$ and since $7 \leq m$, we have $m = 12$, *i.e.*, $I = \{0, 2, 4\}$. Therefore, $P' = a^{\{0,2,4\}} + P''$ with $P'' \subseteq a^*ba^*$ and equation (7) can be written as $PP' = a^{\{0,2,4\}} + a^{\{0,2,4\}}ba^{\{0,7,9,11\}} = Pa^{\{0,2,4\}} + PP''$. Now either $P'' \neq 0$, so $P \subseteq a^*$ which implies $P = 1$, a contradiction, or $P'' = 0$ which is also a contradiction since we should have $Pa^{\{0,2,4\}} = a^{\{0,2,4\}} + a^{\{0,2,4\}}ba^{\{0,7,9,11\}}$.

Now, suppose that case 2) holds, *i.e.*, $S' = a^J$. Looking at equation (8), we have $a^J S = a^{\{0,1,6,7\}} + a^Mba^{19}$. Then, $S = a^{J'} + a^{M'}ba^{19}$ with J', M' such that $J + M' = M$ and $J + J' = \{0, 1, 6, 7\}$. We get $|J||M'| = 13$, $|J||J'| = 4$ which yield $|J| = 1$. Thus, since $1 \in S'$, we have $S' = 1$. This means that C' is a prefix code and P' is prefix-closed, *i.e.*, for each z in P' , all the prefixes of z are in P' (see Sect. 2.2). Therefore, $P' \cap a^* = \{a^0, a, \dots, a^k\}$ for a nonnegative integer k . Looking at equation (7), we have that $PP' \cap a^* = \{0, 2, 4\}$ which implies $k \leq 4$. Furthermore, $k = 0$ since, otherwise, $a \in P'$, $1 \in P$ imply $a \in PP'$ which is a contradiction. In conclusion, $P' \cap a^* = 1 = S'$. Thus, since $C' \neq A$, we have $P' = 1 + P''$, $P'' \subseteq a^*ba^*$, $P'' \neq 1$. Once again, since P' is prefix-closed, we have $P'' = ba^{\{0,1,\dots,k\}}$. As a consequence, $PP' = P + PP'' = P + Pba^{\{0,1,\dots,k\}} = a^{\{0,2,4\}} + a^{\{0,2,4\}}ba^{\{0,7,9,11\}}$. The last relation implies $P \subseteq a^*$ and so $P = a^{\{0,2,4\}}$. This is a contradiction since for no k could we have $a^{\{0,2,4\}} + a^{\{0,2,4\}}ba^{\{0,1,\dots,k\}} = a^{\{0,2,4\}} + a^{\{0,2,4\}}ba^{\{0,7,9,11\}}$. \square

In the next three lemmata, we will prove that for the unique factorization P, S of C we cannot write $P = P' + wP''$ with $C' = P'(A-1)S + 1 \geq 0$, $C'' = P''(A-1)S + 1 \geq 0$, $w \in C'$, $w \in a^* \cup a^*ba^*$; nor can we write $S = S' + S''w$ with $C' = P(A-1)S' + 1 \geq 0$, $C'' = P(A-1)S'' + 1 \geq 0$, $w \in C'$, $w \in a^* \cup a^*ba^*$. We will also use the following notations: $I = \{0, 2, 4\}$, $J = \{0, 1, 6, 7\}$, $P' \cap a^* = a^{I'}$, $P'' \cap a^* = a^{I''}$, $S' \cap a^* = a^{J'}$, $S'' \cap a^* = a^{J''}$.

Lemma 3.4. *Let:*

$$\begin{aligned} M &= \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}, \\ P &= a^{\{0,2,4\}} + a^{\{0,2,4\}}ba^{\{0,7,9,11\}}, \\ S &= a^{\{0,1,6,7\}} + a^Mba^{19}. \end{aligned}$$

For no word $w \in a^*$ can we have $P = P' + wP''$ with $C' = P'(A-1)S + 1 \geq 0$, $C'' = P''(A-1)S + 1 \geq 0$, $w \in C'$. Furthermore, for no word $w \in a^*$ can we have $S = S' + S''w$ with $C' = P(A-1)S' + 1 \geq 0$, $C'' = P(A-1)S'' + 1 \geq 0$, $w \in C'$.

Proof. By contradiction, let $w = a^m$ be such that we have either $P = P' + wP''$ with $C' = P'(A-1)S + 1 \geq 0$, $C'' = P''(A-1)S + 1 \geq 0$, $w = a^m \in C'$, or $S = S' + S''w$ with $C' = P(A-1)S' + 1 \geq 0$, $C'' = P(A-1)S'' + 1 \geq 0$, $w = a^m \in C'$. Then, C' and C'' are n -codes with $n \leq 3$. In addition, since $P, S \subseteq a^* \cup a^*ba^*$ and $P, S \not\subseteq a^*$, condition a) or condition a') in Proposition 3.1 describes (P', S) , (P'', S) or (P, S') , (P, S'') (see Prop. 3.1 and Rem. 3.1).

Suppose that $P = P' + wP''$. Then, (I', J) is a Krasner factorization of \mathbf{Z}_m , (I'', J) is also a Krasner factorization and $J = \{0, 1, 6, 7\}$, $a^I = a^{\{0,2,4\}} = a^{I'} + a^m a^{I''}$. This yields a contradiction since on the one hand $|I'| < |I|$, $|I''| < |I|$, on the other hand for each Krasner factorization (K, J) we must have $I \subseteq K$.

Thus, we have $S = S' + S''w$. Analogously, (I, J') is a Krasner factorization of \mathbf{Z}_m , $I = \{0, 2, 4\}$ and $|J'| < |J|$, $J = \{0, 1, 6, 7\}$. Now, a unique subset J' of \mathbf{N} exists such that $(\{0, 2, 4\}, J')$ is a Krasner factorization with $|J'| < |J|$, namely $J' = \{0, 1\}$. On the other hand, we have $a^{\{0,1\}} + a^{\{0,1\}}(a-1)a^{\{0,7,9,11\}} = a + a^2 + a^{13} - a^7$. Then, in virtue of Proposition 3.1 and Remark 3.1, no $S' \subseteq a^* \cup a^*ba^*$ exists such that $S' \cap a^* = a^{\{0,1\}}$ and $P(A-1)S' + 1 \geq 0$. \square

Lemma 3.5. *Let:*

$$\begin{aligned} M &= \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}, \\ P &= a^{\{0,2,4\}} + a^{\{0,2,4\}}ba^{\{0,7,9,11\}}, \\ S &= a^{\{0,1,6,7\}} + a^M ba^{19}. \end{aligned}$$

For no word $w \in a^*ba^*$ can we have $P = P' + wP''$ with $C' = P'(A-1)S + 1 \geq 0$, $C'' = P''(A-1)S + 1 \geq 0$, $w \in C'$.

By contradiction, suppose that $w \in a^*ba^*$ exists such that $P = P' + wP''$ with $C' = P'(A-1)S + 1 \geq 0$, $C'' = P''(A-1)S + 1 \geq 0$, $w \in C'$. Then, C' and C'' are n -codes with $n \leq 3$. In addition, since $S \subseteq a^* \cup a^*ba^*$ and $S \not\subseteq a^*$, condition a) or condition a') in Proposition 3.1 describes (P', S) and (P'', S) (see Prop. 3.1 and Rem. 3.1). We will see that only three cases can occur for P' and P'' and each of these three cases leads to a contradiction.

Since $P \cap a^* = P' \cap a^*$ we must have $I = I'$. On the other hand, $P = P' + wP''$ and $w \in a^*ba^*$ imply $P'' = a^{I''}$ and $wa^{I''} \subseteq a^{\{0,2,4\}}ba^{\{0,7,9,11\}}$. So we must have $|I''| \leq 4$. In addition, since $(I'', \{0, 1, 6, 7\})$ is a Krasner pair, a positive integer k exists such that $k|i$, for all $i \in I''$, $k = \min I'' \setminus \{0\}$, $k|6$. Thus, $k = 2$, $I'' = \{0, 2, 4\} = I' = I$ and $w \in a^{\{0,2,4\}}ba^7$. If $w = ba^7$ we have $wP'' = wa^{I''} = ba^{\{7,9,11\}}$ and, consequently, $P' = P - wP'' = a^{\{0,2,4\}} + b + a^{\{2,4\}}ba^{\{0,7,9,11\}}$. By using a similar argument, we conclude that one of the following three cases holds

with $P'' = a^{I''} = a^{\{0,2,4\}}$ and $w \in a^{\{0,2,4\}}ba^7$:

$$\begin{aligned} P' &= a^{\{0,2,4\}} + b + a^{\{2,4\}}ba^{\{0,7,9,11\}}, & w &= ba^7; \\ P' &= a^{\{0,2,4\}} + a^2b + a^{\{0,4\}}ba^{\{0,7,9,11\}}, & w &= a^2ba^7; \\ P' &= a^{\{0,2,4\}} + a^4b + a^{\{0,2\}}ba^{\{0,7,9,11\}}, & w &= a^4ba^7. \end{aligned}$$

Let us prove that we cannot have $P'(A-1)S+1 \geq 0$ with P' given by one of the relations above and $S = a^{\{0,1,6,7\}} + a^Mba^{19}$. Indeed, since $19 \notin J$, (P', S) satisfies condition a) in Proposition 3.1. Let us adopt the same notations used in this proposition. We must have in each of the three cases above, $a^M(a-1)a^I + a^{I_M} \geq 0$, where $I_M = I_{19} = \{i \in I \mid 19 \in T_i\}$. Now, in the first case we have $L_0 = \{0\}$, $T_0 = \{1, 2, 7, 8\}$ and $L_2 = L_4 = \{0, 7, 9, 11\}$, $T_2 = T_4 = \{1, 2, 8, 19\}$ so, $I_M = \{2, 4\}$; in the second case we have $I_M = \{0, 4\}$ since $T_2 = \{1, 2, 7, 8\}$ and $T_0 = T_4 = \{1, 2, 8, 19\}$, in the third case we have $I_M = \{0, 2\}$ since $T_4 = \{1, 2, 7, 8\}$ and $T_0 = T_2 = \{1, 2, 8, 19\}$. Consequently, in each of the three cases above, I_M is a proper subset of I . This is a contradiction, since we have $a^M(a-1)a^I + a^{I_M} = a^{13}a^I - a^I + a^{I_M} = a^{13} + a^{15} + a^{17} - a^0 - a^2 - a^4 + a^{I_M} \geq 0$ if and only if $I \subseteq I_M$, *i.e.*, $I = I_M$. \square

Lemma 3.6. *Let:*

$$\begin{aligned} M &= \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}, \\ P &= a^{\{0,2,4\}} + a^{\{0,2,4\}}ba^{\{0,7,9,11\}}, \\ S &= a^{\{0,1,6,7\}} + a^Mba^{19}. \end{aligned}$$

For no word $w \in a^*ba^*$ can we have $S = S' + S''w$ with $C' = P(A-1)S' + 1 \geq 0$, $C'' = P(A-1)S'' + 1 \geq 0$, $w \in C'$.

Proof. By contradiction, suppose that $w \in a^*ba^*$ exists such that $S = S' + S''w$ with $C' = P(A-1)S' + 1 \geq 0$, $C'' = P(A-1)S'' + 1 \geq 0$, $w \in C'$. Then, C' and C'' are n -codes with $n \leq 3$. In addition, since $P \subseteq a^* \cup a^*ba^*$ and $P \not\subseteq a^*$, condition a) or condition a') in Proposition 3.1 describes (P, S') and (P, S'') (see Prop. 3.1 and Rem. 3.1). As in the previous lemma we will see that only three cases can occur for S' and S'' and each of these three cases leads to a contradiction.

Since $S \cap a^* = S' \cap a^*$ we must have $J' = J = \{0, 1, 6, 7\}$. On the other hand, $S = S' + S''w$, $w \in a^*ba^*$, $S \subseteq a^* \cup a^*ba^*$ imply $S'' = a^{J''}$ and $S''w \subseteq a^Mba^{19}$ implies $\max J'' \leq \max M = 12$. Furthermore, since $(\{0, 2, 4\}, J'')$ must be a Krasner pair, we have $\{0, 1\} \subseteq J''$. In virtue of Proposition 3.1, no $S'' \subseteq a^*$ exists such that $P(A-1)S'' + 1 \geq 0$, $a^{J''} = S'' = a^{\{0,1\}}$, since, as we have already stated, $a^{\{0,1\}} + a^{\{0,1\}}(a-1)a^{\{0,7,9,11\}} = a + a^2 + a^{13} - a^7$. Thus, $\{0, 1, 6, 7\} \subseteq J''$. So, we must have $J'' = \{0, 1, 6, 7\} = J = J' = S''$ since, otherwise $\{0, 1, 6, 7, 12, 13\} \subseteq J''$, which is a contradiction with $\max J'' \leq 12$.

Furthermore, since $S''w = a^{J''}w = a^{\{0,1,6,7\}}w = a^{\{0,1,6,7\}}a^m ba^j \subseteq S = a^{\{0,1,6,7\}} + a^M ba^{19}$, we must have $w = a^m ba^{19}$ with $m \leq 5$. In addition, we have $S' = S - S''w = a^{\{0,1,6,7\}} + a^{M'} ba^{19}$ with $M' = M \setminus (m + \{0, 1, 6, 7\})$ and $a^{M'}(a-1)a^I + a^I \geq 0$, $I = \{0, 2, 4\}$ (Prop. 3.1). Looking at the characterization of the solutions of this equation given in Proposition 3.2, we see that $M' = M'_1 \cup M'_2$ with M'_1 being a (possibly empty) set of consecutive integers and $|M'_2|$ being an even number ($k = 2$). Since $|m + \{0, 1, 6, 7\}|$ is an even number and $|M|$ is an odd number, we must have that $|M'|$ and so $|M'_1|$ are odd too (and $M'_1 \neq \emptyset$). Consequently $m \in \{1, 3, 5\}$ (we have $0 \in M'_1$ and $m = 2$ or $m = 4$ imply $M'_1 = \{0, 1\}$ or $M'_1 = \{0, 1, 2, 3\}$ respectively, which is impossible since $|M'_1|$ would be an even number).

If $w = aba^{19}$, we get $S' = S - S''w = S - a^{\{0,1,6,7\}}aba^{19} = S - a^{\{1,2,7,8\}}ba^{19}$. Using an analogous argument when $w \in a^{\{3,5\}}ba^{19}$, we can conclude that, with $S'' = a^{\{0,1,6,7\}}$ and $w \in a^{\{1,3,5\}}ba^{19}$, one of the following three cases holds:

$$\begin{aligned} w &= aba^{19}, S' = a^{\{0,1,6,7\}} + a^{\{0,3,4,5,6,9,10,11,12\}}ba^{19} = a^{\{0,1,6,7\}} + a^{M'}ba^{19}; \\ w &= a^5ba^{19}, S' = a^{\{0,1,6,7\}} + a^{\{0,1,2,3,4,7,8,9,10\}}ba^{19} = a^{\{0,1,6,7\}} + a^{M'}ba^{19}; \\ w &= a^3ba^{19}, S' = a^{\{0,1,6,7\}} + a^{\{0,1,2,5,6,7,8,11,12\}}ba^{19} = a^{\{0,1,6,7\}} + a^{M'}ba^{19}. \end{aligned}$$

Let us prove that we cannot have $P(A-1)S' + 1 \geq 0$ with $P = a^{\{0,2,4\}} + a^{\{0,2,4\}}ba^{\{0,7,9,11\}}$ and S' given by one of the relations above. Indeed, since $19 \notin J = J'$, (P, S') satisfies condition *a*) in Proposition 3.1. Hence, thanks to the fourth requirement in this condition and since, as already observed, $19 \notin J'$, in each of the three cases above, we must have $a^{\{0,7,9,11\}}(a-1)a^{M'} + a^{M'} \geq 0$. On the other hand an easy computation shows that this condition is not satisfied, since we have

$$\begin{aligned} &\left(a^{\{0,7,9,11\}}(a-1)a^{\{0,3,4,5,6,9,10,11,12\}} + a^{\{0,3,4,5,6,9,10,11,12\}}, a^9\right) = -1, \\ &\left(a^{\{0,7,9,11\}}(a-1)a^{\{0,1,2,3,4,7,8,9,10\}} + a^{\{0,1,2,3,4,7,8,9,10\}}, a^7\right) = -1, \\ &\left(a^{\{0,7,9,11\}}(a-1)a^{\{0,1,2,5,6,7,8,11,12\}} + a^{\{0,1,2,5,6,7,8,11,12\}}, a^{11}\right) = -1. \end{aligned}$$

□

Proposition 3.6. *The relation:*

$$C = (a^{\{0,2,4\}} + a^{\{0,2,4\}}ba^{\{0,7,9,11\}})(a+b-1)(a^{\{0,1,6,7\}} + a^{\{0,1,2,3,4,5,6,7,8,9,10,11,12\}}ba^{19}) + 1$$

defines a 3-code C which cannot be obtained by substitution (with other codes).

Proof. By Lemmata 3.2, 3.3, C is a 3-code which has only one factorization, namely $P = a^{\{0,2,4\}} + a^{\{0,2,4\}}ba^{\{0,7,9,11\}}$, $S = a^{\{0,1,6,7\}} + a^{\{0,1,2,3,4,5,6,7,8,9,10,11,12\}}ba^{19}$.

Looking at Lemma 3.1, if C could be obtained by substitution then there should exist $w \in a^* \cup a^*ba^*$ such that $P = P' + wP''$ with $C' = P'(A-1)S + 1 \geq 0$, $C'' = P''(A-1)S + 1 \geq 0$, $w \in C'$ or $S = S' + S''w$ with $C' = P(A-1)S' + 1 \geq 0$, $C'' = P(A-1)S'' + 1 \geq 0$, $w \in C'$. By using Lemmata 3.4, 3.5 and 3.6, neither the first case nor the second case can occur. \square

4. FINAL COMMENTS

As stated explicitly in [17, 18], it is clear that we obtain a recursive construction of the factorizing codes if we show the existence of an algorithm \mathcal{P}_a which allows us to construct each factorizing code C with $a^n \in C$ starting with factorizing codes C' with $a^{n'} \in C'$ and $n' < n$. On the other hand, under the hypotheses of Proposition 2.2, C is obtained by substitution of codes $C^{(k)}$ with $a^h \in C^{(k)}$ and $h < n$ (Rem. 2.2). As a consequence, in order to state the existence of \mathcal{P}_a what is missing is a transformation MOD which allows us to go from a factorizing code C with $a^n \in C$ to a (factorizing) code, say $C^{(\text{mod } n)}$, which should satisfy the above-mentioned hypotheses in Proposition 2.2. We already know that this objective cannot be reached by using the composition operation only and in Section 3.2 we saw that it cannot be reached by substitution either. So, a new operation, say MOD , should be defined so that each factorizing code can be obtained by using composition, substitution and the MOD operation, starting with simpler factorizing codes. More generally (and in a more precise way), a natural question which arises is the investigation of the existence of a finite set \mathcal{O} of operations (*complete set of operations*) such that each factorizing code can be obtained by using operations in \mathcal{O} and starting with prefix or suffix codes. We do not know whether \mathcal{O} exists and we will outline how difficult it is to answer this question. We will also point out that finding a procedure that allows us to construct each m -code starting with m' -codes, with $m' < m$ is not easier.

Let us restrict ourselves to 3-codes. Let C be such a code and let (P, S) be a factorizing pair for C . It is already known that if (P, S) satisfies condition $b)$ or $b')$ in Proposition 3.1 then C can be obtained by substitution and starting with prefix/suffix codes [17, 18]. Then, we will always suppose that for (P, S) we have $P = a^I + \sum_{i \in I} a^i ba^{L_i}$, $S = a^J + \sum_{j \in J'} a^{M_j} ba^j$, with $I, J, J', L_i, M_j \subseteq \mathbf{N}$ satisfying condition $a)$ in Proposition 3.1 (the other case, *i.e.*, when (P, S) satisfies condition $a')$ is analogous).

A first natural attempt to define MOD is by looking for a transformation which allows us to go from a factorizing code C with $a^n \in C$ to a (factorizing) code, say $C^{(\text{mod } n)}$ such that $C_1 = C \cap a^*$ is transformed in $C^{(\text{mod } n)} \cap a^* = C_1^{(\text{mod } n)} = \{a^{i'} ba^{j'} \mid i' < n, j' < n, \exists a^i ba^j \in C_1, i' = i \pmod{n}, j' = j \pmod{n}\}$. When n is a prime number, this attempt already fails for 3-codes C . Indeed, let us consider the code:

$$C = \left(a^{\{0,1,2,3,4\}} + a^{\{0,1,2,3,4\}} ba^{\{0,1\}} \right) (A-1) \left(1 + a^{\{0,1,2,3,4\}} ba^2 \right) + 1.$$

Then, it is easy to see that a factorizing code $C^{(\text{mod } 5)}$ such that $C^{(\text{mod } 5)} \cap a^* = C_1^{(\text{mod } 5)} = \{a^{i'}ba^{j'} \mid i' < 5, j' < 5, \exists a^{i'}ba^{j'} \in C_1, i' = i \pmod{5}, j' = j \pmod{5}\}$ must be a 2-code.

However, suppose that we eliminate the particular case of prime numbers. We could go from $C_1 = C \cap a^*$ to $C_1^{(\text{mod } n)}$ by using Lemma 2.1. Indeed, looking at Proposition 3.1, we could consider the relation $C^{(\text{mod } n)} = P^{(\text{mod } n)}(A-1)S^{(\text{mod } n)} + 1$, where $P^{(\text{mod } n)} = a^I + \sum_{i \in I} a^i ba^{L_i}$, $S^{(\text{mod } n)} = a^J + \sum_{j \in J'_n} a^{M'_j} ba^{j(n)}$, M'_j (resp. L'_i) is the subset of M_j (resp. L_i) defined by Lemma 2.1 and, for each $j \in J'$, $j(n) \in J'_n$ is such that $j(n) = j \pmod{n}$. We will illustrate this construction in Example 4.1. Unfortunately this transformation yields a 3-code only in some special cases, since as Remark 4.1 shows, $C^{(\text{mod } n)}$ is a polynomial with not necessarily positive coefficients.

Example 4.1. Let C be the 3-code considered in Section 3.2 and reported below:

$$C = \left(a^{\{0,2,4\}} + a^{\{0,2,4\}}ba^{\{0,7,9,11\}} \right) (a+b-1) \left(a^{\{0,1,6,7\}} + a^{\{0,1,2,3,4,5,6,7,8,9,10,11,12\}}ba^{19} \right) + 1.$$

It is easy to see that we have $C^{(\text{mod } n)} = C'$, where C' is the factorizing code considered in Remark 3.4 and reported below:

$$C' = (a^{\{0,2,4\}} + a^{\{0,2,4\}}b)(a+b-1)(a^{\{0,1,6,7\}} + ba^7) + 1.$$

Indeed, by using the same notations as in Lemma 2.1, for $L_0 = L_2 = L_4 = \{0, 7, 9, 11\}$ we have $L'_0 = L'_2 = L'_4 = \{0\}$. Analogously, for $M_{19} = M = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ we get $M'_{19} = \{0\}$.

Remark 4.1. Note that 3-codes C exist with a factorizing pair (P, S) such that we have $P = a^I + \sum_{i \in I} a^i ba^{L_i}$, $S = a^J + \sum_{j \in J'} a^{M'_j} ba^j$, where $I, J, J', L_i, M_j \subseteq \mathbf{N}$ satisfy condition a) in Proposition 3.1 and with $a^{M'_j}(a-1)a^{L'_i} + a^{M'_j} < 0$ for $j \in J' \setminus J, i \in I$. In this case, the result $C^{(\text{mod } n)}$ of the operation defined above and applied to C , is a polynomial with negative coefficients. For instance, it is easy to see that the relation $D = (a^{\{0,2,4,12,14,16\}} + a^{\{0,2,4,12,14,16\}}ba^{\{1,3,5,7,9,11,13,15,17,19\}})(A-1)(a^{\{0,1,6,7\}} + a^{\{2,3\}}ba^{21}) + 1$ defines a 3-code. Set $M = \{1, 3, 5, 7, 9, 11, 13, 15, 17, 19\}$. Since $a^M = a^{M'} + a^3a^{\{0,2,4,12,14,16\}}$, where $M' = \{1, 9, 11, 13\}$, the result $D^{(\text{mod } n)}$ of the above-mentioned operation applied to D is the polynomial defined by the relation $D^{(\text{mod } n)} = (a^{\{0,2,4,12,14,16\}} + a^{\{0,2,4,12,14,16\}}ba^{\{1,9,11,13\}})(A-1)(a^{\{0,1,6,7\}} + a^{\{2,3\}}ba^{21}) + 1$. Note that the sets $\{1, 9, 11, 13\}, \{2, 3\}$ belong to a family of pairs (L, M) constructed in [27] and quoted in Remark 3.2. As a consequence, $D^{(\text{mod } n)}$ is a polynomial with negative coefficients.

The search for a complete set of operations for factorizing codes can be carried out starting from a symmetric viewpoint to the one presented above. Namely, we

can try to find a procedure that allows us to construct each m -code C starting with m' -codes, with $m' < m$. An algorithm doing so exists when C has a factorizing pair (P, S) with $P \subseteq a^*$ or $S \subseteq a^*$ [14]. On the other hand, the argument below should give evidence that in the general case this is not an easy task either.

Let C be a 4-code, let (P, S) be a factorizing pair for C with $P \not\subseteq a^*$, $S \not\subseteq a^*$. Under particular additional hypotheses on (P, S) , we can prove that (P', S) is a factorizing pair for a 3-code, where $P' = P_0 + P_1 = P \cap (a^* \cup a^*ba^*)$. However, this method does not construct all 4-codes. Examples 4.2 and 4.3 illustrate the first and the second case respectively.

Example 4.2. Let us consider the polynomials:

$$\begin{aligned} P &= a^{\{0,2,4,12,14,16\}} + a^{\{0,2,4,12,14,16\}}ba^{\{1,3,5,7,9,11,13,15,17,19\}} \\ &\quad + a^{\{0,2,4,12,14,16\}}ba^{\{1,3,5,7,9,11,13,15,17,19\}}ba^{\{1,3,5,7,9,11,13,15,17,19\}}, \\ S &= a^{\{0,1,6,7\}} + a^{\{2,3\}}ba^{21}. \end{aligned}$$

An easy computation shows that the relation $C = P(A - 1)S + 1$ defines a 4-code. In addition, the pair (P', S) , with $P' = P \cap (a^* \cup a^*ba^*)$ defines a 3-code (Rem. 4.1).

Example 4.3. Let us consider the polynomials:

$$\begin{aligned} P &= a^{\{0,2,4,12,14,16\}} + a^{\{0,2,4,12,14,16\}}ba^{\{1,9,11,13\}} \\ &\quad + a^{\{0,2,4,12,14,16\}}ba^{\{1,9,11,13\}}ba^{\{1,3,5,7,9,11,13,15,17,19\}}, \\ S &= a^{\{0,1,6,7\}} + a^{\{2,3\}}ba^{21}. \end{aligned}$$

An easy computation shows that the relation $C = P(A - 1)S + 1$ defines a 4-code. However, the pair (P', S) , with $P' = a^{\{0,2,4,12,14,16\}} + a^{\{0,2,4,12,14,16\}}ba^{\{1,9,11,13\}}$ does not define a 3-code (Rem. 4.1).

Acknowledgements. The author wishes to thank J. Berstel and D. Perrin for discussions concerning the origin of the factorization conjecture and for pointing out Ref.[31]. She is also grateful to the anonymous referees and to one of them in particular for his/her constructive criticism. Many thanks to V. Bruyère for her helpful observations.

REFERENCES

- [1] M. Anselmo, A Non-Ambiguous Decomposition of Regular Languages and Factorizing Codes, in *Proc. DLT'99*, G. Rozenberg, W. Thomas Eds. World Scientific (2000) 141–152.
- [2] M. Anselmo, A Non-Ambiguous Decomposition of Regular Languages and Factorizing Codes. *Discrete Appl. Math.* **126** (2003) 129–165.
- [3] J. Berstel and D. Perrin, *Theory of Codes*. Academic Press, New York (1985).
- [4] J. Berstel and D. Perrin, Trends in the Theory of Codes. *Bull. EATCS* **29** (1986) 84–95.
- [5] J. Berstel and C. Reutenauer, Rational Series and Their Languages. *EATCS Monogr. Theoret. Comput. Sci.* **12** (1988).

- [6] J.M. Boë, Une famille remarquable de codes indécomposables, in *Proc. Icalp 78. Lect. Notes Comput. Sci.* **62** (1978) 105–112.
- [7] J.M. Boë, Sur les codes factorisants, in *Théorie des Codes, Actes de la 7^e École de Printemps d'Informatique Théorique*, edited by D. Perrin. LITP and ENSTA, Paris (1980) 1–8.
- [8] V. Bruyère and C. De Felice, Synchronization and decomposability for a family of codes. *Intern. J. Algebra Comput.* **4** (1992) 367–393.
- [9] V. Bruyère and C. De Felice, Synchronization and decomposability for a family of codes: Part 2. *Discrete Math.* **140** (1995) 47–77.
- [10] V. Bruyère and M. Latteux, Variable-Length Maximal Codes, in *Proc. Icalp 96. Lect. Notes Comput. Sci.* **1099** (1996) 24–47.
- [11] M.G. Castelli, D. Guaiana and S. Mantaci, Indecomposable prefix codes and prime trees, in *Proc. DLT 97* edited by S. Bozapadilis-Aristotel (1997).
- [12] Y. Césari, Sur un algorithme donnant les codes biprefixes finis. *Math. Syst. Theory* **6** (1972) 221–225.
- [13] Y. Césari, Sur l'application du théorème de Suschkevitch à l'étude des codes rationnels complets, in *Proc. Icalp 74. Lect. Notes Comput. Sci.* (1974) 342–350.
- [14] C. De Felice, Construction of a family of finite maximal codes. *Theoret. Comput. Sci.* **63** (1989) 157–184.
- [15] C. De Felice, A partial result about the factorization conjecture for finite variable-length codes. *Discrete Math.* **122** (1993) 137–152.
- [16] C. De Felice, An application of Hajós factorizations to variable-length codes. *Theoret. Comput. Sci.* **164** (1996) 223–252.
- [17] C. De Felice, Factorizing Codes and Schützenberger Conjectures, in *Proc. MFCS 2000. Lect. Notes Comput. Sci.* **1893** (2000) 295–303.
- [18] C. De Felice, On some Schützenberger Conjectures. *Inform. Comp.* **168** (2001) 144–155.
- [19] C. De Felice, An enhanced property of factorizing codes. *Theor. Comput. Sci.* **340** (2005) 240–256.
- [20] C. De Felice and A. Restivo, Some results on finite maximal codes. *RAIRO-Inform. Theor. Appl.* **19** (1985) 383–403.
- [21] C. De Felice and C. Reutenauer, Solution partielle de la conjecture de factorisation des codes. *C.R. Acad. Sci. Paris* **302** (1986) 169–170.
- [22] D. Derencourt, A three-word code which is not prefix-suffix composed. *Theor. Comput. Sci.* **163** (1996) 145–160.
- [23] L. Fuchs, *Abelian groups*. Pergamon Press, New York (1960).
- [24] G. Hajós, Sur la factorisation des groupes abéliens. *Casopis Pest. Mat. Fys.* **74** (1950) 157–162.
- [25] M. Krasner and B. Ranulac, Sur une propriété des polynômes de la division du cercle. *C.R. Acad. Sci. Paris* **240** (1937) 397–399.
- [26] N.H. Lam, A note on codes having no finite completions. *Inform. Proc. Lett.* **55** (1995) 185–188.
- [27] N.H. Lam, Hajós factorizations and completion of codes. *Theor. Comput. Sci.* **182** (1997) 245–256.
- [28] J. Neraud and C. Selmi, Locally complete sets and finite decomposable codes. *Theor. Comput. Sci.* **273** (2002) 185–196.
- [29] M. Nivat, Éléments de la théorie générale des codes, in *Automata Theory*, edited by E. Caianiello. Academic Press, New York (1966) 278–294.
- [30] D. Perrin, Codes asynchrones. *Bull. Soc. Math. France* **105** (1977) 385–404.
- [31] D. Perrin, Polynôme d'un code, in *Théorie des Codes, Actes de la 7^e École de Printemps d'Informatique Théorique*, edited by D. Perrin. LITP and ENSTA, Paris (1980) 169–176.
- [32] D. Perrin and M.P. Schützenberger, Un problème élémentaire de la théorie de l'information, *Théorie de l'Information, Colloques Internat. CNRS, Cachan* **276** (1977) 249–260.
- [33] A. Restivo, On codes having no finite completions. *Discrete Math.* **17** (1977) 309–316.
- [34] A. Restivo, Codes and local constraints. *Theor. Comput. Sci.* **72** (1990) 55–64.

- [35] A. Restivo, S. Salemi and T. Sportelli, Completing codes. *RAIRO-Inf. Theor. Appl.* **23** (1989) 135–147.
- [36] A. Restivo and P.V. Silva, On the lattice of prefix codes. *Theor. Comput. Sci.* **289** (2002) 755–782.
- [37] C. Reutenauer, Sulla fattorizzazione dei codici. *Ricerche di Mat.* **XXXII** (1983) 115–130.
- [38] C. Reutenauer, Non commutative factorization of variable-length codes. *J. Pure Appl. Algebra* **36** (1985) 167–186.
- [39] A.D. Sands, On the factorisation of finite abelian groups. *Acta Math. Acad. Sci. Hungaricae* **8** (1957) 65–86.
- [40] M.P. Schützenberger, Une théorie algébrique du codage, *Séminaire Dubreil-Pisot* 1955–56, exposé No. 15 (1955), 24 p.
- [41] M. Vincent, Construction de codes indécomposables. *RAIRO-Inf. Theor. Appl.* **19** (1985) 165–178.
- [42] L. Zhang and C.K. Gu, Two classes of factorizing codes – (p, p) -codes and $(4, 4)$ -codes, in *Words, Languages and Combinatorics II*, edited by M. Ito and H. Jürgensen. World Scientific (1994) 477–483.

Communicated by J. Berstel.

Received April 20, 2004. Accepted February 15, 2005.