Monoids and Maximal Codes

Fabio Burderi
Università Degli Studi
Palermo, Italy
Dipartimento di Matematica ed Applicazioni,
Università Degli Studi di Palermo,
Via Archirafi 34, 90123 Palermo, Italy
burderi@math.unipa.it

In recent years codes that are not Uniquely Decipherable (*UD*) have been studied partitioning them in classes that localize the ambiguities of the code. A natural question is how we can extend the notion of maximality to codes that are not *UD*. In this paper we give an answer to this question.
To do this we introduce a partial order in the set of submonoids of a monoid showing the existence, in this poset, of maximal elements that we call *full* monoids. Then a set of generators of a full monoid is, by definition, a maximal code. We show how this definition extends, in a natural way, the existing definition concerning *UD* codes and we find a characteristic property of a monoid generated by a maximal *UD* code.

1 Introduction

At the beginning, in the context of information theory, the word *code* has denoted what we call here Uniquely Decipherable (*UD*) code, that is a set of words with the property that every concatenation of words of the set (called *message*) has an unique decomposition in code words. This notion, in the next years, has been weakened so we call here code just a set of non-empty words.

A notion weaker than uniquely decipherability has been used in several situations: to investigate natural languages (see [7]) or to study situations in which it is allowed to recover the original message up to a permutation of the code words (see [10], [11], [9]) or even when the only information to recover is the number of code words (see [12]). In other cases the study has been oriented toward sets of words with a constraint source (see [5]). In [8], Guzmán has been introduced the notion of *variety* of codes to study, in a general approach, decipherability conditions weaker than *UD*.

In [4], studying varieties of codes under the aspect of uniform distribution of probability, we noted that the construction, introduced by Ehrenfeucht and Rozemberg in [6], for embedding a regular *UD* code in a complete and regular *UD* code, also works in the ambit of varieties of codes: the new words, introduced by the construction, do not create new relations between code words. Indeed the only relations between the code words are that existing before the construction.
This observation has lead to deepen the study of the relations that arise in a set of non-empty words and so in [3], generalizing a construction used in [4], we introduced the notion of *coding partition*. Roughly speaking a partition of a code is a coding partition if any message has a unique factorization in blocks: a block is the concatenation of words from one class of the partition, and consecutive blocks are composed by words from different classes of the partition. In this case the possible ambiguities of the code are confined in the classes of the partition.

In [2], the very important class of maximal *UD* codes is studied. In the case of thin *UD* codes, is known, for example, the equivalence between maximality and completeness.

P. Ambrož, Š. Holub and Z. Masáková (Eds.):
8th International Conference WORDS 2011
EPTCS 63, 2011, pp. 83–92, doi:10.4204/EPTCS.63.12

In this paper we define the maximality of a code by an algebraic property of the monoid generated by the code itself. We show that this definition of maximality generalizes the existing one concerning *UD* codes. We present, moreover, some classical result on *UD* codes that we can easily re-establish in the general case.

2 Partitions of a code

Let A be an alphabet. We denote by A^* the set of finite words over the alphabet A, and by A^+ the set of non-empty finite words. A^* is a monoid under the concatenation operation of two words, with the empty word as the neutral element. A *code* X is here a subset of A^+. Its elements are called *code words*, the elements of X^* *messages* .

A code X is said to be *uniquely decipherable* (*UD*) if every message has a unique factorization into code words, i.e. the equality

$$x_1x_2\cdots x_n = y_1y_2\cdots y_m,$$

$x_1,x_2,\ldots,x_n,\ y_1,y_2,\ldots,y_m \in X$, implies $n = m$ and $x_1 = y_1,\ldots,x_n = y_n$.

Let X be a code and let

$$P = \{X_i \mid i \in I\}$$

be a partition of X i.e., $\bigcup_{i\in I} X_i = X$ and $X_i \cap X_j = \emptyset$ iff $i \neq j$.

A *P-factorization* of a message $w \in X^+$ is a factorization $w = z_1z_2\cdots z_t$, where

- for each $i,\ z_i \in X_k^+$, for some $k \in I$
- if $t > 1,\ z_i \in X_k^+ \Rightarrow z_{i+1} \notin X_k^+ \quad (1 \leq i \leq t-1)$.

The partition P is called a *coding partition* if any element $w \in X^+$ has a *unique P-factorization*, i.e. if

$$w = z_1z_2\cdots z_s = u_1u_2\cdots u_t,$$

where $z_1z_2\cdots z_s,\ u_1u_2\cdots u_t$ are *P-factorizations* of w, then $s = t$ and $z_i = u_i$ for $i = 1,\ldots,s$.
We observe that the trivial partition $P = \{X\}$ is always a coding partition.

Let $w \in A^+$ be a word. A *factorization* of w is a sequence of words $(v_i)_{1\leq i\leq s}$ such that $w = v_1v_2\cdots v_s$. Let X be a code. A *relation* is a pair of factorizations $x_1x_2\cdots x_s = y_1y_2\cdots y_t$ into code words of a same message $z \in X^+$; the relation is said non-trivial if the factorizations are distinct. In the sequel, when no confusion arises, sometimes we will denote by z both the "word" z and the *relation* $x_1x_2\cdots x_s = y_1y_2\cdots y_t$. We say that the relation $x_1x_2\cdots x_s = y_1y_2\cdots y_t$ is *prime* if for all $i < s$ and for all $j < t$ one has $x_1x_2\cdots x_i \neq y_1y_2\cdots y_j$.

In [3], the following theorem is proved.

Theorem 2.1 *Let $P = \{X_i \mid i \in I\}$ be a partition of a code X. The partition P is a coding partition iff for every prime relation $x_1x_2\cdots x_s = y_1y_2\cdots y_t$, the code words x_i, y_j belong to the same component of the partition.*

Recall that there is a natural partial order between the partitions of a set X: if P_1 and P_2 are two partitions of X then $P_1 \leq P_2$ if the elements of P_1 are unions of elements of P_2 and we say that P_2 is *finer* then P_1. Then from Theorem 2.1 we have the following corollary.

Corollary 2.2 *Let P and P' be two partitions of a code X with $P \leq P'$. If P' is a coding partition then P also.*

What follows, till Theorem 2.7, is stated in [3].

Theorem 2.3 *The set of the coding partitions of a code X is a complete lattice.*

As a consequence of previous theorem we can give the following definition. Given a code X, *the* finest coding partition P of X is called the *characteristic* partition of X and it is denoted by $P(X)$.

A code X is called *ambiguous* if it is not *UD*. It is called *totally ambiguous* (*TA*) if $|X| > 1$ and $P(X)$ is the trivial partition: $P(X) = \{X\}$.

Remark 2.4 *So UD codes and TA codes correspond to the two extremal cases since a code is UD iff $P(X) = \{\{x\} \mid x \in X\}$.*

Let X be a code and let $P(X)$ be the characteristic partition of X. Let X_0 be the union of all classes of $P(X)$ having only one element, i.e. of all classes $Z \in P(X)$ such that $|Z| = 1$. The code X_0 is a *UD* code and is called the *unambiguous component* of X. From $P(X)$ one then derives another partition of X

$$P_C(X) = \{X_i \mid i \geq 0\},$$

where $\{X_i \mid i \geq 1\}$ is the set of classes of $P(X)$ of size greater than 1. If there are such sets X_i with $i \geq 1$, then they are *TA*. They are called the *TA components* of X. By Corollary 2.2 we have that $P_C(X)$ is a coding partition (indeed $P_C(X) \leq P(X)$) and it is called the *canonical coding partition* of X: it defines a *canonical decomposition* of a code X in at most one unambiguous component and a (possibly empty) set of *TA* components. Roughly speaking, if a code X is not *UD*, then its canonical decomposition, on one hand separates the unambiguous component of the code (if any), and, on the other, localizes the ambiguities inside the *TA* components of the code. On the contrary, if X is *UD*, then its canonical decomposition contains only the unambiguous component X_0. Moreover if X is *UD* then every partition of X is a coding partition.

Theorem 2.5 *There is a Sardinas-Patterson like algorithm to compute the canonical coding partition of a finite code X.*

Example 2.6 *Let us consider the code $X \subseteq \{0,1\}^*$, $X = \{00, 0010, 1000, 11, 1111,$ $010, 011\}$. In [3] it is shown that the canonical coding partition of X is $P_C(X) = \{X_0, X_1, X_2\}$ with $X_0 = \{010, 011\}$, $X_1 = \{00, 0010, 1000\}$, $X_2 = \{11, 1111\}$.*

Theorem 2.7 *Given a regular code X and a partition $P = \{X_1, \ldots, X_n\}$ of X such that X_i, for $i = 1, \ldots, n$, is a regular set, it is decidable whether P is a coding partition of X.*

Still in [3], it was conjectured that *if X is regular, the number of classes of $P_C(X)$ is finite and each class of $P_C(X)$ is a regular set.*

Finally, the positive answer has given in [1] where the following theorem and corollary are proved.

Theorem 2.8 *The canonical partition of a regular code is finite and regular. Its classes can be effectively computed.*

Corollary 2.9 *Given a regular code X and a regular partition $P = \{X_1, X_2, \ldots, X_n\}$ of X, it is decidable whether P is the canonical coding parition of X.*

From the definition of coding partition we deduce immediately the next theorem that gives a tool to construct infinitely many *UD* codes starting from any non-*TA* code with more than one code word.

Theorem 2.10 *Let $P = \{X_i \mid i \in I\}$ be a coding partition of a code with $|I| > 1$.*
Then the sets $\{X_{i_1}^+ X_{i_2}^+ \cdots X_{i_n}^+ \mid n \geq 2, i_j \in I, i_j \neq i_{j+1}\ \forall\, 1 \leq j < n, i_n \neq i_1\}$ are UD codes.

We conclude this section with the following theorem concerning the regularity of the classes of a finite coding partition of a regular code.

Theorem 2.11 *Let $\{Y_j \mid j \in J\}$ be a coding partition of a regular code X and let X_0 be the unambiguous component of X. If there exists $j_1 \in J$ such that Y_{j_1} is not regular then we have $Y_{j_1} \cap X_0 \neq \emptyset$. Moreover if J is finite then there exists $j_2 \in J$, $j_2 \neq j_1$ such that also Y_{j_2} is not regular and $Y_{j_2} \cap X_0 \neq \emptyset$.*

Proof. Let $P_C(X) = \{X_0, X_1, \ldots, X_n\}$ be the regular and finite canonical coding partition of X. If, by contradiction, $Y_{j_1} \cap X_0 = \emptyset$ then, recalling how $P_C(X)$ rises from $P(X)$ and recalling that $P(X)$ is the finest coding partition of X, we see that Y_{j_1} is a finite union of some of the regular codes $\{X_1, \ldots, X_n\}$ and so it is regular: a contradiction. Then $Y_{j_1} \cap X_0 \neq \emptyset$. Moreover if J is finite then if, by contradiction, all the Y_j for $j \neq j_1$ where regular, then Y_{j_1} where the complement, with respect to the regular code X of a regular code and so Y_{j_1} where regular against the hypothesis. Then there exists $j_2 \in J$, $j_2 \neq j_1$ such that also Y_{j_2} is not regular and, by the first part of the proof, $Y_{j_2} \cap X_0 \neq \emptyset$. □

Example 2.12 *Let X be the regular UD code $X = a^+b^+$. Then $X_0 = X$ and put $Y_1 := \{a^n b^n \mid n \geq 1\}$, $Y_2 := X \setminus Y_1$ we have that $P = \{Y_1, Y_2\}$ is a coding partition of X in two non-regular classes.*

3 Free factorizations of a monoid

In this section the previous results are restated in an algebraic setting making use of the free product of monoids.

Given a code $X \subseteq A^*$ we can study the properties of the monoid $M = X^*$. On the contrary, if we start with a monoid $M \subseteq A^*$, we can study the characteristic properties of the different sets $X \subseteq A^+$ of generators of M. We recall that any submonoid M of A^* has a unique minimal set of generators $X = (M \setminus 1) \setminus (M \setminus 1)^2$, where 1 is the empty word (see [2]); in such a case we say that X is the base of M. In general we say that a code X is *a base* if X is the base of X^*.
It is natural to investigate how the properties of a partition of a code are related to those of the monoids generated by the classes of the partition.

Given a partition $P = \{X_i \mid i \in I\}$ of a code $X \subseteq A^+$, the condition that every word $w \in X^+$ admits a unique *P-factorization* has a natural algebraic interpretation in terms of free product of monoids.

Let M be a monoid generated by submonoids $M_\lambda, \lambda \in \Lambda$, and let $m \in M$. An expression of m of the form $m_1 m_2 \cdots m_r$, where $r \geq 0$, $1 \neq m_i \in M_{\lambda_i}$, $\lambda_i \neq \lambda_{i+1}$, is said in *reduced form* with respect to M_λ's. By definition, M is the free product of the M_λ's iff every element of M has an unique expression in reduced form with respect to M_λ's and we write $M = Fr_{\lambda \in \Lambda} M_\lambda$. In the finite case we also write $M = M_{\lambda_1} * \cdots * M_{\lambda_n}$.

The previous results can be expressed then in the following form.

Theorem 3.1 *Let $X \subseteq A^+$ be a code, let $P = \{X_i \mid i \in I\}$ be a partition of X and let $M = X^*$, $M_i = X_i^*$ with $i \in I$. If P is a coding partition of X then M is the free product of the M_i's. Conversely let M be the free product of the submonoids M_i's, let X_i be sets of generators of M_i and let $X = \bigcup_{i \in I} X_i$. Then $P = \{X_i \mid i \in I\}$ it is a coding partition of X.*

It's natural at this point to introduce the notion of free factorizations of a monoid.

Definition 3.2 *A family $\{M_\lambda \mid \lambda \in \Lambda\}$ of submonoids of M is a* free factorization *of M if M is the free product of the M_λ's. The M_λ's are called the* free factors *of the free factorization; moreover we say that a monoid M is* freely indecomposable *if M cannot be expressed as a free product of nontrivial monoids.*

We stress that a free factor is not, in general, a free monoid.

Remark 3.3 *We note that a monoid M is freely indecomposable iff any set of generators of M is a totally ambiguous code. From another hand we have that a code X is UD iff $X^* = Fr_{x\in X}\{x\}^*$ so, in particular, the monoid X^* is free.*

The next proposition comes directly from the definition of free product of monoids: it is the Corollary 2.2 restated in terms of monoids.

Proposition 3.4 *Let $M = Fr_{\lambda\in\Lambda} M_\lambda$ and let $\{\Lambda_\mu \mid \mu \in \Gamma\}$ be a partition of Λ. Set $\forall \mu \in \Gamma$, M_μ the monoid generated by $\{M_\lambda \mid \lambda \in \Lambda_\mu\}$ then $M_\mu = Fr_{\lambda\in\Lambda_\mu} M_\lambda$ and $M = Fr_{\mu\in\Gamma} M_\mu$.*

Starting with an arbitrary family of submonoids of A^*, analogously to what we have made with a code X, we can partition the family in classes in such a way that the monoid generated by the family is the free product of the monoids generated by each class of the partition. On the contrary, if we have a monoid M, we can consider the family of all the free factorizations of M and define a partial order on this family.

Definition 3.5 *Let $F_1 = \{M_\mu \mid \mu \in \Lambda_1\}$, $F_2 = \{M_\lambda \mid \lambda \in \Lambda_2\}$ be two free factorizations of a monoid M. We say that $F_1 \leq F_2$ if there exists a partition $\{\Lambda_\mu \mid \mu \in \Lambda_1\}$ of Λ_2 such that for each μ, $M_\mu = Fr_{\lambda\in\Lambda_\mu} M_\lambda$.*

By Theorem 2.3 and Theorem 3.1 we deduce the following theorem.

Theorem 3.6 *Given a monoid M the family of the free factorizations of M is a complete lattice.*

As in the case of the canonical partition of a code, *the* finest free factorization of a monoid M is called the *characteristic* free factorization of M and it is denoted by $\mathscr{F}(M)$ or, if we want to make the free factors explicit, $\mathscr{F}(M) = Fr_{\lambda\in\Lambda} M_\lambda$.

Now let M_0 be the monoid generated by all the free factors of $\mathscr{F}(M)$ having only one generator. The monoid M_0 is then a free monoid and it is called the *free component* of M. From $\mathscr{F}(M)$ one then derives another decomposition of M

$$\mathscr{F}_C(M) = M_0 * Fr_{\lambda\in\Lambda} M_\lambda,$$

where the M_λ 's are the free factors of $\mathscr{F}(M)$ having more then one generator. If there are such monoids M_λ then they are not free and they are, of course, freely indecomposable. They are called the *freely indecomposable components* of M. By Proposition 3.4 we have that $\mathscr{F}_C(X)$ is a free factorization of M (indeed $\mathscr{F}_C(M) \leq \mathscr{F}(M)$) and it is called the *canonical free factorization* of M: it defines a *canonical decomposition* of a monoid M in at most one free component and a (possibly empty) set of freely indecomposable components.

Example 3.7 *Let $A = \{a_1, a_2, \dots\}$. Then $\mathscr{F}(A^*) = (a_1^*) * (a_2^*) * \cdots$, and $\mathscr{F}_C(A^*) = \{A^*\}$. Then the poset of the free factorizations of A^* are in bijection with the poset of the alphabet A.*

Already in [1], the following equivalent formulation of Theorem 2.8 is given.

Theorem 3.8 *Any regular submonoid $M \subseteq A^*$ admits a canonical decomposition into a free product of at most one regular free submonoid and finitely many (possibly zero) regular freely indecomposable submonoids.*

Example 3.9 *Let $A = \{a,b,c,d\}$ and let $X \subseteq A^+$ be the following regular code: $X = a + bb + c + ad^*b + bc^*bb$.*
In [1] it is shown that $P_C(X) = \{X_0, X_1\}$ where $X_0 = ad^+b$ and $X_1 = a + ab + bb + c + bc^*bb$. Then the canonical decomposition of the regular submonoid X^* is $X^* = (X_0^*) * (X_1^*)$.

4 Full monoids and maximal codes

Using ideas of previous section we introduce a partial order in the family of the submonoids of A^*. We will prove that, in this poset, there exist maximal elements. We call this maximal elements *full* monoids and we will say that a code is *maximal* if it is the base of a full monoid. We show that this definition of maximality extends that concerning *UD* codes and, with Theorem 4.14, we will give a characterization of maximal *UD* codes depending only on the monoid they generate.

Definition 4.1 *Let $M, N \subseteq A^*$ be monoids we say that $M \preceq N$ if there exists a monoid $L \subseteq A^*$ such that $N = M * L$.*

Proposition 4.2 *The relation $\preceq$ is a partial order on the set of submonoids of A^*.*

Proof. We need to prove that $\preceq$ is transitive and antisymmetric. If $L \preceq M$ and $M \preceq N$ then $\exists L', M'$ such that $M = L * L'$ and $N = M * M'$. Then $N = (L * L') * M' = L * (L' * M')$ and so $N \preceq L$. Now let $M \preceq N$ and $N \preceq M$ so $M = N * N'$ and $N = M * M'$ for some monoids M', N'. Then $M = M * M' * N'$ thus M', N' are trivial monoids and so $M = N$. □

The first question is, given a monoid N, if there exists a monoid M with $N \subseteq M$ and M maximal with respect to the partial order $\preceq$.

To answer to the previous question we first prove the following lemma.

Lemma 4.3 *Let $M = M_1 * M_2$ and let X, X_1, X_2 be the base of M, M_1, M_2 respectively. Then $X = X_1 \cup X_2$.*

Proof. Since $M = M_1 * M_2$ and X_1, X_2 are the bases of M_1 and M_2 respectively, it is clear that $X_1 \cup X_2$ is a set of generators of M. Let, by contradiction, $X \subsetneq X_1 \cup X_2$ and let $x' \in (X_1 \cup X_2) \setminus X$. We can assume that $x' \in X_1$. Since X is a set of generators of M, $x' = x_1 x_2 \cdots x_n$ with $x_i \in X$. But $x' \in M_1$ and, by the uniqueness of the reduced form with respect to M_1 and M_2, we have $x_i \in M_1$, $\forall\, 1 \leq i \leq n$, and so $x_i \in X_1$, $\forall\, 1 \leq i \leq n$. This shows that $X_1 \setminus \{x'\}$ is a set of generators of M_1: a contradiction. Thus $X_1 \cup X_2$ is a minimal set of generators of M and we have the thesis. □

As an obvious generalization we have the following

Corollary 4.4 *Let $M = Fr_{\lambda \in \Lambda} M_\lambda$ and let X_λ, $\lambda \in \Lambda$ and X be the bases of M_λ, $\lambda \in \Lambda$ and M respectively. Then $X = \cup_{\lambda \in \Lambda} X_\lambda$.*

We note that without Lemma 4.3, by Theorem 3.1, we only say that $Y := X_1 \cup X_2$ is a set of generators of M and that $P = \{X_1, X_2\}$ is a coding partition of Y. Lemma 4.3 says that Y is the base of M.

Now we can prove the following theorem.

Theorem 4.5 *Any submonoid $M \subseteq A^*$ is contained in a submonoid $N \subseteq A^*$, which is maximal with respect to $\preceq$ and such that $M \preceq N$.*

Proof. We will make use of Zorn's lemma. Let $\mathfrak{F}$ be the family of all the submonoids $P \subseteq A^*$, ordered by $\preceq$, such that $M \preceq P, \forall P \in \mathfrak{F}$ and let $\{M_\lambda \mid \lambda \in \Lambda\}$ be a chain in $\mathfrak{F}$. If $\lambda < \gamma$ then there exists a submonoid $H_{\lambda,\gamma} \subseteq A^*$ such that $M_\gamma = M_\lambda * H_{\lambda,\gamma}$ and so if we call X_γ, X_λ and $X_{\lambda,\gamma}$ the bases of $M_\gamma, M_\lambda, H_{\lambda,\gamma}$ respectively, by Lemma 4.3, $X_\gamma = X_\lambda \cup X_{\lambda,\gamma}$ and then $X_\lambda \subsetneq X_\gamma$. Now, $\forall\, \lambda \in \Lambda$, let X_λ the base of M_λ, $Y := \cup_{\lambda\in\Lambda} X_\lambda$ and let N be the monoid generated by Y. We show that $M_\lambda \preceq N$, $\forall\, \lambda \in \Lambda$. Let $Z_\lambda := Y \setminus X_\lambda$ and let H_λ the submonoid generated by Z_λ. We will prove that $N = M_\lambda * H_\lambda$. Let $m \in N$ and let us suppose, by contradiction, that m has two different expressions in reduced form with respect to M_λ, H_λ so $m = m_1 m_2 \cdots m_r = m'_1 m'_2 \cdots m'_s$ with $r, s \geq 1$. Since N is generated by Y then $m = y_1 y_2 \cdots y_h = y'_1 y'_2 \cdots y'_k$ for certain $y_i, y'_j \in Y$ and, since the two expressions in reduced form with respect to M_λ, H_λ are different, $\exists y \in \{y_1, y_2, \ldots, y_h, y'_1, y'_2, \ldots, y'_k\}$ such that $y \notin X_\lambda$. Let $\lambda_1 \in \Lambda$ such that $\lambda_1 > \lambda$ and $y_i, y'_j \in X_{\lambda_1}$, $\forall\, i, j$. Then $M_{\lambda_1} = M_\lambda * H_{\lambda,\lambda_1}$ for a certain $H_{\lambda,\lambda_1} \subseteq A^*$. Since $m, m_i, m'_j \in M_{\lambda_1}$, $\forall\, i, j$, then the two different expressions of m in reduced form with respect to M_λ, H_λ are still two different expressions in reduced form with respect to $M_\lambda, H_{\lambda,\lambda_1}$. This contradiction shows that $N = M_\lambda * H_\lambda$ and thus $M_\lambda \preceq N$, $\forall\, \lambda \in \Lambda$. Since $M \preceq M_\lambda$, $\forall\, \lambda \in \Lambda$ then $M \preceq N$ so $N \in \mathfrak{F}$ and it is a upper bound for the chain $\{M_\lambda \mid \lambda \in \Lambda\}$. Invoking Zorn's lemma we have the thesis. □

Remark 4.6 *By Example 3.7 we see that if M is not generated by a subset of the alphabet A, then the maximal monoid N which the previous theorem refers to, is properly contained in A^* i.e. $M \preceq N \subsetneq A^*$.*

We give now the following definition.

Definition 4.7 *We say that a submonoid M of A^* is* full *if it is maximal with respect to the partial order $\preceq$.*

Remark 4.8 *From the definition we have that if $M' \subseteq M$ and M' is full then also M is full.*

A first statement on full monoids is given by the following proposition.

Proposition 4.9 *Let $M \subseteq A^*$ be a monoid. If M is maximal with respect to the inclusion order $\subseteq$ then it is full.*

Proof. We will prove that if M is not full then it is not maximal with respect to the inclusion order $\subseteq$. If M is not full then there exist a monoid $N \subseteq A^*$ and a non trivial monoid $M_1 \subseteq A^*$ such that $N = M * M_1$. Let X the base of M_1, $x \in X$ and let M_2 be the monoid $(x^2)^*$. Then we have $M \subsetneq M * M_2 \subsetneq N$. □

We recall that the submonoids of A^* maximal with respect to the inclusion order $\subseteq$ are "few": in fact it is easy to see that a submonoid M of A^* is maximal with respect to the inclusion order $\subseteq$ iff $M = A^* \setminus \{a\}$ for a certain $a \in A$.

A *UD* code $X \subseteq A^+$ is said to be a *maximal UD* code if X is not properly contained in any other *UD* code over A.

Now we extend the notion of maximality to codes that are not *UD*.

Definition 4.10 *A code $X \subseteq A^+$ is said* maximal *if the monoid X^* is full.*

The next theorem shows how this notion generalizes that of maximality given for *UD* codes.

Theorem 4.11 *Let X be a UD code. Then X is a maximal UD code iff X^* is a full monoid.*

Proof. If X is a maximal *UD* code then $\forall w \in A^+$, $X' := X \cup \{w\}$ is not a *UD* code and, by Remark 3.3 and Proposition 3.4, this imply that $\forall w \in A^+$, $(X')^*$ is not the free product of X^* and $\{w\}^*$ and this is true iff X^* is full. □

A free monoid $M \subseteq A^*$ is said *maximal free* if $M \neq A^*$ and M is not properly contained in any other free monoid different from A^*.

If a free monoid is maximal free then it is full. Indeed if a free monoid is maximal free then its base is a maximal *UD* code (see [2]) so by Theorem 4.11 the monoid is full.
We have proved then the following theorem.

Theorem 4.12 *Let M be a free monoid. If M is maximal free then it is full.*

Remark 4.13 *In [2] it is proved that uniform codes A^n are maximal UD codes $\forall n \geq 1$ and it is been underlined that with $n = lm, l, m > 1$, we have $(A^n)^* \subsetneq (A^m)^* \subsetneq A^*$. This has two consequences: from one hand, by Theorem 4.11, we can see that the inverse of Proposition 4.9 is false, moreover, since the monoids $(A^n)^*$ are free, again by Theorem 4.11, also the inverse of Theorem 4.12 is false.*

Recalling that if M is a free monoid then its base is a *UD* code, then from Theorem 4.11 we have the following characterization of a maximal *UD* codes in terms of algebraic properties of the monoid generated by the code itself.

Theorem 4.14 *Let $X \subseteq A^+$ be a code that is a base. Then X is a maximal UD code iff X^* is a full and free submonoid of A^*.*

We see now how with this notion of maximality we will recover some results concerning the *UD* codes.

We first recall some definitions.
A word $w \in A^*$ is a *factor* of a word $z \in A^*$ if there exist $u, v \in A^*$ such that $z = uwv$. For any $X \subseteq A^*$ let $F(X)$ denote the set of factors of words in X.
A set $X \subseteq A^*$ is *dense* if $F(X) = A^*$. A set that is not dense is called *thin*.
Finally, a set $X \subseteq A^*$ is *complete* if X^* is dense.

Theorem 4.15 *Let $X \subseteq A^+$ be a maximal code then it is a complete set.*

Proof. Let X be a code over the alphabet A, with $card(A) \geq 2$ (the case $card(A) < 2$ is trivial). We will prove that if X is not complete then X^* is not full. If X is not complete, there exists a word $v \in A^*$ such that v does not belong to $F(X^*)$. Let a be the first letter of v and let $b \in A \setminus \{a\}$. Consider the word $w = vb^{|v|-1}$. By construction, w is *unbordered*, i.e. no proper prefix of w is a suffix of w. Since v does not belong to $F(X^*)$, we have that also w does not belong to $F(X^*)$. Let $M := (X \cup \{w\})^*$ we now prove that every word $t \in (X \cup \{w\})^*$ has an unique expression in reduced form with respect to $X^*, \{w\}^*$. Indeed, since w is unbordered, we can uniquely distinguish all occurrences of w in t, i.e. t has a unique factorization of the form

$$t = u_1 w u_2 w \cdots w u_n,$$

with $n \geq 1$ and $u_i \in X^*$, for $i = 1, \ldots, n$.
This shows that $M = (X^*) * (w^*)$ and X^* is not full. □

By the previous theorem we deduce the following corollary.

Corollary 4.16 *Any full monoid $M \subseteq A^*$ is dense in A^*.*

The inverse of previous corollary is not true. Indeed the Dyck code D over $A = \{a,b\}$ is a *UD* dense code and for each $x \in D$ the code $D \setminus \{x\}$ remains dense (see [2]) but it is no more a maximal *UD* code and so by Theorem 4.14 $(D \setminus \{x\})^*$ it is not full in A^*.

The next lemma holds (see [2]).

Lemma 4.17 *Let $X \subseteq A^+$ be a thin and complete code. Then all words $w \in A^*$ satisfy*

$$(X^* w X^*)^+ \cap X^* \neq \emptyset.$$

Then we can prove the following theorem.

Theorem 4.18 *Let X be a thin code. If X is complete then it is maximal.*

Proof. Let $M \subseteq A^+$ be a monoid and let $1 \neq w \in M$. By previous lemma there exist $v_1, v_2 \in X^*$ and $z \in X^+$ such that $z = (v_1 w v_2)^+$. From this z has not a unique expression in reduced form with respect to X^* and M. Then X^* is full and X is a maximal code. □

Putting together the last two results we have:

Theorem 4.19 *Let $X \subseteq A^+$ be a thin code. Then X is complete iff it is maximal.*

Again in [2], the following result is proved.

Proposition 4.20 *Any regular UD code is thin.*

Indeed the proof of the cited result shows the following more general proposition.

Proposition 4.21 *Any regular code that is a base is thin.*

Then we can conclude with the following corollary.

Corollary 4.22 *Let $X \subseteq A^+$ be a regular code that is a base. Then X is complete iff it is maximal.*

5 Concluding remarks

In this paper we have given a definition of maximality that extends the existing one for *UD* codes re-establishing, in the general case, some classical results valid for *UD* codes. At this point it is interesting to understand which, among the deep results concerning maximal *UD* codes, can be recovered from the more general definitions of maximality and coding partition. (We emphasize that the notion of coding partition generalizes that of *UD* code: the "uniquely decipherability" at the level of classes of the partition takes the place of the uniquely decipherability existing between the words of a *UD* code.) Two subjects that it is possible to deepen are composition of codes and probability distributions.

References

[1] Marie-Pierre Béal, Fabio Burderi & Antonio Restivo (2009): *Coding partitions of regular sets*. *Inter. Jour. Alg. Comput.* Vol 19, No 8(8), pp. 1011–1023, doi:10.1142/S0218196709005457.

[2] Jean Berstel, Dominique Perrin & Christophe Reutenauer (2010): *Codes and Automata*. *Encyclopedia of Mathematics and its Applications* 129, Cambridge University Press.

[3] Fabio Burderi & Antonio Restivo (2007): *Coding partitions*. *Discret. Math. Theor. Comput. Sci.* Vol 9, No 2(2), pp. 227–240.

[4] Fabio Burderi & Antonio Restivo (2007): *Varieties of Codes and Kraft Inequality*. *Theory Comput. Systems* Vol 40, pp. 507–520, doi:10.1007/s00224-006-1320-0.

[5] M. Dalai & R. Leonardi (2005): *Non prefix-free codes for constrained sequences*. In: *International Symposium on Information Theory, 2005. ISIT 2005*, IEEE, pp. 1534–1538, doi:10.1109/ISIT.2005.1523601.

[6] A. Ehrenfeucht & G. Rozemberg (1986): *Each regular code is included in a maximal regular code*. *RAIRO Inform. Theor. Appl.* 20, pp. 89–96.

[7] Güney Gönenç (1973): *Unique decipherability of codes with constraints with application to syllabification of Turkish words*. In: *COLING 1973: Computational And Mathematical Linguistics: Proceedings of the International Conference on Computational Linguistics*, 1, pp. 183–193.

[8] Fernando Guzmán (1999): *Decipherability of codes*. *J. Pure Appl. Algebra* 141(1), pp. 13–35, doi:10.1016/S0022-4049(98)00019-X.

[9] Tom Head & Andreas Weber (1995): *Deciding Multiset Decipherability*. *IEEE Trans. Inform. Theory* 41(1), pp. 291–297, doi:10.1109/18.370097.

[10] Abraham Lempel (1986): *On multiset decipherable codes*. *IEEE Trans. Inform. Theory* 32(5), pp. 714–716, doi:10.1109/TIT.1986.1057217.

[11] Antonio Restivo (1989): *A note on multiset decipherable codes*. *IEEE Trans. Inform. Theory* 35(3), pp. 662–663, doi:10.1109/18.30991.

[12] Andreas Weber & Tom Head (1996): *The Finest Homophonic Partition and Related Code Concepts*. *IEEE Trans. Inform. Theory* 42(5), pp. 1569–1575, doi:10.1109/18.532902.