TAMKANG JOURNAL OF MATHEMATICS
Volume 29, Number 2, Summer 1998

SOME PROPERTIES OF MAXIMAL COMMA-FREE CODES

CHEN-MING FAN AND H. J. SHYR

Abstract. A code A over a finite alphabet X is a comma-free code if $A^2 \cap X^+AX^+ = \emptyset$, where X is a finite alphabet containing more than one letter. This paper is a study of some algebraic properties of finite maximal comma-free codes. We give several characterizations on two-element comma-free codes and maximal comma-free codes. Let $X_n^m = X^n \cup X^{n+1} \cup \cdots \cup X^m$. We prove that for $n \geq 3$, a maximal comma-free code in X^n is a maximal comma-free code in the region $X_1^m \cup X^n$, $m < n/2$. We also obtain that for $X = \{a, b\}$, a maximal comma-free code in X^3 is a maximal comma-free code; a maximal comma-free code in X^4 is a maximal comma-free code in X_1^4; for every $n \geq 4$, there is a maximal comma-free code in X^n which is not a maximal comma-free code.

1. Introduction and Definitions

In this paper we let X be a finite alphabet consisting of at least two letters. Let X^* be the free monoid generated by X. Every element of X^* is called a *word* and every subset of X^* is called a *language*. The empty word is denoted by 1 and we let $X^+ = X^* \backslash \{1\}$. The length of a word w is denoted by $lg(w)$. A word $f \in X^+$ is a *primitive word* if f is not a power of any other word. The set of all primitive words over X will be denoted by Q. The catenation of two languages A and B is the set $AB = \{xy | x \in A, y \in B\}$. A language $A \subseteq X^+$ is a code if for any $x_i, y_j \in A$, the condition $x_1x_2 \cdots x_n = y_1y_2 \cdots y_m$ always implies that $m = n$ and $x_i = y_i$, $i = 1, 2, \cdots, n$. Prefix codes, infix codes, comma-free codes and intercodes of index n are known codes. We give their definitions as follows: Let $A \subseteq X^+$.

(1) A is a prefix code if $A \cap AX^+ = \emptyset$.
(2) A is an infix code if for $x, y, u \in X^*, u \in A$ and $xuy \in A$ together imply that $xy = 1$.
(3) A is a comma-free code if $A^2 \cap X^+AX^+ = \emptyset$.
(4) A is an intercode of index n if $A^{n+1} \cap X^+A^nX^+ = \emptyset$.

An intercode of index 1 is a comma-free code. A comma-free code A is said to be a *maximal comma-free code* if for any $u \in Q \setminus A, A \cup \{u\}$ is not a comma-free code. ([2], [3]) If a prefix code is used, the decoding process has to be done starting from the very

Received September 17, 1996; revised January 8, 1997.
1991 *Mathematics Subject Classification.* 68R15,68Q45
Key words and phrases. Free monoid, code, primitive word, maximal comma-free code.
* This Research is supported by UGC grant No. F:8-13/87(SR-III).

left end. But for the comma-free code case, we can pick up a code word in the sequence formed by a comma-free code from anywhere in the sequence. Therefore, for the decoding preocess, the job can be done not only start by any end of the sequence but also can be done from anywhere inside the sequence. In this paper we study some properties of comma-free codes and maximal comma-free codes in a certain region K, where $K \subseteq X^+$. For convinence we let $Q^{(1)} = Q \cup \{1\}$ and for $n \geq 2$, $Q^{(n)} = \{f^n | \ f \in Q\}$. In the sequel we need the following known results (see [3]):

Lemma 1.1. *(1) If $x, y \in X^+$ and $xy = yx$, then x and y are powers of a common primitive word. (2) For any $n \geq 1$, $x, y \in X^+$, $xy \in Q^{(n)}$ if and only if $yx \in Q^{(n)}$.*

In order to develop our theory we need to define the following terms: In the last part of this paper, we need the concept of skeleton of a word and some sets L_p, L_s. Every $w \in X^+$ can be written in the form

$$w = a_1^{k_1} a_2^{k_2} \cdots a_n^{k_n},$$

where $k_i \geq 1$ and $a_i \neq a_{i+1} \in X, i = 1, 2, \cdots, n-1$. The *skeleton* of w is the word

$$sk(w) = a_1 a_2 \cdots a_n.$$

Let $u, v \in X^+$. We say that

(i) u is a *prefix* of v if $v = ux$ for some $x \in X^*$;
(ii) u is a *suffix* of v if $v = xu$ for some $x \in X^*$;
(iii) u is a *subword* of v if $v = xuy$ for some $x, y \in X^*$.

We use the notation $u \leq_p v$ ($u \leq_s v$) to means that u is a prefix (suffix) of v. We denote the set of all subwords of v by

$$E(v) = \{u \in X^+ \mid xuy = v, x, y \in X^*\};$$

The word u is an *inter-subword* of v if $v = xuy$ for some $x, y \in X^+$. We denote the set of all inter-subwords of v by

$$\bar{E}(v) = \{u \in X^+ \mid xuy = v, x, y \in X^+\}.$$

For a language $L \subseteq X^+$, let

$$E(L) = \{E(v) | v \in L\}; \ \bar{E}(L) = \{\bar{E}(v) \mid v \in L\}.$$

We also define the sets L_p and L_s as follows,

$$L_p = \{x \in X^+ \mid xy \in L \text{ for some } y \in X^+\};$$
$$L_s = \{y \in X^+ \mid xy \in L \text{ for some } x \in X^+\};$$

2. Elementary Properties of Comma-free Codes

In this sequel we will deal with comma-free codes which are maximal in K, that is $A \subseteq K$, A is a comma-free code such that for any $u \in K \setminus A$, the set $A \cup \{u\}$ is not a comma-free code. In particular, we are interested in the case that $K = X_n^m$ for some $m \geq n$, where $X_n^m = X^n \cup X^{n+1} \cup \cdots \cup X^m$. In this section, we deal with $K = X_1^2$. Let $X = \{a, b\}$. We will see that, for $K = X_1^2$, the comma-free codes $\{a, b\}, \{ab\}, \{ba\}$ are maximal comma-free codes in K.

A singleton set is not always a comma-free code. The following proposition is a characterization of one word which is a comma-free code.

Proposition 2.1. *Let $u \in X^+$. Then the following are equivalent:*

(1) $u \notin Q$;
(2) $u^2 = xuy$ *for some* $x, y \in X^+$;
(3) $\{u\}$ *is not a comma-free code.*

Proof. (1) $\Rightarrow$ (2) and (2) $\Rightarrow$ (3) are immediate. (3) $\Rightarrow$ (1) follows from Proposition 2.4 ([2]).

From the above result we see that for a word $u \in X^+$, u is primitive if and only if $u^2 \neq xuy$ for any $x, y \in X^+$. Also a language $\{u\}$ is a comma-free code if and only if u is primitive.

The following is immediate.

Proposition 2.2. *Let $X = \{a, b\}$ and let $A \subseteq X^+$ be a comma-free code.*

(1) *If $X \cap A \neq \emptyset$, then $A \subseteq X$.*
(2) *If $X^2 \cap A \neq \emptyset$, then $A = \{ab\}$ or $A = \{ba\}$.*

We note that the sets $X, \{ab\}$ and $\{ba\}$ are in fact maximal comma-free codes over $X = \{a, b\}$. Let $A \subseteq X^+$. We call A an *n-comma-free code* (see [3]) if every n elements from A is a comma-free code. It is known that for $n \geq 3$, every n-comma-free code is a comma-free code (see [2]). A word $u \in X^+$ is said to be *non-overlapping* if $u = vx = yv$, $v, x, y \in X^*$ implies $v = 1$ (see [3]). The set of all non-overlapping words over X will be, as usual, denoted by $D(1)$.

Proposition 2.3. ([2]) *Every 2-comma-free code is an infix code.*

Since every infix code is a prefix code which is also a suffix code (see [3]), we see that every 2-comma-free code is a prefix code and also a suffix code.

Proposition 2.4. ([1]) *Let $X = \{a, b\}$ be an alphabet and $u \in Q$. Then $\{u\}$ is a maximal comma-free code if and only if $lg(u) = 2$.*

Lemma 2.5. ([3], Proposition 11.2) *Let $A \subseteq X^+$. Then A is a comma-free code if and only if A is an infix code and $A \cap A_s A_p = \emptyset$.*

We note that if a language $A \subset X^+$ is such that $A \cap A_s A_p = \emptyset$, then $A \subset Q$. Since every uniform code is an infix code, the following lemma is a direct consequence of Lemma 2.5.

Lemma 2.6. *Let $A \subseteq Q$ be a uniform code. Then A is a comma-free code if and only if $A \cap A_s A_p = \emptyset$.*

3. Characterizations of Two-element Comma-free Codes

In this section we give the characterization for two-element comma-free codes. First we define some terms.

Let $u, v \in X^*$. We say that the word v is a *conjugate* of u if $u = xy$ and $v = yx$ for some $x, y \in X^*$. The conjugate relation is then an equivalence relation defined on X^*. Thus, for $u \in X^+$, if we let $[u] = \{v \in X^+ | \ v \text{ is a conjugate of } u\}$, then $[u]$ forms an equivalence class. If $u \in Q$, then we call $[u]$ a *primitive conjugate class.* If two words $u \neq v$ are such that u is a conjugate of v, then we call the pair $\{u, v\}$ a *conjugate pair.* It is clear that if $\{u, v\}$ is a conjugate pair, then $lg(u) = lg(v)$.

We call the pair $\{u, v\}$ a *coconjugate pair* if there exist $x, y, z \in X^+$ and $k \geq 1$ such that one of u, v is of the form, say $u = (xy)^k x$ and the other $v = z(xy)$ or $v = (yx)z$. We note that for $k = 1$ and $z = x$, the coconjugate pair $\{u, v\}$ is a conjugate pair. There exists a conjugate pair which is not a coconjugate pair. For example the pair $\{ab, ba\}$ is such an example.

Let $u, v \in X^+$. We say that $\{u, v\}$ is a *quasi-conjugate pair* if there exist $x, y, z \in X^*$ such that (1) one of u, v is $zyxz$ and the other is xzy; or (2) one of u, v is yzx and the other is xy. We note that for the case $z = 1$ in the above definition, the length of u is the same as the length of v and the quasi-conjugate pair is then a conjugate pair.

The following is a characterization of a two-element comma-free code.

Proposition 3.1. *Let $u, v \in Q, u \notin E(v)$ or $v \notin E(u)$. Then the set $\{u, v\}$ is a comma-free code if and only if the following two conditions hold:*

(i) *$\{u, v\}$ is not a quasi-conjugate pair;*
(ii) *$\{u, v\}$ is not a coconjugate pair.*

Proof. Without loss of generality, we let $lg(u) \geq lg(v)$ and $v \notin E(u)$.

($\Rightarrow$) Suppose (i) is not true. Then there exist $x, y, z \in X^*$ such that (say) (1) $u = zyxz$, $v = xzy$. We have that $v^2 = xzyxzy = xuy$ holds. If it is the case (say) (2) $u = yzx$, $v = xy$, then $u^2 = yzxyzx = yzvzx$. Both cases then lead to that the set $\{u, v\}$ is not a comma-free code.

Suppose the condition (ii) is not true. Then, without loss of generality, there exist $x, y, z \in X^+$, $k \geq 1$ such that (say) $u = (xy)^k x$, and $v = z(xy)$ or $v = (yx)z$. (a) If $v = z(xy)$, then $vu = z(xy)(xy)^k x = z(xy)^k xyx = zuyx$. (b) If $v = (yx)z$, then $uv = (xy)^k xyxz = (xy)(xy)^k xz = (xy)uz$. Both cases imply that $\{u, v\}$ is not a comma-free code and condition (ii) holds.

($\Leftarrow$) Suppose $\{u,v\}$ is not a comma-free code. Since u and v are primitives, for some $x,y \in X^+$, one of the following holds:
(1) $u^2 = xvy$; (2) $v^2 = xuy$; (3) $uv = xuy$; (4) $uv = xvy$; (5) $vu = xuy$; (6) $vu = xvy$.

If (1) holds, since $v \notin E(u)$, then $v = v_1v_2$, for some $v_1, v_2 \in X^+$, such that

$$u = xv_1 = v_2y. \tag{3-1}$$

Since $lg(u) \geq lg(v)$, we have $lg(v_1) \leq lg(y)$ and $lg(v_2) \leq lg(x)$. Thus, from Equation (3-1) there exists a word $u' \in X^*$ such that $x = v_2u'$, $y = u'v_1$, and then

$$u = v_2u'v_1, \ v = v_1v_2.$$

Hence the set $\{u,v\}$ is a quasi-conjugate pair, a contradiction.

If (2) holds, then $u = u_1u_2, u_1, u_2 \in X^+$, such that

$$v = xu_1 = u_2y. \tag{3-2}$$

Since $lg(u) \geq lg(v)$ we have $lg(u_1) \geq lg(y)$, $lg(u_2) \geq lg(x)$. Thus, from Equation (3-2), there exists a word $u' \in X^*$ such that $u_1 = u'y$, $u_2 = xu'$, and then

$$u = u'yxu', \ v = xu'y.$$

Hence the set $\{u,v\}$ is a quasi-conjugate pair, a contradiction.

If (3) holds, then there exist $u_1, u_2, v_1, v_2 \in X^+$, such that $u = u_1u_2$, $v = v_1v_2$ and

$$uv = u_1u_2v_1v_2 = xuy,$$

where $lg(u_1) = lg(x)$, $lg(v_2) = lg(y)$. Thus we have

$$u = u_1u_2 = u_2v_1.$$

By Lemma 1.6 ([3]) there exist $w, z \in X^+$, such that

$$u_1 = wz, u_2 = (wz)^k w, v_1 = zw, \text{for some } k \geq 0,$$

and then

$$u = (wz)^{k+1}w, v = zwv_2.$$

This shows that $\{u,v\}$ is a coconjugate pair, a contradiction.

Similarly if (4) holds, then

$$v = (wz)^{k+1}w, \ u = zwu_2$$

for some $u_2, w, z \in X^*$ and $\{u,v\}$ is a coconjugate pair, a contradiction.

Conditions (5) and (6) are the interchange of u and v of conditions (3) and (4). Thus if conditions (5) or (6) hold, then $\{u,v\}$ is not a comma-free code. We then conclude that if both the conditions (i) and (ii) hold, then $\{u,v\}$ is a comma-free code.

Corollary 3.2. *Let $u, v \in Q \cap X^n, n \geq 1$. Then the set $\{u,v\}$ is a comma-free code if and only if the following two conditions hold:*

(i) $\{u, v\}$ *is not a conjugate pair;*
(ii) $\{u, v\}$ *is not a coconjugate pair.*

Proposition 3.3. *Let* $A \subseteq Q \cap X^n, n \geq 2$. *Then* A *is a comma-free code if and only if the following three conditions hold:*

(i) $A \cap A_s A_p = \emptyset$;
(ii) *For every* $u, v \in A$, $\{u, v\}$ *is not a conjugate pair;*
(iii) *For every* $u, v \in A$, $\{u, v\}$ *is not a coconjugate pair.*

Proof. ($\Rightarrow$) Suppose A is a comma-free code. Then by Proposition 3.1 [2], (i) holds. By Corollary 3.2, conditions $(ii), (iii)$ hold.

($\Leftarrow$) Let $u, v, w \in A$. By Corollary 3.2, each pair $\{u, v\}$, $\{u, w\}$, $\{v, w\}$ is a comma-free code. This fact together with the condition (i) on the set A, we see that $\{u, v, w\}$ is a comma-free code. It follows that the set A is a 3-comma-free code. Since every 3-comma-free code is a comma-free code (Lemma 4.3, [2]), A is then a comma-free code.

4. Maximal Comma-free Codes in a Certain Region K, $K \subseteq X^+$

In this section we show that a maximal comma-free code in X^3 is a maximal comma-free code over $X = \{a, b\}$. We also prove that a maximal comma-free code in X^4 is a maximal comma-free code in X_1^4 which then may not be a maximal comma-free code. In general, a maximal comma-free code in $X^n, n \geq 4$, may not be a maximal comma-free code.

Let $X = \{a, b\}$ and let $x \in X^+$. We call the word $y \in X^+$ the *dual word* of x if y is the word obtained from x by replacing each letter a occurs in x by the letter b and each letter b occurs in x by the letter a. We denote the dual word y of x by $Du(x)$. For example, if $x = a^3b$, then $Du(x) = b^3a$ is the dual word of x. If L is a language, then the *dual language* of L is the language

$$Du(L) = \{Du(x) | x \in L\}.$$

It is clear that if L is a comma-free code, then $Du(L)$ is also a comma-free code. Moreover, if L is a maximal comma-free code in a region X_n^m, then $Du(L)$ is also a maximal comma-free code in the same region.

Proposition 2.2 gave some information on maximal comma-free codes over X, where $X = \{a, b\}$. In fact, it pointed out that the sets X, $\{ab\}$ and $\{ba\}$ are maximal comma-free codes over $X = \{a, b\}$. From this fact, we are able to state the following:

Proposition 4.1. *Let* $X = \{a, b\}$. *If* $A \subset X^+$ *is a maximal comma-free code in a region* $K, K \subseteq X^+$, *then* A *is a comma-free code in the region* $K \cup X_1^2$.

In the following we deal with the maximal comma-free codes in X^3 and in X^4. In 1989, C. Y. Hsieh has listed all the maximal comma-free codes in X^3 and in X^4 in her

M.S. Thesis. We now do a further investgation on the properties of maximal comma-free codes in this area.

Proposition 4.2. *Let $X = \{a, b\}$. If $A \subseteq X^3$ is a comma-free code, then $|A| = 1$ or 2. Each comma-free code $A \subseteq X^3$ with two elements is a maximal comma-free code.*

Proof. By Lemma 2.1, $A = \{u\}$ is a comma-free code if and only if $u \in Q$. Thus a comma-free code with one element in X^3 exists. Now consider the case $A = \{u, v\} \subseteq X^3$. All the primitive words of length three over $X = \{a, b\}$ are in the set

$$B = \{a^2b, aba, ab^2, ba^2, bab, b^2a\}.$$

Let $A = \{u, v\} \subset B$. If the number of the letter a in u is the same as the number of the letter a in v, then $A = \{u, v\}$ is a conjugate pair and hence, by Corollary 3.2, $\{u, v\}$ is not a comma-free code. It is clear that $\{aba, bab\}$ is not a comma-free code. The rest of the pairs are

$$\{a^2b, ab^2\}, \{a^2b, bab\}, \{a^2b, b^2a\}, \{aba, ab^2\}, \{aba, b^2a\}, \{ab^2, ba^2\}, \{ba^2, bab\}, \{ba^2, b^2a\}.$$

Since, in each pair in the above, the number of the letter a in u is not equal the number of the letter a in v, hence $\{u, v\}$ is not a conjugate pair. By observation these pairs are also not coconjugate pairs. Hence by Corollary 3.2, the above pairs are comma-free codes. It is then clear that the 8 pairs in the above are the only comma-free codes with two elements of length 3.
It remains to see the case that $A \subseteq X^3 \cap Q$ with $|A| \geq 3$. In fact if $|A| \geq 3$, then at least two words in A have the property that the number of letter a in both words u and v are equal. By the previous argument we see that A is not a comma-free code. We then conclude that the only comma-free codes with two elements in X^3 are those 8 pairs listed in above. It is also clear that the two elements comma-free codes are maximal in X^3.

The above 8 pairs of maximal comma-free codes in X^3 are in fact maximal comma-free codes over X. To show this fact, suppose A is a maximal comma-free code in X^3. Then

$$A \cap \{a^2b, aba, ba^2\} \neq \emptyset \tag{4-1}$$

and $A \cap \{ab^2, bab, b^2a\} \neq \emptyset$. To show that A is a maximal comma-free code, by Propositions 2.1 and 2.2, we need only show that for any $v \in X^n \cap Q, n \geq 4$, $A \cup \{v\}$ is not a comma-free code.

We consider for $a^2 \in E(v)$ and $a^2 \notin E(v)$.

(1) If $a^2 \in E(v)$, then clearly a^2b, ba^2 are both in $E(v^2)$. Thus both $a^2b \in A$ and $ba^2 \in A$ imply that $A \cup \{v\}$ is not a comma-free code. Hence by Equation (4-1), $aba \in A$ and by observation,

$$b^2a \in A \text{ or } ab^2 \in A. \tag{4-2}$$

But since $aba \in A$, we have that all of $aba \in \bar{E}(v), ab \leq_s v$ and $ba \leq_p v$ imply $A \cup \{v\}$ is not a comma-free code. Hence $b^2 \in E(v)$. But this implies that $b^2a, ab^2 \in E(v^2)$. By Equation (4-2), $A \cup \{v\}$ is not a comma-free code.

(2) If $a^2 \notin E(v)$, then $b^2 \in E(v)$ or $v = (ab)^k a, k \geq 2$. If $b^2 \in E(v)$, then by similar argument to (1), we can show that $A \cup \{v\}$ is not a comma-free code. If $v = (ab)^k a$, then $aba \in E(v)$ and $a^2b, ba^2 \in E(v^2)$. And all cases imply $A \cup \{v\}$ is not a comma-free code.

From (1) and (2), it is then clear that A is a maximal comma-free code over $X = \{a, b\}$.

Remark. The above result is in fact pointed out by Hsieh in [1] without proof.

The following is a characterization of a maximal comma-free code.

Lemma 4.3. *Let A be a comma-free code. Then A is maximal if and only if for any $u \in Q \setminus A$, one of the following conditions holds true:*

(i) $u \in \bar{E}(A^2)$,
(ii) $\bar{E}(Au) \cap A \neq \emptyset$,
(iii) $\bar{E}(uA) \cap A \neq \emptyset$,
(iv) $\bar{E}(u^2) \cap A \neq \emptyset$,
(v) $u \in \bar{E}(Au)$,
(vi) $u \in \bar{E}(uA)$.

Proof. Since A is a comma-free code, by definition, $A^2 \cap X^+AX^+ = \emptyset$. Now if A is maximal, then, since $u \in Q\backslash A$, by the definition of maximality,

$$(A \cup \{u\})^2 \cap X^+(A \cup \{u\})X^+ \neq \emptyset.$$

The word u is primitive, by Proposition 2.1, $u^2 \neq xuy$ for any $x, y \in X^+$. Thus one of the following conditions must hold:

(1) $A^2 \cap X^+uX^+ \neq \emptyset$,
(2) $Au \cap X^+AX^+ \neq \emptyset$,
(3) $Au \cap X^+uX^+ \neq \emptyset$,
(4) $uA \cap X^+AX^+ \neq \emptyset$,
(5) $uA \cap X^+uX^+ \neq \emptyset$,
(6) $u^2 \cap X^+AX^+ \neq \emptyset$,

It is clear that one of the conditions $(i) - (vi)$ must hold.

Conversely if one of the conditios $(i) - (vi)$ holds, then A is a maximal comma-free code.

We present a known lemma which we need in the following proposition.

Lemma 4.4 (see [4]) *Let $X = \{a, b, ...\}$ such that $a \neq b$. Then for any $u \in X^+$, either ua or ub is primitive.*

Proposition 4.5. *Let $n > 2$. If A is a maximal comma-free code in X^n and $v \in Q$ with $lg(v) < \frac{n}{2}$, then $A \cup \{v\}$ is not a comma-free code.*

Proof. Suppose $A \subseteq X^n$ is a maximal comma-free code in X^n. Let $v \in Q$ with $lg(v) < \frac{n}{2}$. We want to show that $A \cup \{v\}$ is not a comma-free code. If $v \in E(A)$, then clearly $A \cup \{v\}$ is not a comma-free code. Thus in this proof we need only to consider for the case that $v \in Q$ and $v \notin E(A)$.

Consider a word $u = vzv \in X^n \cap Q$, where $z \in X^+$. Since, by Lemma 1.1 (2), the word vzv is primitive if and only if v^2z is primitive, by the above lemma, such a word u exists.

If $u \in A$, then $v \in E(A)$ and which is not the case. Thus $u \notin A$. As A is a maximal comma-free code in X^n, we have that $A \cup \{u\}$ is not a comma-free code. It follows that one of the following conditions must hold.

(1) $w_1w_2 = xuy$, for some $w_1, w_2 \in A, x, y \in X^+$;
(2) $u^2 = xwy$, for some $w \in A, x, y \in X^+$;
(3) $w_1u = xw_2y$, for some $w_1, w_2 \in A, x, y \in X^+$;
(4) $uw_1 = xw_2y$, for some $w_1, w_2 \in A, x, y \in X^+$;
(5) $uw = xuy$, for some $w \in A, x, y \in X^+$;
(6) $wu = xuy$, for some $w \in A, x, y \in X^+$;

We will show that some of the conditions can never valid and the others will lead to that $A \cup \{v\}$ is not a comma-free code.

If the condition (1) holds, then $w_1w_2 = xuy = xvzvy$, and either $v \in E(w_1)$ or $v \in E(w_2)$. This is not the case. Hence the condition (1) can never be true.

If the condition (2) holds, then $vzvvzv = xwy$. Since $lg(w) = lg(vzv)$, we see that $v \in E(w)$ and $w \in A$, not possible.

If the condition (3) holds, then $w_1vzv = xw_2y$. Since $v \notin E(A)$, there exist $v_1, v_2, w_{11}, w_{12} \in X^+$ such that $v = v_1v_2$ and $w_1 = w_{11}w_{12}$, and $w_2 = w_{12}v_1$. Thus $w_1v = w_{11}w_2v_2$, with $w_1, w_2 \in A$.

If condition (4) holds, then $vzvw_1 = xw_2y$ and since $v \notin E(A)$, there exist $v_1v_2, w_{11}, w_{12} \in X^+$ such that $v = v_1v_2$ and $w_1 = w_{11}w_{12}$ and $w_2 = v_2w_{11}$. It follows that $vw_1 = v_1w_2w_{12}$ with $w_1, w_2 \in A$.

If condition (5) holds, then $vzvw = xvzvy$. Since $v \notin E(A)$ we must have $lg(vy) > lg(w)$ and there exists $r \in X^+$ such that $rw = vy$. If $lg(r) = lg(v)$, then $r = v$ which then implies that $w = y$. From the equation $vzv = xvzv$ we see that $x = 1$, not possible. If $lg(r) > lg(v)$, from $vzvw = xvzvy = xvzrw$, we have $vzv = xvzr$ and hence $lg(vz) > lg(xvz)$, not possible. If it is the case that $lg(r) < lg(v)$, then, from the condition that $vzvw = xvzvy$ along with the fact that $lg(vy) > lg(w)$ and $rw = vy$, we see that $v = v'r$, for some $v' \in X^+$. It follows that $vw = v'rw = v'vy$.

If the condition (6) holds, then $wvzv = xvzvy$. Since $v \notin E(A)$, we must have $lg(xv) > lg(w)$ and there exists $r \in X^+$ such that $wr = xv$. If $lg(r) = lg(v)$, then $v = r$ and $x = w$. From the equation $vzv = vzvy$, we have $y = 1$, not possible. Now if $lg(r) > lg(v)$, then $lg(w) < lg(x)$ and from the equation $wvzv = xvzvy$, we see that $lg(zv) > lg(zvy)$, not possible. If it is the case that $lg(r) < lg(v)$, then from the condition that $wvzv = xvzvy$ along with the fact that $lg(xv) > lg(w)$ and $wr = xv$, we see that $v = rv'$ for some $v' \in X^+$. It follows that $wv = wrv' = xvv'$.

Summing up our proof, we have that the conditions (1) and (2) can never hold true. Thus one of the conditions (3) - (6) must hold. In either case we see that the set $A \cup \{v\}$ is not a comma-free code. This completes the proof of the proposition.

Proposition 4.6. *If $n > 2$ and $A \subseteq X^+$ is a maximal comma-free code in X^n, then A is a maximal comma-free code in $X_1^m \cup X^n$, where $m = \frac{n}{2} - 1$ if n is even and $m = \frac{n-1}{2}$ if n is odd.*

We remark here that by the duality of the letters a and b, that the number of (maximal) comma-free codes in X^n, $n \geq 1$ is always even.

Lemma 4.7. *Let X be an alphabet and let A be a comma-free code in $X^n, n \geq 1$. Let $v \in X_1^{n-1}$ and let $A \cup \{v\}$ be a comma-free code. Then for some $z \in X^+, vz \in X^n \cap Q$, $A \cup \{vz\}$ is not a comma-free code if and only if*
(*) $vzA \cap X^+(A \cup \{vz\})X^+ \neq \emptyset$.

Proof. ($\Leftarrow$) If the condition (*) holds, then either $(vz)(w) = xw'y,\ w, w' \in A, x, y \in X^+$; or $(vz)(w) = xvzy,\ w \in A, x, y \in X^+$. But both cases imply that $A \cup \{vz\}$ is not a comma-free code.
($\Rightarrow$) Suppose $A \cup \{vz\}$ is not a comma-free code. Since A is a comma-free code and $vz \in Q$, then one of the following conditions holds :

(1) $A^2 \cap X^+vzX^+ \neq \emptyset$,
(2) $vzA \cap X^+AX^+ \neq \emptyset$,
(3) $vzA \cap X^+vzX^+ \neq \emptyset$,
(4) $Avz \cap X^+AX^+ \neq \emptyset$,
(5) $Avz \cap X^+vzX^+ \neq \emptyset$,
(6) $(vz)^2 \cap X^+AX^+ \neq \emptyset$.

If (1) holds, then $A^2 \cap X^+vX^+ \neq \emptyset$. This contradicts that $A \cup \{v\}$ is a comma-free code.

If (4) holds, then there exist $x, y \in X^+, w_1, w_2 \in A$, such that $w_1vz = xw_2y$. Now if $lg(z) = lg(y)$, then v is a suffix of $w_2 \in A$. This is not possible, since $A \cup \{v\}$ is a comma-free code. If $lg(z) < lg(y)$, then $w_1v = xw_2y'$ for some $y' \in X^+$. This is also not possible, since $A \cup \{v\}$ is a comma-free code. Finally, if $lg(z) > lg(y)$, we have then $w_1vz' = xw_2, z' \in X^+$. In here $lg(vz') < lg(vz) = lg(w_2)$ and v is an inter-subword of w_2, not possible, since $A \cup \{v\}$ is a comma-free code. This shows that condition (4) can never be true.

If (5) holds, then there exist $x, y \in X^+, w \in A$ such that $wvz = xvzy$. Since $lg(zy) > lg(z)$, there exist $z' \in X^+$ such that $wv = xvz'$. The set $\{w, v\}$ is not a comma-free code, contradicting that $A \cup \{v\}$ is a comma-free code. Therefore condition (5) can never be true.

If (6) holds, then there exist $x, y \in X^+$ and $w \in A$, such that $vzvz = xwy$. It is easy to see that $\{vz, w\}$ is a conjugate pair. Hence $w^2 \in X^+vX^+$ and $A \cup \{v\}$ is then not a comma-free code.

By above discussion, we have that condition (2) or (3) must hold. Thus $(*)$ holds.

Remark. In the above result if in addition the word vz is non-overlapping i.e., $vz \in D(1)$, then $A \cup \{vz\}$ is not a comma-free code if and only if $vzA \cap X^+AX^+ \neq \emptyset$.

To prove this fact, we first

Claim: If $vz \in X^n \cap D(1)$, then $vzA \cap X^+vzX^+ = \emptyset$.

For this proof, if $vzA \cap X^+vzX^+ \neq \emptyset$, then there exists $w \in A$ such that $vzw = xvzy$, $x, y \in X^+$. Now let $u = vz$. Then there exist $u_1, u_2, w_1, w_2 \in X^+$, such that $u = u_1u_2$, $w = w_1w_2$ and

$$uw = u_1u_2w_1w_2 = xu_1u_2y,$$

where $lg(u_1) = lg(x)$, $lg(w_2) = lg(y)$. Since $u, w \in X^n$, we have

$$u_1u_2 = u_2w_1.$$

By Lemma 1.6 ([3]) there exist $z_1, z_2 \in X^+$, such that

$$u_1 = z_1z_2, u_2 = (z_1z_2)^k z_1, w_1 = z_2z_1, \text{ for some } k \geq 0.$$

Thus $u = u_1u_2 = (z_1z_2)^{k+1}z_1$, and $vz = u \notin D(1)$. Hence $vzA \cap X^+vzX^+ = \emptyset$.

Now it is clear that for the case $vz \in D(1)$, if $A \cup \{vz\}$ is not a comma-free code, then condition (3) does not hold and hence condition (2) must hold. We note that if condition (2) holds, then clearly $A \cup \{vz\}$ is not a comma-free code.

Corollary 4.8. *Let X be an alphabet and A be a comma-free code in $X^n, n \geq 1$, $v \in X_1^{n-1}$. Suppose $A \cup \{v\}$ is a comma-free code. Then if for some $z \in X^+, zv \in X^n \cap Q$, we have $A \cup \{zv\}$ is a not a comma-free code if and only if*

$(**)$ $Azv \cap X^+(A \cup \{zv\})X^+ \neq \emptyset$.

Moreover, if $zv \in D(1)$, then we have $A \cup \{zv\}$ is a not a comma-free code if and only if

$(**)'$ $Azv \cap X^+AX^+ \neq \emptyset$.

The *mirror image of a word* $x = a_1a_2 \ldots a_n$, where $a_i \in X$, denoted by $\tilde{x}$, is the word $\tilde{x} = a_na_{n-1} \ldots a_1$. It is immediate that for any $f \in Q, \tilde{f} \in Q$. Let $L \subseteq X^+$. We let $\tilde{L}$ be the set consist of all mirror image of words in L.

Lemma 4.9. *Let X be an alphabet and A be a maximal comma-free code in $X^n, n \geq 1$, $v \in Q \cap X_1^{n-1}$. Let $\tilde{v}$ be the mirror image of v. Then $A \cup \{v\}$ is not a comma-free code if and only if $\tilde{A} \cup \{\tilde{v}\}$ is not a comma-free code.*

Proof. Since the mirror of $A \cup \{v\}$ is $\tilde{A} \cup \{\tilde{v}\}$, by Proposition 11.2 ([4]), The lemma is immediate.

Lemma 4.10. *Let $X = \{a, b\}$ and let A be a maximal comma-free code in $X^n, n \geq 3$. Then for $v \in \{a^ib, ab^i, ba^i, b^ia | 1 \leq i \leq n-2\}$, $A \cup \{v\}$ is not a comma-free code.*

Proof. Suppose $A \cup \{v\}$ is a comma-free code.

The case $v = a^i b$.

Let A be a maximal comma-free code in X^n. If $a^{n-1}b \in A$, then $a^i b$ is a suffix of $a^{n-1}b$ and hence $A \cup \{v\}$ is not a comma-free code. Now suppose $a^{n-1}b$ is not in A. Then since A is a maximal comma-free code in X^n, we have $A \cup \{a^{n-1}b\}$ is not a comma-free code. Then by the Corollary 4.8(**)$'$, there exist $x, y \in X^+$ such that

$$w_1 a^{n-1} b = x w_2 y, \text{where } w_1, w_2 \in A.$$

Let $k = lg(x)$. Then $k < n$ and $w_1 a^k = x w_2$. If $k \leq i$, then $w_1 a^i b = x w_2 a^{i-k} b$. The set $A \cup \{a^i b\}$ is then not a comma-free code. If $k > i$, then since $a^k \leq_s w_2$ and $w_2 \in Q$, $a^i b \in E(w_2^2)$ is true. Hence $A \cup \{a^i b\}$ is not a comma-free code.

The case $v = b^i a$.

This is a dual case and so by a similar argument we can show that $A \cup \{v\}$ is not a comma-free code.

The case $v = ba^i$.

Since ba^i is a mirror image of $a^i b$, if ba^{n-1} is in A, then $A \cup \{v\}$ is not a comm-free code. Otherwise $A \cup \{ba^{n-1}\}$ is not a comma-free code. By the Remark after Lemma 4.7, there exist $x, y \in X^+$, such that

$$ba^{n-1} w_1 = x w_2 y, \text{where } w_1, w_2 \in A.$$

Let $k = lg(y)$. Then $k < n$ and $a^k w_1 = w_2 y$. If $k \leq i$, then $ba^i w_1 = ba^{i-k} w_2 y$. Hence $A \cup \{ba^i\}$ is not a comma-free code. If $k > i$, then $a^k \leq_p w_2$. Since $w_2 \in Q$, $ba^i \in E(w_2^2)$ is true. It follows that, the set $A \cup \{ba^i\}$ is not a comma-free code.

The case $v = ab^i$.

This is a dual case and by a similar argument as the above we can show that $A \cup \{v\}$ is not a comma-free code.

Corollary 4.11. *Let $X = \{a, b\}$ and let A be a maximal comma-free code in $X^n, n \geq 3$. Then A is a maximal comma-free code in $X^n \cup \{a^i b, ab^i, ba^i, b^i a | 1 \leq i \leq n-2\}$.*

Lemma 4.12. *Let $X = \{a, b\}$ and let A be a maximal comma-free code in $X^n, n \geq 4$, $v \in Q \cap X^3$. Then $A \cup \{v\}$ is not a comma-free code.*

Proof. First we note that

$$X^3 \cap Q = \{a^2 b, ab^2, ba^2, b^2 a\} \cup \{aba, bab\}.$$

If $v \in \{a^2 b, ab^2, ba^2, b^2 a\}$, then by Lemma 4.10, $A \cup \{v\}$ is not a comma-free code.

If $v \in \{aba, bab\}$, then by the definition of dual word, we need only consider $v = aba$. Suppose $A \cup \{v\}$ is a comma-free code. Then we have that $ba \notin A_p$, $ab \notin A_s$ and $aba \notin E(A)$.

(1) n is odd.

In this case we consider the word $(ab)^k a = vz \in X^n \cap Q$, $k \geq 2$. As $aba \notin E(A)$, we see that $(ab)^k a \notin A$. Since A is a maximal comma-free code in X^n, $A \cup \{vz\}$ is not a comma-free code. Thus by Lemma 4.7, there exist $x_1, y_1 \in X^+$ such that one of the following conditions holds:

(1-1) $w(ab)^k a = x_1(ab)^k a y_1$, where $w \in A$;

(1-2) $w_1(ab)^k a = x_1 w_2 y_1$, where $w_1, w_2 \in A$.

If (1-1) holds, then $\{w, (ab)^k a\}$ is a coconjugate pair and then $ab \leq_s w$. Thus $wv \in X^+ v X^+$. The set $A \cup \{v\}$ is then not a comma-free code, a contradiction.

If (1-2) holds, then since $ab \notin A_s$ and $aba \notin E(A)$, then we have that $w_1 a = x_1 w_2$ and then $w_1 v = x_1 w_2 ba \in X^+ w_2 X^+$. It follows that the set $A \cup \{v\}$ is not comma-free code, a contradiction. This show that n odd can never be true.

(2) n is even.

We consider the word $a(ab)^k a = zv \in X^n \cap Q$, $k \geq 1$. As $aba \notin E(A)$, we see that $a(ab)^k a \notin A$. Since A is a maximal comma-free code in X^n, we have that the set $A \cup \{a(ab)^k a\}$ is not a comma-free code. By Corollary 4.8, there exist $x_1, y_1 \in X^+$ such that one of the following conditions holds :

(2-1) $wa(ab)^k a = x_1 a(ab)^k a y_1$, where $w \in A$;

(2-2) $w_1 a(ab)^k a = x_1 w_2 y_1$, where $w_1, w_2 \in A$.

If (2-1) holds, then $\{w, a(ab)^k a\}$ is a coconjugate pair and $a(ab)^k \leq_s w$. Thus $wv \in X^+ v X^+$.

If (2-2) holds, then since $ab \notin A_s$ and $aba \notin E(A)$, then we have that $w_1 a^2 = x_1 w_2$. Thus a^2 is suffix of w_2.

Now consider $vz = (ab)^k a^2 \in X^n \cap Q, k \geq 1$. As $(ab)^k a^2 \notin A$, the set $A \cup \{(ab)^k a^2\}$ is not a comma-free code. Again by Lemma 4.7 there exist $x_2, y_2 \in X^+$ such that one of the following conditions holds :

(2-2-1) $(ab)^k a^2 w = x_2 (ab)^k a^2 y_2$, where $w \in A$;

(2-2-1) $(ab)^k a^2 w_3 = x_2 w_4 y_2$, where $w_3, w_4 \in A$.

If (2-2-1) holds, then $\{w, (ab(ab)^{k-1}a)a\}$ is a coconjugate pair and $b(ab)^{k-1}a^2 \leq_p w$. Thus $vw \in X^+ v X^+$. It follows that the set $A \cup \{v\}$ is not a comma-free code, a contradiction.

If (2-2-2) holds, then, again, since $ba \notin A_p$ and $aba \notin E(A)$, then we have that $a^2 w_3 = w_4 y_2$. But $a^2 \leq_s w_2$, we have that $w_2 w_3 \in X^+ w_4 X^+$. This then implies that A is not a comma-free code, a contradiction. We then conclude that n even is also not possible. Therefore under the assumption in the lemma, $A \cup \{v\}$ is not a comma-free code.

Th following is now clear.

Proposition 4.13. *Let $X = \{a, b\}$ and let $A \subseteq X^n, n \geq 4$ be a maximal comma-free code in X^n. Then A is a maximal comma-free code in $X^n \cup X_1^3$.*

Corollary 4.14. *Let $X = \{a, b\}$ and $A \subseteq X^4$ be a maximal comma-free code in X^4. Then A is a maximal comma-free code in X_1^4.*

If $n \geq 4$, then not every maximal comma-free code $A \subset X^n$ is a maximal comma-free code. We will prove this fact after we present the following proposition.

Proposition 4.15. *Let $X = \{a, b\}$. Then for every $n \geq 4$, there is a maximal comma-free code $A \subseteq X^n$ with*

$$|A| = \begin{cases} 2^{\frac{n-1}{2}}, & \text{if n is odd;} \\ 2^{\frac{n}{2}} - 1, & \text{if n is even.} \end{cases}$$

Proof. For the case $n = 4$, it is easy to see that the language $\{a^3b, ab^2a, ab^3\}$ is a maximal comma-free code in X^4. We then assume that $n \geq 5$.

Case I: $n = 2k+1, k \geq 2$.

Let $A = \{a^k xb \mid x \in X^k\}$. Then $|A| = 2^k = 2^{\frac{n-1}{2}}$. We show that A is a maximal comma-free code in X^n. Since $A \subset X^n$ is a uniform code, A is an infix code. It is clear that $A \subset Q$. By observation the particular set A has the property that $A_s A_p \cap A = \emptyset$. By Proposition 2.6, A is a comma-free code.

For the maximality of A, suppose $u \in X^n \cap (Q \setminus A)$. Then the word u may be one of the following two forms (1) and (2).

(1) $u = a^k xa$, $x \in X^k$. In this case, since $u \in Q$, u is not a power of the letter a. Then $x = x'ba^r$, for some $r \geq 0$, $x' \in X^*$. It is easy to see that, for some $v \in A$, $\{u, v\}$ is a conjugate pair. By Corollary 3.2, $A \cup \{u\}$ is not a comma-free code.

(2) $u = a^i bx, 0 \leq i < k$, $x \in X^{n-i-1}$. In this case, since $n - 1 \geq k$ and also $n - i - 2 \geq k$, we have $a^{n-1}b, a^{n-i-2}ba^ib \in A$
and from

$$a^{n-1}bu = a^{n-1}ba^ibx = a^{i+1}(a^{n-i-2}ba^i)bx,$$

we see that $Au \cap X^+AX^+ \neq \emptyset$ and $A \cup \{u\}$ is not a comma-free code. This shows that the language A is a maximal comma-free code in X^n.

Case II: $n = 2k$, $k \geq 2$. Let $A = \{a^{k-1}xb | x \in X^k \setminus \{ba^{k-1}\}\}$. Then $|A| = 2^k - 1 = 2^{\frac{n}{2}} - 1$. We show that A is a maximal comma-free code in X^n.

First, we show that A is a comma-free code. Since $A \subseteq X^n$, A is an infix code. It is clear that $A \subseteq Q$ and by observation, the condition $A \cap A_s A_p = \emptyset$ holds true for A. By Proposition 2.6, A is a comma-free code.

For the maximality of A, suppose $u \in X^n \cap (Q \setminus A)$. The word u may be one of the following two forms (1) and (2).

(1) $u = a^{k-1}xa$, $x \in X^k$. In this case, again since $u \in Q$, u is not a power of the letter a. Then $x = x'ba^r, r \geq 0, x' \in X^*$. We have then for some $v \in A$, the set $\{u, v\}$ is a conjugate pair and hence $A \cup \{u\}$ is not a comma-free code.

(2) $u = a^i bx, 0 \leq i < k-1$, $x \in X^{n-i-1}$. In this case, since $n - 1 \geq k - 1$ and also $n - i - 2 \geq k - 1$, and $a^{n-1}b, a^{n-i-2}ba^ib \in A$ such that

$$a^{n-1}bu = a^{n-1}ba^ibx = a^{i+1}(a^{n-i-2}ba^i)bx.$$

Thus $Au \cap X^+AX^+ \neq \emptyset$ and $A \cup \{u\}$ is not a comma-free code. This completes the proof that A is a maximal comma-free code in X^n.

Proposition 4.16. *Let $X = \{a, b\}$ and let $n \geq 4$. There exists a maximal comma-free code in X^n which is not a maximal comma-free code.*

Proof. Suppose $n = 4$.

The language $A = \{a^3b, ab^2a, ab^3\}$ is a maximal comma-free code in X^4. Since $A \cup \{aba^2b\}$ is a comma-free code and so A is not a maximal comma-free code.

We consider for the case $n \geq 5$.

Case I: $n = 2k + 1, k \geq 2$.

In this case, by Proposition 4.15, the language $A = \{a^kxb | x \in X^k\}$ is a maximal comma-free code in X^n. Let $u = a^kb^kab \in X^{n+1}$ and let $L = A \cup \{u\}$. First, since $A \cap E(u) = \emptyset$, L is an infix code; second, since A is a comma-free code and $u \in Q$, together by obvervation, $L \cap L_sL_p = \emptyset$. By Proposition 2.6, L is a comma-free code.

Case II: $n = 2k, k \geq 3$.

In this case, by Proposition 4.15, $A = \{a^{k-1}xb | x \in X^k \setminus \{ba^{k-1}\}\}$ is a maximal comma-free code in X^n. Let $u = a^{k-1}b^kab \in X^{n+1}$ and consider the language $L = A \cup \{u\}$: First, since $A \cap E(u) = \emptyset$, L is an infix code; second, again since A is a comma-free code and $u \in Q$, together by observation, $L \cap L_sL_p = \emptyset$. By Proposition 2.6, we have L is a comma-free code. This completes the proof that A is not a maximal comma-free code.

References

[1] C. Y. Hsieh, *Some Algebraic Properties of Comma-Free Codes*, M.S. Thesis, Institute of Applied mathematics, National Chung-Hsing University (1989).

[2] C. Y. Hsieh, S. C. Hsu and H. J. Shyr, "Some Algebraic Properties of Comma-Free Codes, RIMS, Kenkyuroku," 697, *Japan*,(1989), 57-66.

[3] H. J. Shyr, *Free Monoids and Languages*, Second Edition, HonMin Book Company, Taichung, Taiwan, 1991.

[4] H. J. Shyr and F. K. Tu, "Locall distribution of non-primitive words," *Proceedings of the SEAMS Conference on Ordered Structure and Algebra of Computer Languages (June 1991)*, 202-217, 1993.

Management Information Department, National Chin-Yi Institute of Technology, Taichung, Taiwan.

Institute of Applied Mathematics, National Chung-Hsing University, Taichung, Taiwan.