Pcodes of Partial Words

F. Blanchet-Sadri*and Margaret Moorefield
Department of Mathematical Sciences
University of North Carolina
P.O. Box 26170
Greensboro, NC 27402–6170, USA
blanchet@uncg.edu

October 14, 2005

Abstract

Codes play an important role in the study of combinatorics on *words*. Recently, we introduced *pcodes* that play a role in the study of combinatorics on *partial words*. Partial words are strings over a finite alphabet that may contain a number of "do not know" symbols. In this paper, the theory of codes of words is revisited starting from pcodes of partial words. We present some important properties of pcodes. We give several equivalent definitions of pcodes and the monoids they generate. We investigate in particular the Defect Theorem for partial words. We describe an algorithm to test whether or not a finite set of partial words is a pcode. We also discuss two-element pcodes, complete pcodes, maximal pcodes, and the class of circular pcodes. A World Wide Web server interface has been established at http://www.uncg.edu/mat/pcode/ for automated use of the program.

Keywords: Partial word, pcode, monoid, complete pcode, maximal pcode, circular pcode.

1 Introduction

Partial words are strings of symbols from a finite alphabet that may have a number of "do not know" symbols. While a word can be described by a total function, a partial word can be described by a partial function (see Section 2). Some combinatorial properties of partial words have been investigated [2, 4, 5, 6, 7, 8, 9, 10], and more research is underway.

*This material is based upon work supported by the National Science Foundation under Grant CCF–0207673. We thank the referees of preliminary versions of this paper for their very valuable comments and suggestions.

Codes play an important role in the study of the combinatorics on words. In our recent paper [5], we introduced *pcodes* that play a role in the study of combinatorics on *partial words*. While a code of words X does not allow two distinct decipherings of some word in X^+, a pcode of partial words Y does not allow two distinct *compatible* decipherings in Y^+. We described various ways of defining and analyzing pcodes. In particular, many pcodes can be obtained as antichains with respect to certain partial orderings. Adapting a technique related to dominoes [3, 12, 13], we showed that the pcode property is decidable. In this paper, the theory of codes of words, as exposited in [1], is revisited starting from pcodes of partial words. Proofs are omitted when they follow the word case very closely. Section 2 reviews preliminary notions on partial words. In particular, the roles of *compatibility* and *conjugacy* are described. The definitions and some important general properties of pcodes are presented in Section 3. Section 4 contains several equivalent definitions of pcodes and the monoids they generate. There, we investigate in particular the *Defect Theorem* for partial words, a problem that is proposed in [15]. In Section 5, we give an analog of the Sardinas-Paterson algorithm for testing whether or not a given finite set of partial words is a pcode. Section 6 discusses the concept of a *complete pcode*. In Section 7, we introduce the *circular pcodes* which take into account, in a natural way, the operation of conjugacy. The main feature of these pcodes is that they define a unique factorization of partial words written on a circle. Section 8 discusses two-element pcodes.

2 Preliminaries

In this section, we give a short review of some basic notions on partial words that will be used throughout the paper.

Let A be a nonempty finite set of symbols, which we call an *alphabet*. An element $a \in A$ is called a *letter*. A *word* over the alphabet A is a finite sequence of elements of A. The empty sequence is called the *empty word* and is denoted by ϵ. The set of all words over A is denoted by A^* and is equipped with the associative operation defined by the concatenation of two sequences. The empty word is the neutral element for concatenation. Thus the set A^* is equipped with the structure of a monoid. It is called the *free monoid* over A. The set of nonempty words over A is denoted by A^+. Thus we have $A^+ = A^* \setminus \{\epsilon\}$, and A^+ is equipped with the structure of a semigroup. It is called the *free semigroup* over A.

A word of length n over A can be defined by a total function $u : \{0, \ldots, n-1\} \to A$ and is usually represented as $u = a_0 a_1 \ldots a_{n-1}$ with $a_i \in A$. A *partial word* of length n over A is a partial function $u : \{0, \ldots, n-1\} \to A$. For $0 \leq i < n$, if $u(i)$ is defined, then we say that i belongs to the *domain* of u (denoted by $i \in D(u)$), otherwise we say that i belongs to the *set of holes* of u (denoted by $i \in H(u)$). A word over A is a partial word over A with an empty set of holes (we sometimes refer to words as *full* words). For any partial word u over A, $|u|$ denotes its length. Clearly, $|\epsilon| = 0$. We denote by $W_0(A)$ the set A^*, and for $i \geq 1$, by $W_i(A)$ the set of partial words over A with at most i holes. We put $W(A) = \bigcup_{i \geq 0} W_i(A)$, the set of all partial words over A with an arbitrary number of holes.

If u is a partial word of length n over A, then the *companion* of u (denoted by $u_\diamond$) is the total function $u_\diamond : \{0, \ldots, n-1\} \to A \cup \{\diamond\}$ defined by

$$u_\diamond(i) = \begin{cases} u(i) & \text{if } i \in D(u), \\ \diamond & \text{otherwise.} \end{cases}$$

The symbol $\diamond \notin A$ is viewed as a "do not know" symbol. The word $u_\diamond = abb\diamond b\diamond cb$ is the companion of the partial word u of length 8 where $D(u) = \{0, 1, 2, 4, 6, 7\}$ and $H(u) = \{3, 5\}$. The bijectivity of the map $u \mapsto u_\diamond$ allows us to define for partial words concepts such as concatenation and powers in a trivial way. More specifically, for partial words u, v, the concatenation of u and v is defined by $(uv)_\diamond = u_\diamond v_\diamond$, and the i-power of u is defined by $(u^i)_\diamond = (u_\diamond)^i$ where $(u_\diamond)^0 = \epsilon$ and $(u_\diamond)^i = u_\diamond (u_\diamond)^{i-1}$. The set $W(A)$ is a monoid under the concatenation of partial words (ϵ serves as identity). For convenience in the sequel, we consider a partial word over A as a word over the enlarged alphabet $A \cup \{\diamond\}$, where the additional symbol $\diamond$ plays a special role. Thus, we say for instance "the partial word $\diamond ab\diamond b$" instead of "the partial word with companion $\diamond ab\diamond b$".

Given two subsets X, Y of $W(A)$, we define

$$XY = \{uv \mid u \in X \text{ and } v \in Y\}.$$

We sometimes write $X \sqsubset Y$ if $X \subset Y$ but $X \neq Y$. For a subset X of $W(A)$, we use the notation $\|X\|$ for the cardinality of X.

A *factorization* of a partial word u is any sequence $u_1, u_2, \ldots, u_i$ of partial words such that $u = u_1 u_2 \ldots u_i$. For a subset X of $W(A)$ and integer $i \geq 0$, we denote by X^i the set

$$\{u_1 u_2 \ldots u_i \mid u_1, \ldots, u_i \in X\}.$$

We denote by X^* the submonoid of $W(A)$ generated by X, or $X^* = \bigcup_{i\geq 0} X^i$ where $X^0 = \{\epsilon\}$, and by X^+ the subsemigroup of $W(A)$ generated by X, or $X^+ = \bigcup_{i>0} X^i$. By definition, each partial word u in X^* admits at least one factorization $u_1, u_2, \ldots, u_i$ whose elements are all in X. Such a factorization is called an *X-factorization*.

A partial word u is a *factor* of a partial word v if there exist partial words x, y such that $v = xuy$. The factor u is *proper* if $u \neq v$. The partial word u is a *prefix* (respectively, *suffix*) of v if $x = \epsilon$ (respectively, $y = \epsilon$). For a subset X of $W(A)$, we denote by $F(X)$ the set of factors of elements in X. More specifically,

$$F(X) = \{u \mid u \in W(A) \text{ and there exist } x, y \in W(A) \text{ such that } xuy \in X\}.$$

The *reverse* of a word $u = a_0a_1 \ldots a_{n-1}$ is $rev(u) = a_{n-1} \ldots a_1a_0$. The *reverse* of a partial word u is $rev(u)$ where $(rev(u))_\diamond = rev(u_\diamond)$. The *reverse* of a set $X \subset W(A)$ is the set $rev(X) = \{rev(u) \mid u \in X\}$.

A *period* of a partial word u over A is a positive integer p such that $u(i) = u(j)$ whenever $i, j \in D(u)$ and $i \equiv j \bmod p$. In such a case, we call u *p-periodic*.

2.1 Compatibility

In this section, we discuss compatibility on partial words.

If u and v are two partial words of equal length, then u is said to be contained in v, denoted by $u \subset v$, if all elements in $D(u)$ are in $D(v)$ and $u(i) = v(i)$ for all $i \in D(u)$. We sometimes write $u \sqsubset v$ if $u \subset v$ but $u \neq v$. The order $u \subset v$ on partial words is obtained when we let $\diamond \prec a$ and $a \preceq a$ for all $a \in A$. A partial word u is called *primitive* if there exists no word v such that $u \subset v^n$ with $n \geq 2$. Observe that the empty word is not primitive. If u is a nonempty partial word, then there exists a primitive word v and a positive integer n such that $u \subset v^n$ [6]. Uniqueness does not hold since the partial word $u = \diamond a$ is such that $u \subset a^2$ and $u \subset ba$ for distinct letters a, b. A nonempty partial word u is called *unbordered* if no nonempty words x, v, w exist such that $u \subset xv$ and $u \subset wx$ (otherwise, it is called *bordered*) [6]. Unbordered partial words turn out to be primitive.

The partial words u and v are called *compatible*, denoted by $u \uparrow v$, if there exists a partial word w such that $u \subset w$ and $v \subset w$. We denote by $u \vee v$ the least upper bound of u and v (in other words, $u \subset u \vee v$ and $v \subset u \vee v$ and $D(u \vee v) = D(u) \cup D(v)$). As an example, $u = aba\diamond\diamond a$ and $v = a\diamond\diamond b\diamond a$ are compatible and $(u \vee v) = abab\diamond a$. For a subset

X of $W(A)$, we denote by $C(X)$ the set of all partial words compatible with elements of X. More specifically,

$$C(X) = \{u \mid u \in W(A) \text{ and there exists } v \in X \text{ such that } u \uparrow v\}.$$

If $X = \{u\}$, then we denote $C(\{u\})$ simply by $C(u)$. We call a subset X of $W(A)$ *pairwise non compatible* if no distinct partial words $u, v \in X$ satisfy $u \uparrow v$. In other words, X is pairwise non compatible if for all $u \in X$, $X \cap C(u) = \{u\}$.

The following rules are useful for computing with partial words [2].

Multiplication: If $u \uparrow v$ and $x \uparrow y$, then $ux \uparrow vy$.

Simplification: If $ux \uparrow vy$ and $|u| = |v|$, then $u \uparrow v$ and $x \uparrow y$.

Weakening: If $u \uparrow v$ and $w \subset u$, then $w \uparrow v$.

We end this section with the following lemma.

Lemma 1 ([2]) *Let $u, v, x, y \in W(A)$ be such that $ux \uparrow vy$.*

- *If $|u| \geq |v|$, then there exist $w, z \in W(A)$ such that $u = wz$, $v \uparrow w$, and $y \uparrow zx$.*
- *If $|u| \leq |v|$, then there exist $w, z \in W(A)$ such that $v = wz$, $u \uparrow w$, and $x \uparrow zy$.*

2.2 Conjugacy

In this section, we discuss conjugacy on partial words. Two partial words u and v are called *conjugate* if there exist partial words x and y such that $u \subset xy$ and $v \subset yx$ [10].

Lemma 2 ([10]) *Let $u, v \in W(A) \setminus \{\epsilon\}$ and let $z \in A^*$. If $uz \uparrow zv$, then there exist words x, y such that $u \subset xy$, $v \subset yx$, and $z \subset (xy)^n x$ for some integer $n \geq 0$.*

Lemma 2 does not necessarily hold if z is not full even if u, v are full. The partial words $u = a$, $v = b$, and $z = \diamond bb$ provide a counterexample. However, the following holds.

Lemma 3 ([10]) *Let $u, v \in W(A) \setminus \{\epsilon\}$ and let $z \in W(A)$. If $uz \uparrow zv$ and $uz \vee zv$ is $|u|$-periodic, then there exist words x, y such that $u \subset xy$, $v \subset yx$, and $z \subset (xy)^n x$ for some integer $n \geq 0$.*

Throughout the rest of this paper, A denotes a fixed alphabet.

3 Definitions

This section contains the definitions of *pcode*, *prefix* (respectively, *suffix*, *biprefix*) *pcode*, *maximal pcode*, and *pcoding morphism*.

Definition 1 ([5]) *Let X be a subset of $W(A)$. Then X is called a* pcode *over A if for all integers $m, n \geq 1$ and partial words $u_1, \ldots, u_m, v_1, \ldots, v_n \in X$, the compatibility relation*

$$u_1 u_2 \ldots u_m \uparrow v_1 v_2 \ldots v_n$$

implies $m = n$ and $u_i = v_i$ for $i = 1, \ldots, m$.

A nonempty set X satisfying $X \subset \{a, \diamond\}^*$ is a pcode over $\{a\}$ if and only if X is a singleton set distinct from $\{\epsilon\}$. The set $Y = \{a\diamond b, a\diamond\}$ is a pcode over $\{a, b\}$. But the set $Z = \{u_1, u_2, u_3, u_4\}$ where $u_1 = a\diamond b$, $u_2 = aa\diamond bba$, $u_3 = \diamond b$, and $u_4 = ba$ is not a pcode over $\{a, b\}$ since $u_1u_3u_3u_4u_3 \uparrow u_2u_3u_1$ is a nontrivial compatibility relation over Z.

Definition 1 has immediate consequences that should be emphasized:

- A pcode is always a pairwise non compatible set, but the converse is false.
- If X is a pcode over A, then $X_\diamond = \{u_\diamond \mid u \in X\}$ is a code over $A \cup \{\diamond\}$. But the converse does not hold. Consider, for instance, the code $X_\diamond = \{a\diamond a, \diamond a\diamond\}$ over $\{a, \diamond\}$. The underlying set X is not a pcode over $\{a\}$ since its elements are compatible. This fact is important, since it justifies the study of pcodes.

For a monoid M, we call a morphism $\varphi : M \to W(A)$ *pinjective* if for all $m, m' \in M$, $\varphi(m) \uparrow \varphi(m')$ implies $m = m'$. The definition of a pcode can be rephrased according to the following proposition.

Proposition 1 *If a subset X of $W(A)$ is a pcode over A, then a morphism $\varphi : B^* \to W(A)$ which induces a bijection of some alphabet B onto X is pinjective. Conversely, if there exists a pinjective morphism $\varphi : B^* \to W(A)$ such that $X = \varphi(B)$, then X is a pcode over A.*

For an alphabet B, a morphism $\varphi : B^* \to W(A)$ which is pinjective and satisfies $X = \varphi(B)$ is called a *pcoding morphism* for X. For any pcode $X \subset W(A)$, the existence of a pcoding morphism for X is straightforward: it suffices to take any bijection of a set B onto X and to extend it to a morphism from B^* into $W(A)$.

We now introduce the class of *prefix pcodes*. Let X be a subset of $W(A)$. Then X is called a *prefix pcode* if for all $u, v \in X$,

$$ux \uparrow v \text{ for some } x \in W(A) \text{ implies } u = v.$$

It is immediate that a singleton set is a prefix pcode and any subset of a prefix pcode is a prefix pcode. Hence any intersection of prefix pcodes is also a prefix pcode. A prefix pcode is a pcode. Note that a pcode may not be a prefix pcode (take for example $X = \{a\diamond b, a\diamond\}$). *Suffix pcodes* are defined in a symmetric way. Clearly, a set of partial words X is a suffix pcode if and only if $rev(X)$ is a prefix pcode. *Biprefix pcodes* are pcodes that are both prefix and suffix.

If n is a positive integer, then a largest pairwise non compatible set X satisfying $X \subset (A \cup \{\diamond\})^n$ is a biprefix pcode called a *uniform pcode* of partial words of length n. By largest we mean that if u is a partial word of length n over A, then there exists $v \in X$ such that $u \uparrow v$. Such a uniform pcode is maximal over A (a pcode is called *maximal* over A if it is not a proper subset of any other pcode over A).

The following proposition follows as for codes over A.

Proposition 2 *Any pcode X over A is contained in some maximal pcode over A.*

4 Pcodes and Monoids

We can prove that a set X is a pcode by knowing the submonoid X^* of $W(A)$ it generates. In particular, X is a pcode (respectively, prefix pcode, suffix pcode, biprefix pcode) if and only if X^* is a *pfree* monoid (respectively, *right unitary* monoid, *left unitary* monoid, *biunitary* monoid).

Proposition 3 *If M is a submonoid of $W(A)$, then the set $X = (M \setminus \{\epsilon\}) \setminus (M \setminus \{\epsilon\})^2$ is the unique minimal set that generates M.*

We call a submonoid M of $W(A)$ *pfree* if there exists a morphism $\varphi : B^* \to M$ of a free monoid B^* onto M that satisfies

$$\varphi(x) \uparrow \varphi(y) \text{ implies } x = y.$$

For instance, for any partial word $u \in W(A) \setminus \{\epsilon\}$, the submonoid generated by u is pfree.

Proposition 4 *If M is a pfree submonoid of $W(A)$, then its minimal generating set is a pcode. Conversely, if $X \subset W(A)$ is a pcode, then the submonoid X^* of $W(A)$ is pfree and X is its minimal generating set.*

We call the pcode X which generates a pfree submonoid M of $W(A)$ the *base* of M.

Corollary 1 *Let X and Y be pcodes over A. If $X^* = Y^*$, then $X = Y$.*

The set $X = \{a, \diamond b, a\diamond b\}$ is not a pcode over $\{a, b\}$ since it is not the minimal generating set of X^*. The set $Y = \{x, y\}$ where $x = \diamond bb$ and $y = abb\diamond$ is the minimal generating set of Y^*, yet Y is not a pcode over $\{a, b\}$ because $xy \uparrow yx$ is a nontrivial compatibility relation over Y. Here Y^* is not pfree.

Proposition 5 gives a characterization of a pfree submonoid of $W(A)$ that does not depend on its base. This proposition can be used to show that a submonoid is pfree (and consequently that its base is a pcode) without knowing its base. We call a submonoid M of $W(A)$ *stable* (in $W(A)$) if for all $u, u', v, w \in W(A)$ with $u \uparrow u'$, the conditions $u, u'w, v \in M$ and $wv \in C(M)$ imply $u = u'$ and $w \in M$.

Proposition 5 *A submonoid M of $W(A)$ is stable if and only if it is pfree.*

Note that although the monoid A^* is stable, the monoid $W(A)$ is not stable (and hence not pfree).

Returning to the above example, the set $Y^* = \{x, y\}^*$ where $x = \diamond bb$ and $y = abb\diamond$ is not pfree, which can be seen by using Proposition 5. Indeed, Y^* is not stable by setting $u = \diamond bb$, $u' = abb$, $v = \epsilon$, and $w = \diamond\diamond bb$ in the definition of stability.

Let M be a submonoid of $W(A)$. Then we call M *right unitary* (in $W(A)$) if for all $u, u', v \in W(A)$ with $u \uparrow u'$, the conditions $u, u'v \in M$ imply $u = u'$ and $v \in M$. Symmetrically, we call M *left unitary* (in $W(A)$) if for all $u, u', v \in W(A)$ with $u \uparrow u'$, the conditions $u, vu' \in M$ imply $u = u'$ and $v \in M$. The submonoid M is *biunitary* if it is both left and right unitary.

Proposition 6 *Let M be a submonoid of $W(A)$ and let X be its minimal generating set. Then M is right unitary (respectively, left unitary, biunitary) if and only if X is a prefix (respectively, suffix, biprefix) pcode. In particular, a right unitary (left unitary, biunitary) submonoid of $W(A)$ is pfree.*

Proposition 7 *An intersection of pfree submonoids of $W(A)$ is a pfree submonoid of $W(A)$.*

If X is a subset of $W(A)$, the set $\mathbf{M}(X)$ of pfree submonoids of $W(A)$ containing X may be empty. If $\mathbf{M}(X) \neq \emptyset$, then X is pairwise non compatible. If $\mathbf{M}(X)$ is not empty, then we call the intersection of all elements of $\mathbf{M}(X)$, which is the smallest pfree submonoid of $W(A)$ containing X by Proposition 7, the *pfree hull* of X. If X^* is a pfree submonoid of $W(A)$, then X^* coincides with its pfree hull.

Proposition 8 *Let $X \subset W(A)$ be such that $\mathbf{M}(X) \neq \emptyset$. Let Y be the base of the pfree hull of X. Then*

$$Y \subset \{u \mid uy \in X \text{ for some } y \in Y^*\} \cap \{u \mid yu \in X \text{ for some } y \in Y^*\}.$$

Proof. We show that $Y \subset \{u \mid yu \in X \text{ for some } y \in Y^*\}$. Suppose there exists $v \in Y$ such that $v \notin \{u \mid yu \in X \text{ for some } y \in Y^*\}$. Then $X \subset \{\epsilon\} \cup Y^*(Y \setminus \{v\})$. Let Z be defined by $v^*(Y \setminus \{v\})$. We have $Z^+ = Y^*(Y \setminus \{v\})$, and thus $X \subset Z^*$. Now Z is a pcode. Indeed, a compatibility relation $u_1u_2 \dots u_m \uparrow v_1v_2 \dots v_n$ where m, n are positive integers and $u_1, \dots, u_m, v_1, \dots, v_n \in Z$ can be rewritten as

$$v^{k_1}y_1v^{k_2}y_2 \dots v^{k_m}y_m \uparrow v^{\ell_1}z_1v^{\ell_2}z_2 \dots v^{\ell_n}z_n$$

with $u_i = v^{k_i}y_i$, $v_j = v^{\ell_j}z_j$, $y_i \in Y \setminus \{v\}$, $z_j \in Y \setminus \{v\}$, $k_i \geq 0$, $\ell_j \geq 0$ for $i = 1, \dots, m$ and $j = 1, \dots, n$. Since Y is a pcode, we get $k_1 = \ell_1, y_1 = z_1, k_2 = \ell_2, y_2 = z_2, \dots$, and finally $m = n$ and $u_i = v_i$ for $i = 1, \dots, m$. Thus, the set Z^* is a pfree submonoid of $W(A)$ containing X. But we have $Z^* \sqsubset Y^*$, which contradicts the minimality of the pfree submonoid Y^*. We similarly show that $Y \subset \{u \mid uy \in X \text{ for some } y \in Y^*\}$. □

The following result extends the well-known *Defect Theorem* on words to partial words.

Theorem 1 *Let X be a finite subset of $W(A)$ such that $\mathbf{M}(X) \neq \emptyset$. Let Y be the base of the pfree hull of X. If X is not a pcode, then $\|Y\| < \|X\|$.*

5 An Algorithm for Pcodes

In this section, we give an algorithm to test whether or not a finite set is a pcode.

Let X be a subset of $W(A) \setminus \{\epsilon\}$. Let

$$U_1 = \{x \mid x \in W(A) \setminus \{\epsilon\} \text{ and there exists } u \in C(X) \text{ such that } ux \in X\},$$

and for $i \geq 1$, let

$$U_{i+1} = \{x \mid x \in W(A) \text{ and there exists } u \in C(X) \text{ such that } ux \in U_i\} \cup \{x \mid x \in W(A) \text{ and there exist } y \in U_i, y' \in W(A) \text{ such that } y \uparrow y' \text{ and } y'x \in X\}.$$

Lemma 4 *Let $X \subset W(A) \setminus \{\epsilon\}$. For all $n \geq 1$ and $k \in \{1, \ldots, n\}$, we have $\epsilon \in U_n$ if and only if there exist a partial word $x \in U_k$ and integers $i, j \geq 0$ such that $xX^i \cap C(X^j) \neq \emptyset$ and $i + j + k = n$.*

Proof. We prove the statement for all n by descending induction on k. Assume first that $k = n$. If $\epsilon \in U_n$, then the condition is satisfied with $x = \epsilon$ and $i = j = 0$. Conversely, if the condition is satisfied, then $i = j = 0$ and $x = \epsilon$ and $\epsilon \in U_n$.

Now, let $n > k \geq 1$, and suppose that the equivalence holds for $n, n-1, \ldots, k+1$. If $\epsilon \in U_n$, then by the inductive hypothesis, there exist a partial word $x \in U_{k+1}$ and integers $i, j \geq 0$ such that $xX^i \cap C(X^j) \neq \emptyset$ and $i + j + (k+1) = n$. Thus there exist partial words $u_1, \ldots, u_i, v_1, \ldots, v_j \in X$ such that

$$xu_1 \ldots u_i \uparrow v_1 \ldots v_j.$$

Now $x \in U_{k+1}$, and there are two cases: Either there exists $u \in C(X)$ such that $ux \in U_k$, or there exists $y \in U_k$ and a partial word y' such that $y \uparrow y'$ and $y'x \in X$. In the first case, we have $u \uparrow u'$ for some $u' \in X$ and

$$uxu_1 \ldots u_i \uparrow u'v_1 \ldots v_j.$$

Consequently, there exist a partial word $ux \in U_k$ and integers $i, j+1 \geq 0$ such that $uxX^i \cap C(X^{j+1}) \neq \emptyset$ and $i + (j+1) + k = n$, and the condition is satisfied. In the second case, we have

$$y'xu_1 \ldots u_i \uparrow yv_1 \ldots v_j.$$

Consequently, there exist a partial word $y \in U_k$ and integers $j, i+1 \geq 0$ such that $yX^j \cap C(X^{i+1}) \neq \emptyset$ and $j + (i+1) + k = n$, and the condition is satisfied.

Conversely, assume that there exist a partial word $x \in U_k$ and integers $i, j \geq 0$ such that $xX^i \cap C(X^j) \neq \emptyset$ and $i + j + k = n$. Then

$$xu_1 \dots u_i \uparrow v_1 \dots v_j$$

for some $u_1, \dots, u_i, v_1, \dots, v_j \in X$. If $j = 0$, then $i = 0$ and $k = n$. If $j > 0$, then we consider two cases:

Case 1. $|x| \geq |v_1|$

If $x = v_1'y$ for some $y \in W(A)$ and some v_1' satisfying $v_1' \uparrow v_1$, then $y \in U_{k+1}$ and $yu_1 \dots u_i \uparrow v_2 \dots v_j$. Thus $yX^i \cap C(X^{j-1}) \neq \emptyset$ and by the inductive hypothesis $\epsilon \in U_n$.

Case 2. $|x| < |v_1|$

If $v_1 = x'y$ for some $y \in W(A) \setminus \{\epsilon\}$ and some x' satisfying $x \uparrow x'$, then $y \in U_{k+1}$ and $u_1 \dots u_i \uparrow yv_2 \dots v_j$. Thus $yX^{j-1} \cap C(X^i) \neq \emptyset$ and by the inductive hypothesis $\epsilon \in U_n$. □

Note that if X is a finite set, then $\{U_n \mid n \geq 1\}$ is finite (this is because each U_n contains only suffixes of partial words in X). The next theorem hence provides an algorithm for testing whether or not a finite set is a pcode. Note that the algorithm ends immediately for prefix pcodes since $U_1 = \emptyset$ for such sets.

Theorem 2 *Let $X \subset W(A) \setminus \{\epsilon\}$ be pairwise non compatible. The set X is a pcode if and only if none of the sets U_n contains the empty word.*

Proof. If X is not a pcode, then there exists a compatibility relation

$$u_1u_2 \dots u_m \uparrow v_1v_2 \dots v_n,$$

where m, n are positive integers, $u_1 \neq v_1$, and $u_1, \dots, u_m, v_1, \dots, v_n \in X$. Assume first that $|u_1| = |v_1|$. Then $u_1 \uparrow v_1$, a contradiction since X is pairwise non compatible. Now assume that $|u_1| > |v_1|$. Then $u_1 = v_1'x$ for some $x \in W(A) \setminus \{\epsilon\}$ and some v_1' satisfying $v_1' \uparrow v_1$. But then $x \in U_1$ and $xX^{m-1} \cap C(X^{n-1}) \neq \emptyset$. By Lemma 4, $\epsilon \in U_{m+n-1}$.

If X is a pcode and $\epsilon \in U_n$, then put $k = 1$ in Lemma 4. There exist $x \in U_1$ and integers $i, j \geq 0$ such that $i + j = n - 1$ and $xX^i \cap C(X^j) \neq \emptyset$. Since $x \in U_1$, we have $v = ux$ for some $u \in C(X), v \in X$. Furthermore, $u \neq v$ since $x \neq \epsilon$. Since $u \in C(X)$, there exists $u' \in X$ such that $u \uparrow u'$. It follows from $uxX^i \cap uC(X^j) \neq \emptyset$ that $vX^i \cap C(u'X^j) \neq \emptyset$, showing that X is not a pcode. □

As an example, let us consider the set X consisting of the partial words u_1, u_2, u_3, u_4 with $u_1 = a\diamond b$, $u_2 = aa\diamond bba$, $u_3 = \diamond b$, and $u_4 = ba$. This set is not a pcode since $u_2u_3u_1 \uparrow u_1u_3u_3u_4u_3$. In an attempt to discover this nontrivial compatibility relation, we give the following list of compatibilities

$$\begin{array}{ccc} u_2 & \uparrow & u_1\underline{bba} \\ u_2 & \uparrow & u_1u_3\underline{a} \\ u_2\underline{b} & \uparrow & u_1u_3u_3 \\ u_2u_3 & \uparrow & u_1u_3u_3\underline{b} \\ u_2u_3\underline{a} & \uparrow & u_1u_3u_3u_4 \\ u_2u_3u_1 & \uparrow & u_1u_3u_3u_4u_3 \end{array}$$

The underlined word bba is in U_1, next $a \in U_2$, then $b \in U_3$ and $b \in U_4$, finally $a \in U_5$ and $\epsilon \in U_6$. More precisely, we obtain

$U_1 = \{b, bba\}$,

$U_2 = \{a, b\}$,

$U_3 = \{a, b, \diamond b, a\diamond bba\}$,

$U_4 = \{\epsilon, a, b, \diamond b, ba, bba, a\diamond bba\}$,

$U_5 = \{\epsilon, a, b, \diamond b, ba, a\diamond b, bba, a\diamond bba, aa\diamond bba\} = U_6 = U_7 = \cdots$.

Since $\epsilon \in U_4$, the set X is not a pcode by Theorem 2 (see `http://www.uncg.edu/mat/pcode/`).

6 Complete Pcodes

Let X be a subset of $W(A)$. A partial word u over A is *completable in* X if there exist $x, y \in W(A)$ such that $xuy \in C(X)$. It is equivalent to saying that $W(A)uW(A) \cap C(X) \neq \emptyset$, or, in other words, that $u \in F(C(X))$. The set X is *dense* if all elements of $W(A)$ are completable in X, or equivalently $F(C(X)) = W(A)$. Clearly, each superset of a dense set is dense. The set X is *complete* if X^* is dense. Every dense set is also complete.

The proof that a maximal pcode is complete is based on Proposition 9 which describes a method for embedding any pcode in a complete pcode.

Proposition 9 *Let $X \subset W(A)\setminus\{\epsilon\}$ be a pcode, let u be an unbordered word over A such that $u \notin F(C(X^*))$, let U be a largest pairwise non compatible subset of $W(A) \setminus C(W(A)uW(A))$ containing X^*, and let $Y = U \setminus X^*$. Then the set*

$$Z = X \cup \{uy_1u \dots y_nu \mid y_1, \dots, y_n \in Y \text{ and } n \geq 0\}$$

is a complete pcode.

Proof. First, let us show that the set $V = Uu$ is a prefix pcode. To see this, suppose that $vux \uparrow v'u$ for two partial words $v, v' \in U$ and some $x \in W(A)$. If $|vu| > |v'|$, then $vu \uparrow v'y$ with $u = yz$ for some y, z. We deduce that $yz \uparrow z'y$ for some z'. If $z = \epsilon$, then $vu \uparrow v'u$ and $v \uparrow v'$. Since U is pairwise non compatible, we have $v = v'$. If $z \neq \epsilon$, then since y is full, by Lemma 2, there exist words x', y' such that $z' \subset x'y'$, $z \subset y'x'$, and $y \subset (x'y')^n x'$ for some integer $n \geq 0$. But then $u \subset (x'y')^{n+1}x'$, and since u is unbordered, $x' = \epsilon$. If $n > 0$, u is bordered, and if $n = 0$, we get $y = \epsilon$ and so $vu \uparrow v'$. This leads to $v' \in C(W(A)uW(A))$, which is a contradiction. Hence $|vu| \leq |v'|$, and $vuy \uparrow v'$ for some y. But then again v' is in $C(W(A)uW(A))$, a contradiction.

Next, we show that Z is a pcode. Assume the contrary and consider a relation

$$u_1 u_2 \dots u_m \uparrow v_1 v_2 \dots v_n$$

with $u_1, \dots, u_m, v_1, \dots, v_n \in Z$, and $u_1 \neq v_1$. The set X being a pcode, one of these partial words must be in $Z \setminus X$. Assume that one of $u_1, \dots, u_m$ is in $Z \setminus X$, and let i be the smallest index such that u_i matches $u(Yu)^*$. Since $W(A)uW(A) \cap C(X^*) = \emptyset$, it follows that $W(A)u_iW(A) \cap C(X^*) = \emptyset$. Consequently one of $v_1, \dots, v_n$ matches $u(Yu)^*$. Let j be the smallest index such that v_j matches $u(Yu)^*$. Then $u_1 \dots u_{i-1}u, v_1 \dots v_{j-1}u \in V$ whence $u_1 \dots u_{i-1} = v_1 \dots v_{j-1}$ since V is a prefix pcode. The set X being a pcode, thus from $u_1 \neq v_1$ it follows that $i = j = 1$. Put

$$u_1 = uy_1u \dots uy_ku,$$
$$v_1 = uy'_1u \dots uy'_\ell u,$$

with $y_1, \dots, y_k, y'_1, \dots, y'_\ell \in Y$. If $|u_1| = |v_1|$, then $u_1 \uparrow v_1$. Since X is a pcode, we get $u_1 = v_1$, a contradiction. So assume that $|u_1| < |v_1|$. Since V is a prefix pcode, the set V^* is right unitary. Since $Y \subset U$, each y_iu, y'_iu is in V. Consequently $y_1 = y'_1, \dots, y_k = y'_k$. Put $w = y'_{k+1}u \dots uy'_\ell u$. We have $u_2 \dots u_m \uparrow wv_2 \dots v_n$ with $w \in V^*$. The word u is a factor of w, and thus occurs also in $u_2 \dots u_m$. This shows that one of $u_2, \dots, u_m$, say u_r, matches $u(Yu)^*$. Suppose r is chosen minimal. Then $u_2 \dots u_{r-1}u \in V$ and $y'_{k+1}u \in V$, and with the set V being a prefix pcode, we have $y'_{k+1} = u_2 \dots u_{r-1}$. Thus $y'_{k+1} \in X^*$, a contradiction with the fact that $y'_{k+1} \in Y$.

Last, let us show that Z is complete. Let $w \in W(A)$. If $w \in C(W(A)uW(A))$, then

$$w \uparrow u_1uu_2u \dots uu_{n-1}uu_n$$

for some positive integer n and some partial words $u_1, \ldots, u_n \in U$. If $w \notin C(W(A)uW(A))$, then $w \in U$ or $w \in C(U)$, and the abovementioned compatibility relation holds. In any case, $uwu \in C(Z^*)$ and so $w \in F(C(Z^*))$. To see this, let $u_{i_1}, u_{i_2}, \ldots, u_{i_k}$ be those u_i's in X^*. Then

$$uwu \uparrow (uu_1u \ldots uu_{i_1-1}u)u_{i_1}\ (uu_{i_1+1}u \ldots uu_{i_2-1}u)u_{i_2} \ldots u_{i_k}(uu_{i_k+1}u \ldots uu_nu).$$

The parenthesized partial words are in Z and the result follows. □

Theorem 3 ([5]) *Let $X \subset W(A) \setminus \{\epsilon\}$. If X is a maximal pcode, then X is complete.*

7 Circular Pcodes

In this section, we introduce a particular class of pcodes called *circular pcodes*. Circular pcodes turn out to have numerous interesting properties.

Let X be a subset of $W(A)\setminus\{\epsilon\}$. Then X is called a *circular pcode* over A if for all integers $m, n \geq 1$, partial words $u_1, \ldots, u_m, v_1, \ldots, v_n \in X$, and $r \in W(A)$ and $s \in W(A) \setminus \{\epsilon\}$, the conditions

$$su_2u_3 \ldots u_mr \uparrow v_1v_2 \ldots v_n,$$
$$u_1 \subset rs,$$

imply $m = n$, $r = \epsilon$, and $u_i = v_i$ for $i = 1, \ldots, m$. A subset X of A^+ is a circular code if and only if it is a circular pcode. Also, any subset of a circular pcode is a circular pcode. Using arguments as in [5], we can show that a circular pcode does not contain two distinct conjugate partial words, and also that elements of a circular pcode are primitive.

We now characterize in various ways the submonoids generated by circular pcodes.

A submonoid M of $W(A)$ is called *pure* if for all $u \in W(A)$ and integer $n \geq 1$, the conditions $u_1 \ldots u_n \in M$ and $u_i \subset u$ for all $i = 1, \ldots, n$ imply $u_1 = u_2 = \cdots = u_n$ and $u_i \in M$ for all $i = 1, \ldots, n$. A submonoid M of $W(A)$ is called *very pure* if for all $u, v, u', v' \in W(A)$ satisfying $|v'| = |v|$ and $|u'| = |u|$, the conditions $vu \uparrow v'u'$, $uv \in M$, and $v'u' \in M$ imply $u = u'$ and $u, v \in M$. A very pure monoid is pure.

Proposition 10 *A submonoid M of $W(A)$ is very pure if and only if its minimal generating set is a circular pcode.*

Proof. Let M be a very pure submonoid of $W(A)$. Let $u, u', v, w \in W(A)$ with $u \uparrow u'$, $u, u'w, v \in M$, and $wv \in C(M)$. We have $(vu')w = v(u'w) \in M$ and $w(vu') = (wv)u' \in C(M)$. This implies $u = u'$ and $w \in M$. Thus M is stable, hence M is pfree by Proposition 5. Let X be its base. Assume that there exist positive integers m, n, partial words $u_1, \ldots, u_m, v_1, \ldots, v_n \in X$, and $r \in W(A)$ and $s \in W(A) \setminus \{\epsilon\}$ such that

$$su_2u_3 \ldots u_m r \uparrow v_1v_2 \ldots v_n,$$
$$u_1 \subset rs.$$

Put $u_1 = r's'$ where $|r'| = |r|$ and $|s'| = |s|$. Put $u = s'$ and $v = u_2 \ldots u_m r'$. Then $vu \in M$ and by weakening $uv \in C(M)$. Since M is very pure, $u, v \in M$. Since $u_2 \ldots u_m, u_2 \ldots u_m r', s', r's' \in M$, the stability of M implies that $r' \in M$. From $r's' \in X$, it follows that $r' = \epsilon$ (and $r = \epsilon$). By weakening, $u_1u_2 \ldots u_m \uparrow v_1v_2 \ldots v_n$. Since X is a pcode by Proposition 4, this implies $m = n$ and $u_i = v_i$ for $i = 1, \ldots, m$.

Conversely, let X be the minimal generating set for M and assume that X is a circular pcode (here $M = X^*$). To show that M is very pure, consider $u, v \in W(A)$ such that $uv \in M$ and $vu \in C(M)$. The latter implies that $vu \uparrow v'u'$ with u', v' satisfying $v'u' \in M$, $|v'| = |v|$, and $|u'| = |u|$. If $u = \epsilon$ or $v = \epsilon$, then $u = u'$ and $u, v \in M$. If $u \neq \epsilon$ and $v \neq \epsilon$, then put

$$uv = u_1u_2 \ldots u_m,$$
$$vu \uparrow v_1v_2 \ldots v_n$$

with $u_1, \ldots, u_m, v_1, \ldots, v_n \in X$. There exists an integer i, $1 \leq i \leq m$, such that

$$u = u_1u_2 \ldots u_{i-1}r,$$
$$v = su_{i+1} \ldots u_m$$

where $u_i = rs$, $r \in W(A)$, and $s \in W(A) \setminus \{\epsilon\}$. Then

$$su_{i+1} \ldots u_m u_1u_2 \ldots u_{i-1}r \uparrow v_1v_2 \ldots v_n.$$

Since X is a circular pcode, this implies $m = n$, $r = \epsilon$, and $u_i = v_1, u_{i+1} = v_2, \ldots, u_m = v_{n-i+1}, u_1 = v_{n-i+2}, u_2 = v_{n-i+3}, \ldots, u_{i-1} = v_n$. Thus $u = u_1u_2 \ldots u_{i-1} = v_{n-i+2}v_{n-i+3} \ldots v_n = u'$ and $u, v \in M$, showing that M is very pure. □

We now give a characterization of circular pcodes in terms of conjugacy. Let $X \subset W(A) \setminus \{\epsilon\}$ be a pcode. Two partial words $u, v \in X^*$ are called *X–conjugate* if there exist

$x, y \in X^*$ such that $u = xy, v = yx$. Of course, two partial words in X^* which are X–conjugate are conjugate.

Proposition 11 *Let $X \subset W(A) \setminus \{\epsilon\}$ be a pcode. The following conditions are equivalent:*

1. *The set X is a circular pcode.*
2. *The monoid X^* is pure, and any two partial words in X^* which are conjugate are also X-conjugate.*

Proof. We first show that Condition 1 implies Condition 2. Since X^* is very pure, it is pure. Next, let $u, v \in X^*$ be conjugate partial words. Then $u \subset xy, v \subset yx$ for some $x, y \in W(A)$. Put $u = x'y'$ where $|x'| = |x|$ and $|y'| = |y|$. Also, put $v = y''x''$ where $|y''| = |y|$ and $|x''| = |x|$. Since $x' \subset x$ and $y' \subset y$, we get $y'x' \subset yx$. The latter and the fact that $y''x'' \subset yx$ imply that $y'x' \uparrow y''x''$. We get the two conditions $x'y' \in X^*$ and $y'x' \in C(X^*)$. Since X^* is very pure, $x' = x''$ and $x', y' \in X^*$. With a similar reasoning, we can deduce that $y' = y''$. So $u = x'y', v = y'x'$ with $x', y' \in X^*$, showing that u, v are X-conjugate.

Now, we show that Condition 2 implies Condition 1. Let $u, v \in W(A)$ be such that $uv \in X^*$ and $vu \in C(X^*)$. The latter implies that $vu \uparrow v'u'$ with u', v' satisfying $v'u' \in X^*$, $|v'| = |v|$ and $|u'| = |u|$. If $u = \epsilon$, then $u = u'$ and $u, v \in X^*$. If $u \neq \epsilon$ and $v = \epsilon$, then $u, u', v \in X^*$ and $u \uparrow u'$. Since X is a pcode, this yields $u = u'$. If $u \neq \epsilon$ and $v \neq \epsilon$, then by definition, there exists a primitive word x and a positive integer n such that $vu \subset x^n$ and $v'u' \subset x^n$. We get words r, s and integers p, q such that $x = rs$, $u \subset sx^q$, $v \subset x^p r$, and $p + q + 1 = n$. Put $y = sr$ (x being primitive, y is primitive as well). Since $v'u' \subset x^n$ and $uv \subset y^n$, write $v'u' = x_1 x_2 \dots x_n$ and $uv = y_1 y_2 \dots y_n$ where $|x_1| = |x_2| = \cdots = |x_n|$, $|y_1| = |y_2| = \cdots = |y_n|$. Since X^* is pure, we have $x_1 = x_2 = \cdots = x_n$, $y_1 = y_2 = \cdots = y_n$, and $x_1, \dots, x_n, y_1, \dots, y_n \in X^*$. Thus $v'u' = (x')^n$ and $uv = (y')^n$ with $x', y' \in X^*$. Since $x' \subset x \subset rs, y' \subset y \subset sr$, we get that x', y' are conjugate and thus X-conjugate. So there exist $r', s' \in X^*$ such that $x' = r's'$ and $y' = s'r'$. Thus $u = s'(x')^q$, $u' = s'(x')^q$, and $v = (x')^p r'$ showing that $u = u'$ and $u, v \in X^*$. □

The following result is an analogue of Theorem 3.

Proposition 12 *Let $X \subset W(A) \setminus \{\epsilon\}$ be a circular pcode. If X is maximal as a circular pcode, then X is complete.*

8 Two-Element Pcodes

In [5], it was shown that if the two-element set $\{u, v\}$ belongs to the so-called type 1 or type 2 sets of partial words, then $\{u, v\}$ is a pcode if and only if $uv \not\uparrow vu$. In this section, we extend this result further. Note that the above equivalence is not true in general: the set $\{u, v\}$ where $u = a\diamond b$ and $v = abbaab$ satisfies $uv \not\uparrow vu$, but $\{u, v\}$ is not a pcode since $u^2 \uparrow v$.

The following proposition holds.

Proposition 13 ([5]) *Let X be a subset of $W(A)$. Then X is a pcode over A if and only if for every integer $n \geq 1$ and partial words $u_1, \ldots, u_n, v_1, \ldots, v_n \in X$, the compatibility relation*

$$u_1 u_2 \ldots u_n \uparrow v_1 v_2 \ldots v_n$$

implies $u_i = v_i$ for $i = 1, \ldots, n$.

We now assume that $\{u, v\}$ is a set of partial words over an alphabet of size at least two. Otherwise, sets of at least two partial words are obviously non pcodes.

Proposition 14 *Let k be an integer satisfying $k > 1$. Let $u, v \in W(A) \setminus \{\epsilon\}$ be such that $|v| = k|u|$ and $\|H(v)\| = 0$. Then $\{u, v\}$ is a pcode if and only if $u^k v \not\uparrow vu^k$.*

Proof. If $\{u, v\}$ is a pcode, then clearly $u^k v \not\uparrow vu^k$. Conversely, assume that $\{u, v\}$ is not a pcode and $u^k v \not\uparrow vu^k$. Then there exist $n \geq 1$ and $u_1, \ldots, u_n, v_1, \ldots, v_n \in \{u, v\}$ such that

$$u_1 u_2 \ldots u_n \uparrow v_1 v_2 \ldots v_n \tag{1}$$

and with $|u_1 u_2 \ldots u_n|$ as small as possible contradicting Proposition 13. We hence have $u_1 \neq v_1$ and $u_n \neq v_n$, and we may assume that $n \geq 2$. There are four possibilities: $u_1 = u_n = u$, $v_1 = v_n = v$; $u_1 = v_n = u$, $v_1 = u_n = v$; $u_1 = v_n = v$, $v_1 = u_n = u$; and $u_1 = u_n = v$, $v_1 = v_n = u$. In all cases, put $u_2 \ldots u_{n-1} = x$ and $v_2 \ldots v_{n-1} = y$. These possibilities can be rewritten as (a) $uxu \uparrow vyv$; (b) $uxv \uparrow vyu$; (c) $vxu \uparrow uyv$; and (d) $vxv \uparrow uyu$. Since $|v| = k|u|$, for any of the possibilities (a)–(d), there exist $w_1, w_2, \ldots, w_k \in W(A) \setminus \{\epsilon\}$ such that $v = w_1 w_2 \ldots w_k$, $|w_1| = |w_2| = \cdots = |w_k| = |u|$, $w_1 \uparrow u$, and $w_k \uparrow u$. The latter two relations give $u \subset w_1$ and $u \subset w_k$ since v is full.

Let us consider the case where $u_1 = u$ and $v_1 = v$ (the other cases are handled similarly).

Case 1. $u_1 = u_2 = \cdots = u_{k-1} = u$.

In this case $w_1 \uparrow u$, $w_2 \uparrow u$, ..., $w_k \uparrow u$, and by multiplication, $u^k v \uparrow vu^k$, contradicting our assumption.

Case 2. There exists $1 < j < k$ such that $u_1 = u_2 = \cdots = u_{j-1} = u$ and $u_j = v$.

Note that each element in $\{w_1, w_2, \ldots, w_{j-1}\}$ is compatible with u. Here, $k = m(j-1)+r$ with $1 \leq r < j$. We get $w_{k-j+2} = w_{k-2j+3} = \cdots =$ element in the set $\{w_1, w_2, \ldots, w_{j-1}\}$, $w_{k-j+3} = w_{k-2j+4} = \cdots =$ element in the set $\{w_1, w_2, \ldots, w_{j-1}\}$, ..., and $w_k = w_{k-j+1} = \cdots =$ element in the set $\{w_1, w_2, \ldots, w_{j-1}\}$. Thus, $w_1 \uparrow u$, $w_2 \uparrow u$, ..., $w_k \uparrow u$. Hence $u^k v \uparrow vu^k$, contradicting our assumption. □

Proposition 15 *Let k, ℓ be integers satisfying $1 \leq k \leq \ell$. Let $u, v, w, w_1, \ldots, w_k \in W(A) \setminus \{\epsilon\}$ be such that $u = w_1 w_2 \ldots w_k$, $v = w^\ell$, and $|w_1| = |w_2| = \cdots = |w_k| = |w|$. Then $\{u, v\}$ is a pcode if and only if $uv \not\uparrow vu$.*

Proof. We refer the reader to the proof of Proposition 14. Any of the possibilities (a)–(d) imply $w_1 \uparrow w$, $w_2 \uparrow w$, ..., $w_k \uparrow w$. Hence $uv \uparrow vu$, which contradicts our assumption. □

Proposition 16 *Let k be an integer satisfying $k > 1$. Let $u, v, w_1, w_2, \ldots, w_k \in W(A) \setminus \{\epsilon\}$ be such that $v = w_1 w_2 \ldots w_k$, $|w_1| = |w_2| = \cdots = |w_k| = |u|$, $\|H(u)\| = 0$, and $\|H(v)\| = \|H(w_i)\|$ for some $1 \leq i \leq k$. Then $\{u, v\}$ is a pcode if and only if $uv \not\uparrow vu$.*

Proof. We refer the reader to the proof of Proposition 14. Any of the possibilities (a)–(d) imply $w_1 \uparrow u$ and $w_k \uparrow u$. The latter two relations give $w_1 \subset u$ and $w_k \subset u$ since u is full. Let us consider the case where $u_1 = u$ and $v_1 = v$ (the other cases are handled similarly).

Case 1. $u_1 = u_2 = \cdots = u_n = u$.

In this case $w_1 \uparrow u$, $w_2 \uparrow u$, ..., $w_k \uparrow u$, and thus $uv \uparrow vu$ contradicting our assumption.

Case 2. There exists $1 < j \leq n$ such that $u_1 = u_2 = \cdots = u_{j-1} = u$ and $u_j = v$.

Here we consider the cases where $j \geq k$ and $j < k$. Note that each element in the set $\{w_1, w_2, \ldots, w_{j-1}\}$ is compatible with u. If $j \geq k$, then $w_1 \uparrow u$, $w_2 \uparrow u$, ..., $w_k \uparrow u$, and thus $uv \uparrow vu$ contradicting our assumption. Now, if $j < k$, then $k = m(j-1) + r$ and $i = m'(j-1) + r'$ with $1 \leq r < j$ and $1 \leq r' < j$. We get

$$w_i \subset w_{i-j+1} = w_{i-2j+2} = \cdots = w_{r'} \tag{2}$$

We also get

$$w_i \subset w_{i+j-1} = w_{i+2j-2} = \cdots = w_{k-j+1-r+r'} \tag{3}$$

if $r' > r$, and

$$w_i \subset w_{i+j-1} = w_{i+2j-2} = \cdots = w_{k-r+r'} \tag{4}$$

if $r' \leq r$. Moreover, if $1 \leq i' \leq k$ and $i' \not\equiv r' \bmod j-1$, then $w_{i'} = u$. We consider the following three cases.

Case 2.1. $j \leq i \leq k-j+1$

In this case, the compatibility relation (1) yields $w_1 = w_2 = \cdots = w_{j-1} = u = w_{k-j+2} = \cdots = w_{k-1} = w_k$. The relations (2)–(3)–(4) imply that $v = u^{i-1} w_i u^{k-i}$ with $w_i \subset u$. Hence $uv \uparrow vu$, contradicting our assumption.

Case 2.2. $1 \leq i < j$

Here $i = r'$ and $w_i \subset u$. Consider the case where $r' > r$ (the case where $r' \leq r$ is handled similarly). Referring to relations (1)–(3), we have $w_{k-j+1-r+r'} \uparrow u$ or $w_{k-j+1-r+r'} \uparrow w_s$ where $s \not\equiv r' \bmod j-1$. In either case, $w_{k-j+1-r+r'} \uparrow u$ and since u is full, we get $w_{k-j+1-r+r'} \subset u$. Again $uv \uparrow vu$, contradicting our assumption.

Case 2.3. $k-j+2 \leq i \leq k$

This case is symmetric to Case 2.2. □

References

[1] J. Berstel and D. Perrin, *Theory of Codes* (Academic Press, New York, 1985).

[2] J. Berstel and L. Boasson, Partial words and a theorem of Fine and Wilf, *Theoretical Computer Science* **218** (1999) 135–141.

[3] F. Blanchet-Sadri, On unique, multiset, and set decipherability of three-word codes, *IEEE Transactions on Information Theory* **47** (2001) 1745–1757.

[4] F. Blanchet-Sadri, Periodicity on partial words, *Computers and Mathematics with Applications* **47** (2004) 71–82.

[5] F. Blanchet-Sadri, Codes, orderings, and partial words, *Theoretical Computer Science* **329** (2004) 177–202.

[6] F. Blanchet-Sadri, Primitive partial words, *Discrete Applied Mathematics* **148** (2005) 195–213.

[7] F. Blanchet-Sadri and Ajay Chriscoe, Local periods and binary partial words: an algorithm, *Theoretical Computer Science* **314** (2004) 189–216.
`http://www.uncg.edu/mat/AlgBin/`

[8] F. Blanchet-Sadri and S. Duncan, Partial words and the critical factorization theorem, *Journal of Combinatorial Theory, Series A* **109** (2005) 221–245. `http://www.uncg.edu/mat/cft/`

[9] F. Blanchet-Sadri and Robert A. Hegstrom, Partial words and a theorem of Fine and Wilf revisited, *Theoretical Computer Science* **270** (2002) 401–419.

[10] F. Blanchet-Sadri and D.K. Luhmann, Conjugacy on partial words, *Theoretical Computer Science* **289** (2002) 297–312.

[11] C. Choffrut and J. Karhumäki, Combinatorics of Words, in G. Rozenberg and A. Salomaa (Eds.), *Handbook of Formal Languages, Vol. 1, Ch. 6* (Springer-Verlag, Berlin, 1997) 329–438.

[12] F. Guzmán, Decipherability of codes, *Journal of Pure and Applied Algebra* **141** (1999) 13–35.

[13] T. Head and A. Weber, Deciding multiset decipherability, *IEEE Transactions on Information Theory* **41** (1995) 291–297.

[14] M. Ito, H. Jürgensen, H.J. Shyr and G. Thierrin, Anti-commutative languages and n-codes, *Discrete Applied Mathematics* **24** (1989) 187–196.

[15] P. Leupold, Partial words: Results and Perspectives, preprint 2003.

[16] M. Lothaire, *Combinatorics on Words* (Addison-Wesley, Reading, MA, 1983).

[17] J. Setubal and J. Meidanis, *Introduction to Computational Molecular Biology* (PWS Publishing Company, Boston, MA, 1997).

[18] H.J. Shyr, *Free Monoids and Languages* (Hon Min Book Company, Taichung, Taiwan, 1991).