INFORMATION AND CONTROL 7, 23-36 (1964)

On the Synchronizing Properties of Certain Prefix Codes

M. P. Schützenberger

Faculté des Sciences, Poitiers, France

A special family J of prefix codes is considered. It is verified that if $A \in J$ has not a certain synchronizing property, then $A = C^p$ (p > 1), where C is another code from the same family.

I. INTRODUCTION

Let F be the free monoid generated by the set ("alphabet") X; F consists of its neutral element 1 (the so-called "empty word") and of the set $X^* = X \cup X^2 \cup \cdots \cup X^n \cdots$ of all words of positive degree (or "length"). We denote by X^* the collection of all nonempty subsets of X^* and we consider the family J of all prefix codes A that can be defined by taking an arbitrary $H \in X^*$ and by letting a word f belong to A iff f has some right factor (or "final segment") in H, i.e. $f \in FH$, and no proper (i.e., $\neq f$) left factor (or "initial segment") of f has the same property, i.e., $f \in FHX^*$.

This theory is due to B. Mandelbrot, who studied in details the especially important case where H is a particular letter (the so-called "space") of the alphabet (cf. bibliography in Mandelbrot (1957) and Mandelbrot (1961)). A special case obtains by selecting an arbitrary subset of states of a definite automaton, and by defining A as the set (provided it belongs to \mathbf{X}^*) of all words at the last letter of which the distinguished set is reached for the first time. This construction is part of a more general theory, due to P. G. Neuman (Neuman (1962)).

Both of the authors quoted have emphasized the synchronizing properties of the codes of the family J. Indeed, let us say that the prefix code A is almost surely synchronizing if there exists at least one word $a \in F$ such that $fa \in A^*$ (=A U A^2 U ··· U A^n ···) for all $f \in F$. In Winograd's theory (Winograd (1963), cf. also Winograd (1962)) a would be called a universal synchronizing word. If J_1 denotes the subset of all almost surely synchronizing codes of J, we intend to verify J =

 $\{A^p: p>0, A \in J_1\}$. In other terms, if $A \in J$ is not a.s. synchronizing then there exists a unique $C \in J_1$ and natural number p>1 which are such that A consists of all products of p words from C. These notions may be clarified by the following examples in which $X = \{x, y\}$.

- (i) $H = \{x\}$; then $A = FH \setminus FHX^*$ (= $\{f \in FH: f \in FHX^*\}$) consists of all words x, yx, y^2x , \cdots , y^nx , \cdots . Since obviously $FA \subset A^*$, A belongs to J_1 .
- (ii) $H = \{xx, xyx, xyy, yyx, yyyy\}$. The corresponding prefix code A consists of H and the words yxx, yxyx, yxyy, yyyx. In fact, $A = C^2$ where $C = FH' \backslash FH'X^*$ with $H' = \{x, yy\}$. Since $CA^* \subset A^*C$ and $A^* \cap A^*C = \emptyset$, A does not belong to J_1 but it is the square of the code $C \in J_1$.
- (iii) As a related counter example one might consider the prefix code A consisting of x and of all the words of the form y^dxf where $d=1, 2, \dots, n, \dots$ and where f is an arbitrary word of degree ("length") d. Thus $A \in J$ because for instance, $x, yxx \in A$ and $yxx \in FAX^*$. Since for every $f' \in F$ of degree d' one has $f'x^d \in A^*$ when d > d', every word of F can be "resynchronized." However, under the same hypothesis $y^df' \in F \setminus A^*F$, and one sees that there exists no universal synchronizing word, i.e., no word which resynchronizes all the words of F.

II. DEFINITIONS AND NOTATIONS

For the sake of completeness we recall first some well-known facts concerning prefix codes and we summarize some general properties of the family J.

By definition, a prefix code is a set $A \in \mathbf{X}^*$ which satisfies the condition

$$\mathfrak{u}_r':AX^*\cap A=\emptyset.$$

Indeed \mathfrak{A}_r' simply expresses that every word of F has at most one left factor in A. Thus, letting $T = F \backslash AF$, we have $F = T \cup A^*T$ and we can define inductively a mapping $\tau\colon F\to T$ by setting for any word f, $\tau f = f$ if $f\in T$; $\tau f = \tau f'$ if f=af' where $a\in A$. Thus $\tau f=1$, iff $f\in \{1\}\cup A^*$. The identity $\tau ff'=\tau((\tau f)f')$ is easily checked by examining the two cases of $f\in T$ and $f\in T$. By construction, for all $f\in F$, $\tau Ff=(-\{\tau f'f:f'\in F\})$ is the same as $\tau Tf=(-\{\tau ff:f\in T\})$. It follows that for any $f,f',f''\in F$ one has Card $\tau Tff'f''= Card \tau Tf'$. Indeed, on the one hand, $\tau Tff'f''=\tau((\tau Tf)f'f'')\subset \tau'Tf'f''$ and, on the other hand, Card $\tau Tf'f''= Card \tau((\tau Tf')f'') \subseteq Card \tau Tf'$.

Let p denote the minimum value (possibly infinite) of Card τTa over all $a \in A^*$. Since $1 \in \tau Ta$ for $a \in A^*$, p is positive and since

 $\tau Ta = \{1\}$ is equivalent to $Fa \in A^*$, one sees that p = 1 iff A is a.s. synchronizing.

We now return to the family J.

Property 1. For each $H \in \mathbf{X}^*$ the set $A = FH \backslash FHX^*$ belongs to \mathbf{X}^* and it satisfies the conditions:

$$\mathfrak{U}'_{r_j}: FAX^* \cap A = \emptyset;$$

$$\mathfrak{N}_{r_j}: FA \subset AF.$$

Reciprocally, if the prefix code A satisfies \mathfrak{N}_{rj} , then $A = FH \backslash FHX^*$ where $H = A \backslash X^*A$; further, $H \cap (X^*H \cup HX^* \cup X^*HX^*) = \emptyset$ and AF = FHF.

PROOF: Let $H \in \mathbf{X}^*$. The set $A = FH \backslash FHX^*$ is a subset of X^* ; A is not empty since it contains every word of H of minimal degree. Consider a word of the form faf' where $f \in F$, $a \in A$, $f' \in X^*$. By hypothesis, a = f''h for some $f'' \in F$ and $h \in H$; thus $faf' = ff''hf' \in FHX^*$ and, as a result, $faf' \notin A$. This proves that A satisfies \mathfrak{U}'_{rj} ; hence \mathfrak{U}'_{r} , since $AX^* \subset FAX^*$.

Consider now $f \in F$ and $a \in A$. Again a = f''h for some $f'' \in F$ and $h \in H$. Hence $fa \in FH$ and fa has a left factor, say f'''h', of minimal degree that belongs to FH and, by construction, that does not belong to FHX^* . Thus $F'''h' \in A$ and this proves \mathfrak{N}_{r_j} .

Reciprocally consider any set $A \in \mathbf{X}^*$ and define $H = A \setminus X^*A$. By construction $H \cap X^*H = \emptyset$ and the right factor in A of minimal degree of every word of A belongs to H. Thus, $H \in \mathbf{X}^*$, and $H \subset A \subset X^*H \cup H = FH$. In fact, H is the least set H' such that $A \subset FH'$.

Assume now that A satisfies \mathfrak{N}_{r_j} ; if $f \in F$ and $h \in F$ are such that fh has no proper left factor in FH, fh has no proper left factor in A since $A \subset FH$. However, since $H \subset A$, we have $fh \in FA$, hence $fh \in AF$ and thus, $fh \in A$. This proves $FH \setminus FHX^* \subset A$.

Assuming finally that A is a prefix code, we see that $A \cap FHX^* = \emptyset$, i.e., $A = FH \backslash FHX^*$ because every word of FHX^* has a proper left factor in $FH \backslash FHX^*$, hence in A. Thus A satisfies \mathfrak{U}'_{r_i} . Since $H \cap X^*H = \emptyset$ and since $HX^* \cup X^*HX^* = FHX^* \subset FAX^*$, it follows that $H \cap (HX^* \cup X^*H \cup X^*HX^*) = \emptyset$, showing that in fact H is the least set H' such that FH'F = FAF. Since \mathfrak{N}_{r_i} implies FAF = AF, it follows that AF = FHF, or, in equivalent fashion, that $T = F \backslash FHF$.

Remark 1. If A is a prefix code, one has $A^{p}X^{*} \cap A^{p} = \emptyset$ for any

positive p. Thus if $A \in J$ one has also $A^p \in J$ for any positive p because of the relations $FA^p = (FA)A^{p-1} \subset AFA^{p-1} = A(FA)A^{p-2} \subset A^2FA^{p-2} \cdots \subset A^{p-1}FA \subset A^{p-1}AF = A^pF$ which show that A^p satisfies \mathfrak{N}_{r_i} . Observing that for p > p' > 0, $FA^pX^* \subset FA^pX^*$, one concludes that $FA^pX^* \cap A^{p'} = \emptyset$. Clearly $A^p \notin J_1$ for p > 1.

As an application let us consider two words $a, a' \in A$, a word $\bar{a} \in A^m$ (where $\bar{a} \in A^\circ$ is understood to mean $\bar{a} = 1$) and two right factors t_i' and t_j' of a' such that $0 \le \deg t_i' \le \deg t_j' < \deg a'$. We verify that, provided $t_j'\bar{a}a \in A^*$, one has $\deg \tau t_i\bar{a}a \le \deg \tau t_j'\bar{a}a$. Indeed, by the definition of τ , there exist two elements u_i and u_j of X^* such that $t_i'\bar{a}u_i$, $t_j'\bar{a}u_j \in A^*$ and $u_i\tau t_i'\bar{a}a = u_j\tau t_j'\bar{a}a = a$. If $t_i' = 1$, the result is proved. Thus we can assume that none of u_i and u_j is equal to a and, as a result, both of the words $a_i = t_i'\bar{a}u_i$ and $a_j = t_j'\bar{a}u_j$ have \bar{a} as a proper factor and are proper factors of $a'\bar{a}a$. However, $\bar{a} \in A^m$; $a'\bar{a}$, $\bar{a}a \in A^{m+1}$; $a'\bar{a}a \in A^{m+2}$. Hence a_i , $a_j \in A^{m+1}$. By hypothesis t_i' is a right factor of t_j' and, thus, $a_j \in Fa_iX^*$ is excluded. It follows that either $u_i = u_j$ (and then $t_i'\bar{a}a = \tau t_j'\bar{a}a$) or u_j is a proper left factor of u_i (and then $t_i'\bar{a}a = \tau t_j'\bar{a}a$). The verification is concluded.

Remark 2. Let B be a prefix code and ξ an epimorphism (homomorphism onto) of $B^{\dagger} = \{1\} \cup B^*$ onto an abelian group G of order p > 1. We suppose that $A \in \mathbf{J}$ is contained in the kernel $B^{\dagger} \cap \xi^{-1} 1$ of ξ and we prove that under these hypotheses

- (i) There exists a prefix code C such that A = CP;
- (ii) $C \in J$ and, moreover, $C \in J_1$ if B is a.s. synchronizing.

VERIFICATION OF (i)

Let $B_0 = B \cap \xi^{-1}1$; $B_1 = B \setminus B_0$; $C = B_0^{\dagger}B_1$ where $B_0^{\dagger} = \{1\} \cup B_0^*$. Since ξ is an epimorphism, B_1 is not empty and, clearly, C is a prefix code. We call C-degree of a word $f \in B^*$, the number of its factor from B_1 and we note that no $a \in A$ has C-degree zero. Indeed, otherwise, we could take some $b \in B_1$, and, since ba has no left factor in $A \subset \xi^{-1}1$, A would not satisfy \mathfrak{N}_{r_j} .

Let $a \in A$ of minimal C-degree q and $g = c_1c_2 \cdots c_q \in C^q$. Applying \mathfrak{N}_{r_j} to c_qa shows that $c_qa = a'f'$ where $a' \in A$ and where $\xi f' = \xi c_q \neq \xi 1$. Thus $f' \in C^*B_0^{\dagger}$. Since a has minimal C-degree we must have in fact $f' \in CB_0^{\dagger}$ and a' has also C-degree q. By reiterating the argument we see that $gb \in A$ for some word $b \in B_0^{\dagger}$ which is necessarily a left factor of a. Thus $g \in A$ would have been proved if we had taken an element $a \in A \cap B_1C^*$ of C-degree q. However, choosing $c_1 \in B_1$, the relation

 $gb \in A$ shows that such an element a does exist and we can conclude that $C^q \subset A$.

It follows that $C_1 \subset \xi^{-1}u$ for some element $u \in G$ of order q. Thus, since ξ is an epimorphism, G is a cyclic group and p = q. Finally since A satisfies \mathfrak{U}_r , the relations $A \cap B_0^{\dagger} = \emptyset$, $C^p \subset A$ and $A \subset (\{1\} \cup (C^p)^*)B_0^{\dagger}$ show that $A = C^p$ and (i) is proved.

VERIFICATION OF (ii)

It suffices to show that C satisfies \mathfrak{N}_{r_j} . Assume $Fc' \subset CF$ already proved for all words $c' \in C$ of degree less than m and consider a word $c \in C$ of degree m and any $f \in F$. If c admits another word $c' \in C$ as a proper right factor we have $fc \in Fc'$ and $fc \in CF$ results from the induction hypothesis. Thus we may assume $c \in X^*C$, and we consider fc^p . By $A = C^p$ and \mathfrak{N}_{r_j} we have $fc^p = c_1c_2 \cdots c_p f'$ where $c_1 , c_2 , \cdots , c_p \in C$ and $f' \in F$. Because of the induction hypothesis, $c_p f'$ cannot be a factor of c, thus deg $c < \deg c_p f'$, and cancelling gives $fc^{p-1} = c_1c_2 \cdots c_{p-1}f''$ for some $f'' \in X^*$. In the same manner, deg $c < \deg c_{p-1}f''$; hence $fc^{p-2} = c_1c_2 \cdots c_{p-2}f'''$, and so on. Finally we obtain $fc = c_1f''''$ and $Fc \in CF$ is proved. This ends the verification.

We shall need later the following formulation of this remark: If $A \in \mathbf{J}$ and if $B = B_0 \cup B_1(B_1 \neq \emptyset)$ is a partition of a prefix code B such that $A \subset C^pB_0^{\dagger}$ where $C = B_0^{\dagger}B_1$, then $A = C^p$ and $C \in \mathbf{J}$. That $C \in \mathbf{J}_1$ when B is a.s. synchronizing is trivial.

The next remarks are not needed for the verification of the main result. Remark 3. For h, $h' \in H (= A \backslash X^*A)$, let $R_{h',h} = \{f \in X^* \backslash Fh: h'f \in Fh\}$. Thus $h'R_{h',h} = (h'F \cap Fh) \backslash Fh'hF$ is a finite set and $Fh'R_{h',h} \subset Fh$. Because of \mathfrak{A}_{r_j} and \mathfrak{A}_{r_j} , any word $f \in Fh$ has a unique maximal left factor $a \in A^*$ (since $A^* = \bigcup \{A^* \cap Fh: h \in H\}$), and, by definition, either $f = a \in A^* \cap Fh$ or $f \in aR_{h',h}$ where h' is determined by $a \in A^* \cap Fh'$. Reciprocally if $a \in A^* \cap Fh'$, one has $aR_{h',h} \subset Fh$. Thus, for each $h \in H$, one has the equation $Fh = (A^* \cap Fh) \cup \{(A^* \cap Fh')R_{h',h}: h' \in H\}$ where, as it is easily checked, every word of F appears at most once in each member. Assuming that the finite sets $R_{h',h}$ $(h, h' \in H)$ are given, this provides a system of equations from which the sets $A^* \cap Fh$ (hence A^* itself) can be computed by standard substitution methods. Another system having the same properties consists of the equation $\{1\} \cup TX = T \cup A$ and the equations

$$Th = (A \cap Fh) \cup \{(A \cap Fh') R_{h',h}: h' \in H\} (h \in H).$$

These systems are due essentially to Von Mises and to W. Feller; the relevant bibliography can be found in (Feller, 1958) and in (David and Barton, 1962).

Remark 4. Let us verify that for $A \in J$ the set $H_p = A^p \backslash X^*A^p$ is equal to $A^p \cap \tilde{A}^p$, where $\tilde{A} = HF \backslash X^*HF$. Observing that the condition $H \cap (X^*H \cup HX^* \cup X^*HX^*) = \emptyset$ on $H = A \backslash X^*A$ is symmetric and recalling that $F \backslash AF = T = F \backslash FHF$, we immediately deduce that $T = F \backslash F\tilde{A}$ and that \tilde{A} itself satisfies the relations $\tilde{A} \cap X^*\tilde{A}F = \emptyset$ and $\tilde{A}F \subset F\tilde{A}$. Thus, using $A^pT \subset (FHF)^p \subset A^pF$ and $A^pT \cap A^{p+1}F = \emptyset$, we obtain the equations

$$(FHF)^p \setminus (FHF)^{p+1} = A^p T = T\tilde{A}^p; (FHF)^p = FA^p F = F\tilde{A}^p F.$$

Since H_p is the least subset H' such that $FH'F = FA^pF$, this shows that $H_p \subset A^p \cap \widetilde{A}^p$. Further, for any $h \in A^p \cap \widetilde{A}^p$, if h = f'f'' $(f' \in X^*, f'' \in F)$, the word f'' belongs to $F \setminus FA^p$, hence it does not belong to $(FHF)^p$. This proves that $A^p \cap \widetilde{A}^p \subset H_p$, and it concludes the verification.

Remark 4bis. In view of the symmetric relation $F \setminus AF = F \setminus F\widetilde{A}$, a close connection between A and \widetilde{A} is to be expected. We verify that there exists a bijection ("1 to 1 mapping onto") $\rho: A \to \widetilde{A}$ sending each $a \in A$ on one of its conjugates (i.e., on a word of the form f''f', where $f', f'' \in F$ satisfy a = f'f''). Indeed, let $a = fh \in A$; $f \in F$, $h \in H$. If h = f'f'' where $f' \in F$ and $f'' \in X^*$, ff' belong to T, hence $ff' \notin HF$. However, for f'' = h, (and f' = 1), $f''f \in HF$. Thus h has a right factor $f'' \in X^*$ of minimal degree which is such that $f''ff' \in HF$, and, because of its minimality, $f''ff' \notin X^*HF$, i.e., $f''ff' \in \widetilde{A}$. We define $\rho a = f''ff'$.

Since another mapping $\tilde{\rho}: \tilde{A} \to A$ can be defined in a perfectly symmetric fashion and since, then, both of the mappings $\tilde{\rho}\rho: A \to A$ and $\rho\tilde{\rho}: \tilde{A} \to \tilde{A}$ are identity mappings, the remark is verified.

Remark 5. We assume here that Card $X=k<\infty$ and we define $\alpha(k,n)$ as the minimum number of words in the sets $H\in X^n$ that satisfy the condition Card $(FH\backslash FHX^*)<\infty$. For instance, $\alpha(1,n)=1$; $\alpha(k,1)=k$; $\alpha(k,2)=2^{-1}k(k+1)$; $\alpha(2,5)=9$. The exact value of $\alpha(k,n)$ is not known in the general case, but we can verify that $\alpha(k,n)\geq n^{-1}k^n$ and that, assuming k,n>1, $\lim nk^{-n}\alpha(k,n)=1$ for $\max(k,n)$

Let us recall that a word f is said to be *primitive* iff $f = f'^p$ $(f' \in F, p > 0)$ implies p = 1. The number of conjugate classes of

primitive words of degree n is

$$\psi_k(n) = n^{-1} \sum \{k^{n|d}\mu(d):d \mid n\}$$

where μ () denotes Möbius function (Moreau quoted in (Lucas, 1891)).

Verification of $\alpha(k, n) \ge n^{-1} k^n$

Observe that for $f \in X^n$ and m > 0, any factor of degree n of f^m has the form f'^d where f' is a primitive word of degree d' and dd' = n. Thus the condition that $FH \setminus FHX^*$ is finite implies that H contains a d-th power of at least one word from each conjugate class of primitive words of degree d'(dd' = n). It follows that

$$\alpha(k, n) \ge \sum \{ \psi_k(d') : d' \mid n \} = n^{-1} \sum \{ k^{n|d} \varphi(d) : d \mid n \} \ge n^{-1} k^n$$

(where $\varphi(\)$ is Euler's function) and the inequality is verified. It follows that, more generally, if $H' \in X \cup X^2 \cdots \cup X^n$ is such that $FH' \backslash FH'X^*$ is finite, one has $\sum \{k^{n-\operatorname{deg} h'} : h' \in H'\} \ge n^{-1}k^n$ since we can derive from H' a subset $H \subset X^n$ (satisfying also $FH \setminus FHX^*$ finite) by replacing each $h' \in H'$ by the set of all words $h \in X^n$ which admit h' as a left factor.

Verification of $\lim nk^{-n}\alpha(k, n) = 1$

Let \leq denote a lexicographic order on F. We use the results given in (Chen, Fox, and Lyndon, 1956) and, following these authors, we define $K \subset X^*$ by the condition that $f \in K$ iff f = f'f'' for $f', f'' \in X^*$ implies f < f''. It is known that K consists of the first word (in lexicographic order) from each conjugate class of primitive conjugate words of positive degree. Together with K we define $\bar{K} = \{ (f'f'')^p f' : f'f'' \in K; p > 0 \}$ and we verify the following statement:

If $f \in X^*$ is such that $f = f_1 f_2 = f_3 f_4$ for $0 < \deg f_1 = \deg f_4$ implies $f_1 \leq f_n$, then $f \in \bar{K}$.

PROOF: If there exists no word $g \neq 1$, h which is at the same time a left and a right factor of f, each relation $f_1 \leq f_4$ can be replaced by $f_1 < f_4$. Then, identically, $f = f_1 f_2 < f_4$, i.e., $f \in K$. Thus we have only to discuss the case where f admits some nontrivial word as a proper left and right factor. Then it is known that f has the form $(g_1g_2)^{1+p}g_1$ where $p \ge 0$ and where g_1g_2 is primitive. Let g_1' and g_2' be defined by the conditions $g_1g_2 = g_1'g_2'$ and $g_2'g_1' \in K$.

If deg $g_1' < \deg g_1$ or if p > 0, $g_2'g_1'$ is itself a factor of f. Because of

our hypothesis on f, it cannot satisfy $g_2'g_1' < g_1g_2$ (because the right factor f_4 of f beginning with $g_2'g_1'$ would be in the relation < with the corresponding left factor f_1 of the same degree). However $g_2'g_1' \in K$ implies $g_2'g_1' \leq g_1'g_2'$ (= g_1g_2). Thus $g_2'g_1' = g_1g_2$ and we have verified $f \in \bar{K}$ for this case.

Finally let us assume p=0 and $\deg g_1' \geq \deg g_1$. Without loss of generality we can further assume that g_1 has maximum degree among the words which are at the same time a proper left and a right factor of f. There exists g_3 such that $g_1'=g_1g_3$ and $g_2=g_3g_2'$. Since $f=g_1g_2g_1=g_1g_3g_2'g_1$, one has $g_1g_4 \leq g_2'g_1$ where g_2 is the left factor of degree $\deg g_2'$ of g_3g_2' . However $g_2'g_1'$ (= $g_2'g_1g_3$) $\leq g_1'g_2'$ (= $g_1g_3g_2'$) from which we conclude that $g_2'g_1 \leq g_1g_4$ and finally that $g_2'g_1 = g_rg_4$. By construction g_1g_4 is a left factor of f. Since we have assumed g_1 to be the common left and right factor of maximal length of f, we must have $g_2'=g_4=1$, hence $g_1g_2=g_1'g_2'=g_2'g_1'\in K$ and the verification is concluded. In fact, K is the set of all left factors of the elements of K.

Consider now $H = \bar{K} \cap X^n$. Each long enough word $s = x_{i_1} x_{i_2} \cdots x_{i_m}$ contains at least one factor $f = x_{i_j} x_{i_{j+1}} \cdots x_{j+n-1} \in X^n$ $(j \leq m-2n+2)$ which is such that $f \leq x_{i_j}, x_{i_{j'+1}} \cdots X_{i_{j'+n-1}}$ for $j \leq j' \leq j+n-1$. What we have just proved shows that $f \in K$. Thus $FH \setminus FHX^*$ is finite.

In a similar manner, let H' consist of all words of the form x^n ($x \in X$) and of the form xh where $x \in X$, $h \in R \cap X^{n-1}$ and h < x. We also have Card $(FH'\backslash FH'X^*) < \infty$. Indeed as it is easily verified, when $x_{i_{j+1}} x_{i_{j+2}} \cdots x_{j+n-1} \in \bar{K} \cap X^{n-1}$, one has $x_{i_j} x_{i_{j+1}} \cdots x_{j+n-1} \in H'$ or $x_{i_j} x_{i_{j+1}} \cdots x_{i_{j+n-2}} \in K \cap X^{n-1}$ depending upon $x_{i_{j+1}} < x_{i_j}$ or not. Thus $\alpha(k, n) \leq \min (\operatorname{Card} H, \operatorname{Card} H')$ identically.

Now, since Card $K \cap X^m = \psi_k$ $(m) \leq m^{-1}k^m$ we have Card $H = \sum_{0 < m \leq n} \psi_k(m) \leq n^{-1}k^n (1 + \sum_{0 < m < n} n m^{-1} k^{m-n})$ from which it follows that for each $\epsilon > 0$ there exists $k_{\epsilon} < \infty$ such that $n k^{-n} \alpha(k, n) < 1 + \epsilon$ for all n and $k > k_{\epsilon}$.

On the other hand,

Card $H' \leq k + (k-1)$ Card $(\bar{K} \cap X^{n-1})$

$$\leq k + (k-1) \sum_{0 \leq m \leq n} m^{-1} k^m = n^{-1} k^n (n(n-1)^{-1} + u_n)$$

where u_n is determined inductively by $u_3 = (2k)^{-1}$ and $u_{n+1} = k^{-1}$ $(u_n + (n-1)^{-1} (n-2)^{-1})$. Since $\lim_{n\to\infty} u_n = 0$, there exists, for each $\epsilon > 0$ and k > 1 some $n_{k,\epsilon} < \infty$ such that $n k^{-n} \alpha(k,n) < 1 + \epsilon$ for all $n > n_{k,\epsilon}$. The verification of Remark 5 is concluded.

III. VERIFICATION OF THE MAIN PROPERTY

We intend to show that if $A \in J \setminus J_1$ there exists another prefix code $C \in \mathbf{J}_1$ and a natural number p > 1 such that $A = C^p$. For this, let $A \in \mathbf{J}$ and observe that condition \mathfrak{N}_{r_i} expresses that for each $a \in A$ and $f \in F$, $\tau \bar{r}a$ is a right factor of a. Thus Card $\tau Ta \leq \deg a$. Recalling the notations introduced at the beginning of Section II, this proves that p is finite and, since p = 1 is equivalent to $A \in J_1$, we assume now p > 1. Letting $Q = \{a \in A^* : \text{Card } \tau Ta = p\}$, we know that $FQF \cap A^* \subset Q$ and, for each $f \in F$ and $q \in Q$, $\tau Tfq = \tau Tq$. The p elements of τTq indexed by increasing degree will be denoted by $\tau_0 q$ (=1 since $Q \subset A^*$), $\tau_1 q, \dots, \tau_{p-1} q$. We shall use repeatedly the fact that an equation like $\tau fq = \tau_i g$ is equivalent to the existence of an element $\bar{a} \in A^*$ such that $fq = \bar{a}\tau_i fq.$

We verify first a few easy consequences of the definitions.

3.1 For all $q, q' \in Q$ and $i \in [0, p-1]$, $\tau((\tau_i q)q') = \tau_i q'$. Proof: Because of the relation $\tau Tq' = \tau Tqq' = \tau((\tau Tq)q')$ and the fact that τTq and $\tau Tq'$ have the same finite cardinality, there corresponds to each $\tau_i q' \in \tau T q'$ one and only one element, say $\tau_i' q$, of $\tau T q$ that satisfies $\tau((\tau_{i'}q)q') = \tau_{i}q'$. However, we have $q' = \bar{a}'a'$ where $\bar{a}' \in \{1\} \cup A^*, \ a' \in A \text{ and the elements } \tau_i q' \text{ are right factors of } a'. A$ similar observation can be made for q. Thus by Remark 1, we know that $i \leq j$ implies $i' \leq j'$. Thus i = i' identically and 3.1 is proved. In fact, if f is any right factor of $\tau_{i}q$, Remark 1 shows that $\tau fq' = \tau_{i}$, q' where $j' \leq j$. Thus, denoting by \bar{B}_j the set of all words \bar{f} that satisfies the

(*) for each $q' \in Q$, $\tau \bar{f}q' = \tau_j q'$; (**) for each $q' \in Q$ and right factor f of \bar{f} , deg $\tau fq' \leq \deg \tau \bar{f}$ we have proved that $\{\tau_j q : q \in Q\} \subset \bar{B}_j$, identically. As a consequence we have 3.2 $A \subset \bar{B}_1^p$.

PROOF: Let $a \in A$ and, taking a fixed $q \in Q$, let the p words u_1 , u_2 , $\tau_3 qa; \cdots; u_{p-1} u_{p-2} \cdots u_2 u_1 = \tau_{p-1} qa; u_p u_{p-1} \cdots u_2 u_1 = a.$ By 1 we know that, for each $i \in [0, p - 1]$, $\tau((\tau_i qa)qa) = \tau_i qa$, or, in other terms, that the word $q_i = u_i u_{i-1} \cdots u_1 qau_p u_{p-1} \cdots u_{i+1}$ belongs to A^* . In fact since it admits q as a factor, it belongs to Q. Since for $i \in$ $u_2 u_1$) = $u_{i+1} u_i \cdots u_2 u_1$, it follows that $\tau_1 q_i = u_{i+1}$ for $i = 1, 2, \cdots$, p-1 and 3.2 is proved.

Using the same notations it is readily seen that if $b, b' \in \bar{B}_1$ then, $bb' \in \bar{B}_2$

 \bar{B}_1 . Indeed we have $b'qa = q_1'u_1$ with $q_1' \in Q$ and $bb'qa = bq_1'u_1 =$ $q_2'u_2u_1$ with $q_2'\in Q$ showing that $\tau bb'qa= au_2qa$. In similar fashion if b $\in \bar{B}_0$ and $b' \in \bar{B}_1$ it is easily seen that $bb', b'b \in \bar{B}_0$.

3.3 If $f \in F$ and $q, q' \in Q$ are such that $\tau f q = \tau_1 q$ and $\tau f q' = \tau_0 q$, then $bf \in \bar{B}_0 \cup \bar{B}_1$ for any $b \in \bar{B}_0 \cup \bar{B}_1$.

Proof: As said before $\tau fq' = \tau_0 q$ is equivalent to the hypothesis that fg' is a certain element, say q_0 , of Q. In similar manner, using 3.1, fq = $\tau_1 q$ implies that $\tau f q q = \tau_1 q$ i.e. that $f q q = q_1 \tau_1 q$ where $q_1 \in Q$.

Let $b \in \bar{B}_0$. This implies $bq_0, bq_0 \in A^*$. Thus $\tau bq_0 = \tau bf^*q' = \tau_0 q'$ and

 $\tau bfqq = \tau(q_1\tau_1q) = \tau_1q$ showing that $bf \in \bar{B}_0 \cup \bar{B}_1$. Let $b \in \bar{B}_1$. This implies $bfq' = bq_0 = q_0' \tau_1q_0 = q_0'\tau_1q'$ and $bq_1 = q_1'\tau q_1$ where $q_0', q_1' \in Q$. Thus $\tau bfq' = \tau_1q'$ and $\tau bfqq = \tau((q_1'\tau q_1)\tau_1q) = \tau(q_1'\tau q_1)\tau_1q'$ $\tau((\tau_1q_1) \ (\tau_1q) \neq \tau_1q$ showing again that $bf \in \bar{B}_0 \cup \bar{B}_1$ and concluding the proof of 3.3.

This practically ends the verification of our main property. Let B = $(\bar{B}_0 \cup \bar{B}_1) \setminus (\bar{B}_0 \cup \bar{B}_1) X^*$. By construction B is a prefix code. Further, if b and bf are two elements of $\bar{B}_0 \cup \bar{B}_1$, the same must be true of f because of 3 and the fact that if condition (***) is satisfied by f it is also satisfied by f. Thus $\bar{B}_0 \cup \bar{B}_1 \subset B^*$. Let now $B_0 = \bar{B}_0 \cap B$ and $B_1 = \bar{B}_0 \cap B$ $\bar{B}_1 \cap B$. Using the remarks made at the end of 2 we obtain $\bar{B}_0 \subset {B_0}^*$ and $\bar{B}_1 \subset B_1 \cup B_0^* B_1 \cup B_1 B_0^*$. Thus, by Remark 2 and 3.2, $A = C^p$ where the prefix code $C = B_1 \cup B_0^*B_1$ belongs to J. Finally, taking a word $a \in A$ of the form $a = b^p$ where $b \in B_1$, we have $\tau_i f a = b^i \in C^*$ for all $f \in F$. Thus the parameter p associated with C has value 1, that is, $C \in \mathbf{J_1}$ and the proof is concluded.

IV. AN ALTERNATIVE VERIFICATION OF THE MAIN PROPERTY

A more systematic verification of the main property can be given if one uses the theory of monoids instead of insisting on a self-contained argument as it was done above. It will appear that the main property follows instantly from Remark 1 and Remark 2 once proved the simple Property 2 below.

We recall first without proof some classical results on the minimal ideals of a monoid and some of their more or less obvious consequences. The reader is referred for more details to the existing literature and especially to (Clifford and Preston, 1961).

Let us recall that a homomorphism φ of F onto a quotient monoid is said to be compatible with a subset F' of F iff $\varphi^{-1}\varphi F' = F'$. To each $F' \subset F$ one can associate a maximal compatible homomorphism $\varphi = \varphi_{F'}$

by the condition that φF is a homomorphic image of $\varphi' F$ for any homomorphism φ' compatible with F' (Teissier, 1951).

Consider now a set $A \subset X^*$ and, letting $A^{\dagger} = \{1\} \cup A^*$ and $\varphi = \varphi_{A^{\dagger}}$, assume that the following conditions are satisfied:

- (\mathfrak{U}_d). For all $f \in F \setminus A^{\dagger}$, $fA^{\dagger} \cap A^{\dagger}f \cap A^{\dagger} = \emptyset$. (\mathfrak{N}_d). For all $f \in F$, $A^{\dagger} \cap FfF \neq \emptyset$.
- (\mathfrak{M}_k) . φF admits minimal right ideals R_i $(i \in I)$ and minimal left ideals L_j $(j \in J)$.

Let $I' = \{i \in I : R_i \cap \varphi A^{\dagger} \neq \emptyset\}, J' = \{j \in J : L_j \cap \varphi A \neq \emptyset\}$ and select arbitrarily a pair of indices—(say (1, 1)) in $I' \times J'$. It is classical (Suschkewitsch, 1928) that there exists an isomorphism γ of $R_1 \cap L_1$ onto a group G (which will be identified with a basis of its ring over the integers). Letting $e_{i,j}$ denote the idempotent contained in $R_i \cap L_j$ we define the $J \times I$ matrix Γ by $\Gamma_{j,i} = \gamma(e_{1,j} \cdot e_{i,1})$.

It follows instantly from the hypothesis that there exists an isomorphic representation of φF by pairs of matrices $(\mu f, \nu f)$ where μf (resp. νf) is a $J \times J$ (resp. $I \times I$) matrix with entries in $\{0\} \cup G$, and that one has:

4.1.1. For all $f \in F$, $\mu f \cdot \Gamma = \Gamma \cdot \nu f$.

Consider the restriction Γ' of Γ to $J' \times I'$ (i.e., let Γ' be a $J' \times I'$ matrix such that $\Gamma'_{j,i} = \Gamma_{j,i}$ for $j \in J'$, $i \in I'$); let μ' and ν' be the restrictions of μ and ν to $J' \times J'$ and to $I' \times I'$ respectively. There exists a minimal sub-group G' of G that has the following properties:

4.2. For each $a \in A^{\dagger}$, $\mu'a \cdot \Gamma' = \Gamma' \cdot \nu'a$ and all the entries of Γ' , $\mu'a$

- and $\nu'a$ $(a \in A^{\dagger})$ belong to $\{0\} \cup G'$.
- 4.3. The only invariant subgroup of G contained in G' is the trivial subgroup {e} consisting of the neutral element e of G.
- 4.4. A^{\dagger} consists of all the words $f \in F$ such that both $\mu'f$ and $\nu'f$ have at least one entry in G'.

It is useful to note that since Γ and Γ' have all their entries in G, 4.1 and 4.2 imply that μf and $\mu' a (a \in A \dagger)$ (resp. νf and $\nu' a$) have one and only one nonzero entry in each row (resp. column).

One can also observe that for $f \in R_i \cap L_j$ the matrix μf (resp. νf) has its jth column (resp. ith row) equal to the ith column (resp. jth row) of Γ multiplied on the right (resp. left) by $\gamma(e_{1,1} \cdot \varphi f \cdot e_{1,1})$. Finally, because of 4.3, one has G = G' iff $G = \{e\}$, that is iff there exists at least one word $a \in A^{\dagger}$ such that $aFa \subset A^{\dagger}$.

We assume henceforth that A is a prefix code, i.e., that A^{\dagger} satisfies

 (\mathfrak{A}_r) For all $f \in F \setminus A^{\dagger}$, $A^{\dagger} f \cap A^{\dagger} = \emptyset$ which is obviously more restric-

tive than (\mathfrak{U}_d) . Because of (\mathfrak{N}_d) and (\mathfrak{M}_k) , (\mathfrak{U}_r) is equivalent to

 (\mathfrak{N}_r) For each $f \in F$, $fF \cap A^{\dagger} \neq \emptyset$ (that is, I = I').

It is known that a consequence of I' = I is that μ gives an isomorphic representation of φF . Thus $\mu^{-1}\mu A^{\dagger} = A^{\dagger}$ and, for any $b \in \varphi F$, we can write μb instead of the more cumbersome $\mu \varphi^{-1}b$.

On the other hand, we shall see that $\nu^{-1}\nu A^{\dagger} \subset A^{\prime\dagger}$ where A^{\prime} is a prefix code such that $A \subset A'^*$ and that the following is true.

4.5. $\gamma(R_1 \cap L_1 \cap \varphi A^{\prime\dagger}) = G' \text{ and } J'' = \{j \in J : L_j \cap \varphi A^{\prime\dagger} \neq \emptyset\} \text{ is }$

the set of all $j \in J$ such that $\Gamma_{j,i} \in G'$ for each $i \in I$. PROOF: By construction $A^{\dagger} \subset \nu^{-1} \nu A^{\dagger}$. Since I' = I implies $\nu = \nu'$, the properties 4.1 and 4.2 show directly that $f \in \nu^{-1} \nu A^{\dagger}$ if all the entries of νf belong to $\{0\} \cup G'$. This proves that the set of all $f \in F$ satisfying this last condition is a submonoid, say A'^{\dagger} , of F having the property 4.5 since then, $L_i \cap \varphi A'^{\dagger} \neq \emptyset$ iff $e_{i,j} \in \varphi A'^{\dagger}$ for each $i \in I$.

For the sake of completeness we verify that A^{\dagger} satisfies (\mathfrak{A}_r) . Let $f' \in F \backslash A'^{\dagger}$, i.e. let f be such that, e.g., $(\nu f')_{i,i'} \in G \backslash G'$. Since every matrix νf $(f \in F)$ has one and only one zero entry in each column, it follows that for each $f \in \nu^{-1} \nu A'^{\dagger}$ one has $(\nu \text{ ff}')_{i'',i'} \in G \backslash G'$ for at least one $i'' \in I$. Thus $A'^{\dagger}(F \backslash A'^{\dagger}) \subset F \backslash A'^{\dagger}$ and the verification is concluded.

It follows from the properties of φ that A' = A iff ν is an isomorphic representation and that under the present hypothesis $A^{\prime\dagger}$ can be defined directly as the set of all $f \in F$ such that $fA^{\dagger}a \subset A^{\dagger}$ for at least one $a \in A^{\dagger}$. Clearly $A'^{\dagger} = F$ iff A is a.s. synchronizing.

Property 2. If G is an abelian group there exists an a.s. synchronizing code B and an epimorphism $\xi: B^{\dagger} \to G$ such that $A \subset B^* \cap \xi^{-1}e$.

Proof: Let $L'' = \bigcup \{L_j: j \in J''\}$ and consider an element $b \in \varphi \bar{B}$ where $B = \{ f \in F: L'' \varphi f \cap L'' \neq \emptyset \}$. In equivalent manner, consider an element $b \in \varphi \Gamma$ such that $(\mu b)_{j,j'} = u \in G$ for at least one pair $j,j' \in J''$.

For any other $j'' \in J''$ let $j''' \in J$ and $v \in G$ be defined by $(\mu b)_{j'',j'''} =$ v. Now, the (j, 1) and the (j'', 1) entries of μbe_{11} are respectively equal to u and v since $\mu e_{1,1}$ has all its non zero entries equal to e (and located in its first column). On the other hand, if i'' is defined by $be_{1,1} \in R_{i''}$, these two entries are equal respectively to $\Gamma_{i,i} w$ and $\Gamma_{ir,i} w$ for some $w \in G$. Since $j, j'' \in J''$, and since the hypothesis that G is abelian implies $G'=\{e\}$ (because of 4.3), and consequently $\Gamma_{\bar{\imath},\bar{\imath}}=e$ for $\bar{\jmath}\in J'',\ \bar{\imath}\in I$, we have $\Gamma_{j,i}=\Gamma_{j'',i}=e$. Consequently u=v=w. We set $\xi\varphi^{-1}b=u$.

Consider now an arbitrary $i \in I$ and define $i' \in I$ by $b e_{i,1} \in R_{i'}$. The same argument shows that the (j, 1) entry, and the (j'', 1) entry

of μb $e_{i,1}$ are respectively equal on the one hand to $u\Gamma_{j',i}$ and $v\Gamma_{j''',i}$ and on the other hand to $\Gamma_{j,i'}$ w' and $\Gamma_{j'',i'}$ w' for some $w' \in G$.

Again, $\Gamma_{j',i} = \Gamma_{j'',i'} = e$ and, using u = v shows that $\Gamma_{j''',i} = e$. Since this is true for all $i \in I$, we conclude from 4.5 that $j''' \in J''$. Thus, we have proved that $\bar{B} = \{ f \in F: L'' \cdot \varphi f \subset L''' \}$.

It is classical (Dubreil, 1953) that this relation implies $\bar{B} = B^{\dagger}$ where B is a prefix code. Indeed, if $b, b', bb'' \in \varphi \bar{B}$, we have $L'' \cdot bb' = (L''b)b' \subset L''b' \subset L''$ (showing $bb' \in \bar{B}$, hence $\varphi \bar{B} \varphi \bar{B} \subset \varphi \bar{B}$) and $L''b'' \supset (L''b)b'' = L''bb'' \subset L''$ (showing $b'' \in \varphi \bar{B}$, hence that $\varphi \bar{B}$ satisfies \mathfrak{A}_r).

To complete the verification, we observe that $\varphi F \cdot e_{1,1} \subset \varphi \overline{B}$. Thus \overline{B} is a.s. synchronizing.

Since $L'' \cdot \varphi \bar{B} \subset L''$ and since μ is a representation of F, we have $\xi bb' = \xi b \cdot \xi b'$ for any $b, b' \in \bar{B}$. Finally, $A \subset \bar{B} \cap \xi^{-1}e$ follows from 4.3, $G' = \{e\}$, and $J' \subset J''$.

Let now $A \in J$. In order to be able to apply Property 2 and Remark 2, we need to show that φF satisfies (\mathfrak{M}_k) and that G is abelian.

Recalling the notations introduced at the beginning of Section 2, we define a homomorphism φ' of F by the condition that, for any $f, f' \in F$, $\varphi'f = \varphi'f'$ iff $\tau tf = \tau tf'$ for each $t \in T$. Clearly, φ' is compatible with A^{\dagger} and, consequently, φF is a homomorphic image of $\varphi'F$. Further, since Card $\tau Tf' \leq \text{Card } \tau Tf$ for any $f \in F$ and $f' \in FfF$ and since Card $\tau Ta \leq \text{deg } a$ for any $a \in A$, the ideal $\varphi' FAF$ of $\varphi'F$ contains no infinite strictly decreasing sequence of one sided ideals. It follows immediately that φF satisfies (\mathfrak{M}_k) and that any group in φFAF is the homomorphic image of at least one group in $\varphi'F$.

We show that any group H' in $\varphi'F$ is a finite cyclic group. Of course, this is equivalent to *Remark 1* but it can also be verified directly as follows.

Let $H = \varphi'^{-1}H'$. The hypothesis that H' is a group is equivalent to the existence of a subset $T' \subset T$ such that $\tau Th = \tau T'h = T'$ for any $h \in H$.

Let $\varphi'h'$ $(h' \in F)$ be the neutral element of H'. We have $t = \tau th'$ for every $t \in T'$. Thus $h' \in FAF$ and, by \mathfrak{N}_j , T' is a set of $r < \infty$ right factors of h'. Further $\tau T'h = T'$ for $h \in H$, implies that $r = \operatorname{Card} \tau T'f$ for each $f \in K = \{f' \in F : \mathbf{f}'F \cap \mathbf{A} \neq \emptyset\}$. Thus, for $f \in K$, we can index the r elements of T' and $\tau T'f$ by increasing degree and define a permutation $\gamma'f : i \to i'$ of [1, r] by the identical relation $\tau t_i f = t'_i$, $(t'_i, \in \tau T'f)$. Clearly, $\gamma'ff' = \gamma'f\gamma'f'$ when $ff' \in H$ and H' is isomorphic to $\gamma'H$. Since T' is a set of right factors of h', a straightforward application of \mathfrak{N}_{r_i}

shows that $\gamma' f$ is a cyclic permutation when $f \in X \cap K$ and the verification is concluded.

RECEIVED: June 7, 1963

References

- CHEN, K. T., FOX, R. H., AND LYNDON, R. C. (1958), Free differential calculus. Ann. Math. 68, 82-86.
- CLIFFORD, A. H., AND PRESTON, G. B. (1961), The algebraic theory of semi groups. Math. Surveys of the Am. Math. Soc., Providence, R. I.
- DAVID, F. N., AND BARTON, R. A. (1962), "Combinatorial Chance." (London).
- Dubreil, P., (1953), Contributions à la théorie des demi-groupes III. Bull. Soc. Math. France 81, 289-306.
- Feller, W., (1958), "An Introduction to Probability Theory and Its Applica-
- tions," Chap. 13, sect. 7 and 8. Wiley, New York. Lucas, E. (1891), "Théorie des Nombres," pp. 501–503. Gauthiers Villars, Paris. Mandelbrot, B., (1957), "Linguistique Statistique Macroscopique en Logique, Language et Théorie de l'Information," by L. Apostel, B. Mandelbrot, and A. Morf. P.U.F., Paris.
- MANDELBROT, B., (1961), On the theory of word frequences and on related Markovian models of discourse, in "Structure of language and its mathematical aspects," Twelfth Symp. Appl. Math., Am. Math. Soc., Providence, Rhode Island.
- NEUMAN, P. G. (1962), Efficient error limiting codes. IRE Trans. Inform. Theory 8, 292-304.
- NEUMAN, P. G. (1962), On a class of efficient error limiting codes. IRE Trans. Inform. Theory. 8, 260-266.
- NEUMAN, P.G. (1963), On error limiting variable length codes. IRE Trans. Inform. Theory 9, p. 209.
- Suschkewitsch, A. (1928), Ueber die endlichen Gruppen ohne das Gesetz der eindeutigen Umkehrbarkeit. Math. Ann. 99, 30-50.
- Teissier, M. (1951), Sur les équivalences régulières dans les demi groupes. C. R. Acad. Sci. 232, 1987-1989.
- WINOGRAD, S. (1962), Bounded transient automata. Proc. A.I.E.C. 3rd Switching Theory and Logical Design, pp. 138-141.
- Winograd, S. (1963), Input error limiting automata. IBM Res. Rept. RC 966.