RAIRO. INFORMATIQUE THÉORIQUE

ANTON ČERNÝ

On a class of infinite words with bounded repetitions

RAIRO. Informatique théorique, tome 19, nº 4 (1985), p. 337-349

<http://www.numdam.org/item?id=ITA_1985__19_4_337_0>

R.A.I.R.O. Informatique théorique/Theoretical Informatics
(vol. 19, n° 4, 1985, p. 337 à 349)

ON A CLASS OF INFINITE WORDS WITH BOUNDED REPETITIONS (*)

by Anton ČERNÝ [1]

Communicated by J. BERSTEL

Abstract. — *The well-known sequence of Thue and Morse contains no factor of the form xvxvx, x being a letter and v a word. In the present paper an analogical property of a class of sequences called generalized words of Thue-Morse is proved.*

Résumé. — *La suite bien connue de Thue et Morse ne contient pas de facteur de la forme xvxvx, avec x une lettre et v un mot. Dans le présent article, nous démontrons une propriété analogue pour une classe de suite appelées mots de Thue-Morse généralisés.*

1. INTRODUCTION

Axel Thue in his remarkable works [Th 06], [Th 12] on infinite sequences of symbols has shown the existence of infinite words over three-letter alphabet without squares of non-empty words as factors. The construction of such a word in [Th 12] is based on an infinite sequence $\underline{t}$ over the two-letter alphabet $\{0, 1\}$, not containing a factor of the form $xvxvx$, x being a letter and v a word. The i-th symbol of $\underline{t}$ can be described as the parity of occurrences of the symbol 1 in binary notation of the natural number i (however, this is not the way of its description in [Th 12]). The same sequence $\underline{t}$ appears in the work of Morse [Mo 21] on symbolic dynamics. Therefore we shall call $\underline{t}$ the sequence of Thue-Morse.

(*) Received December 1983, revised February 1985.

([1]) The presented results have been achieved during the author's stay at the Institut de Programmation, Université P. et M. Curie, Paris-VI.

In the sense of Cobham [Co 72], $\underline{t}$ is a simple example of a uniform tag sequence. In [CKM-FR 80], where some algebraic properties of uniform tag sequences are investigated, generalized sequences of Thue-Morse are introduced. In such a generalized sequence the i-th symbol denotes the parity of occurrences of some fixed factor w over $\{0, 1\}$ in binary notation of i. (In fact, a slightly stronger generalization is given in [CKM-FR 80]). In this paper we show that in such generalized words of Thue-Morse all the factors are of a bounded power. More precisely, there are no factors of the form:

$$(xv)^{2^{|w|}}x.$$

2. NOTATIONS AND DEFINITIONS

Let A^* be the free monoid generated by a finite alphabet A, with the neutral element ε. Let $A^+=A^*-\{\varepsilon\}$. Let A^ω be the set of all infinite (to the right) sequences of elements of A. Let $A^\infty=A^*\cup A^\omega$. The elements of A^∞ will be called *words* (finite or infinite). A word $x\in A^*$ is a factor of a word $y\in A^\infty$, iff $y=zxt$ for some $z\in A^*$, $t\in A^\infty$. x is called *initial/terminal/ {proper} factor* of y iff $z=\varepsilon$ $/t=\varepsilon/$ $\{zt\neq\varepsilon\}$. The length $|x|$ of a finite word x is the number of its symbols, $|\varepsilon|=0$.

Let $\varphi: A^*\to B^*$ be a morphism of monoids. φ can be extended to the mapping $\varphi: A^\infty\to B^\infty$ satisfying:

$$\varphi(xy)=\varphi(x)\varphi(y),$$

for all $x\in A^*$, $y\in B^\infty$. φ is called *prolongable in* $a\in A$ iff $\varphi(a)=ax$ for some $x\in A^+$. In this case for each $n\geqq 0$ $\varphi^n(a)$ is a proper initial factor of $\varphi^{n+1}(a)$. There exists a limit:

$$\underline{z}=\lim_{n\to\infty}\varphi^n(a)\in A^\omega,$$

such that each $\varphi^n(a)$ is an initial factor of $\underline{z}$. Moreover, $\underline{z}$ is a fixpoint of φ, i. e. $\varphi(\underline{z})=\underline{z}$. A morphism φ is called *m-uniform* for some $m\geqq 0$ iff $|\varphi(b)|=m$ for all $b\in A$.

Let $\mu: A^i\to A^j$ be a mapping, $i, j\geqq 1$. μ can be extended to the mapping $\mu: A^\omega\to A^\omega$ defined by:

$$\mu(x_0x_1x_2\ldots)=y_0y_1y_2\ldots,$$

where:

$$y_{k.j}y_{k.j+1}\ldots y_{k.j+j-1}=\mu(x_{k.i}x_{k.i+1}\ldots x_{k.i+i-1}),$$

for all $k\geqq 0$. This extension is called *(i, j)-substitution.*

We follow [Co 72] where two devices for formal description of infinite words are used — uniform tag systems and sorting automata.

A *tag system* is a quintuple $T=(\Sigma, a, \sigma, \Gamma, \tau)$, where Σ and Γ are alphabets, $\sigma:\Sigma^* \to \Sigma^*$ is a morphism prolongable in $a\in\Sigma$, $\tau:\Sigma^*\to\Gamma^*$ is a morphism such that $\tau(\Sigma)\subseteq\Gamma$. The *internal /external/ tag sequence* generated by T are respectively:

$$\mathrm{intseq}_T=\lim_{n\to\infty}\sigma^n(a),$$

$$\mathrm{seq}_T=\lim_{n\to\infty}\tau(\sigma^n(a))=\tau(\mathrm{intseq}_T).$$

The tag system and the corresponding sequences are called *m-uniform* iff σ is m-uniform.

Let $m>0$. Denote $[m]=\{0, 1, \ldots, m-1\}$. A *sorting automaton* over $[m]$ is a quintuple $A=(S, \delta, s_0, F, G)$, where S is a finite set (of states), $s_0\in S$ is the initial state, $\delta: S\times[m]\to S$ is the transition function satisfying $\delta(s_0, 0)=s_0$, G is an alphabet, and $F=\{F_g\}_{g\in G}$ is a disjoint partition of S. δ can be extended to the domain $S\times[m]^*$ by setting $\delta(s,\varepsilon)=s$, $\delta(s, xd)=\delta(\delta(s, x), d)$ for $s\in S$, $x\in[m]^*$, $d\in[m]$. The *state /sorting/ sequence* of the automaton A is defined by:

$$\mathrm{state}_A=y_0\,y_1\ldots\in S^\omega,$$

$$/\mathrm{sort}_A=x_0\,x_1\ldots\in G^\omega/,$$

where $y_i=\delta(s_0, i_{[m]})$, $i_{[m]}$ being the m-ary expansion of the integer i (since $\delta(s_0, 0)=s_0$, there are no problems with leading zeros), and where x_i is the letter g such that $y_i\in F_g$.

Thus the sorting automaton is a slight generalization of the notion of finite automaton which in fact sorts to two classes of objects (accepted-rejected).

The relation between tag systems and sorting automata can be expressed as in the following proposition.

PROPOSITION 1 [Co 72]: *Let $T=(\Sigma, a, \sigma, \Gamma, \tau)$ be an m-uniform tag system and let $A=(\Sigma, \delta, s_0, F, \Gamma)$ be a sorting automaton over $[m]$ such that $\delta(s, i)=i$-th symbol of $\sigma(s)$, $s_0=a$, and $s\in F_g$ iff $\tau(s)=g$, where $s\in S$, $i\in[m]$, $g\in\Gamma$.*

Then $\mathrm{intseq}_T=\mathrm{state}_A$, $\mathrm{seq}_T=\mathrm{sort}_A$.

Finally, let us define the generalized words of Thue-Morse. Let $w\in\{0, 1\}^*-0^*$. Denote:

$$\underline{a}_w=a(0)\,a(1)\,a(2)\ldots$$

the infinite word with the i-th symbol:

$$a(i)=\#_w(i_{[2]})\bmod 2,$$

where $\#_w(x)$ denotes the number of occurrences of the factor w in the word x and $i_{[2]}$ is the binary notation of i with at least $|w|$ leading zeros. For example, 000010101010101 contains five occurrences of the factor 0101.

From [CKM-FR 80] we know the following important property of the words $\underline{a}_w$.

PROPOSITION 2 [CKM-FR 80]: *Let* $w \in \{0, 1\}^* - 0^*$, *let* μ *be a* $(2^{|w|-1}, 2^{|w|})$-*substitution on* $\{0, 1\}^*$ *defined by:*

$$\mu(x_0 x_1 \dots x_{2^{|w|-1}-1}) = y_0 y_1 \dots y_{2^{|w|}-1},$$
$$y_i = (x_{i/2} + \chi_w(i)) \bmod 2, \qquad i \in \{0, 1, \dots, 2^{|w|}-1\},$$

where $\chi_w(i) =$ *if* w *is a terminal factor of* $i_{[2]}$ *then* 1 *else* 0.

Then $\mu(\underline{a}_w) = \underline{a}_w$.

As one can easily see, there is exactly one $j \in \{0, 1, \dots, 2^{|w|}-1\}$ such that $\chi_w(j) = 1$.

In the case $w = 1$ we obtain:

$$\underline{a}_1 = \underline{t} = 0110100110010110\dots$$

the word of Thue-Morse. It is well known that $\underline{t}$ does not contain any factor of the form $xvxvx$, $x \in \{0, 1\}$, $v \in \{0, 1\}^*$. In particular, $\underline{t}$ contains no cubes $x^3 = xxx$.

3. PROOF OF THE RESULT

Our goal is to prove the following theorem:

THEOREM 1: $\underline{a}_w$ *does not contain any factor of the form:*

$$(xv)^{2^{|w|}} x \qquad \textit{where} \quad x \in \{0, 1\}, \quad v \in \{0, 1\}^*.$$

In the case $w = 1$ *the theorem states the well-known property of the sequence of Thue and Morse, thus in the following we consider* $w \in \{0, 1\}^* - 0^*$ *to be a fixed word of length at least* 2.

The proof is based on the method from [Pa 81]. The proof is divided to a series of lemmas. In the first of them, the minimal sorting automaton for $\underline{a}_w$ is described. Since the notion of the sorting automaton is derived directly from the notion of the finite automaton, the results from the theory of finite automata concerning the minimality can be applied to sorting automata, too. This fact is used in the proof of the first lemma.

LEMMA 1: *Let* $A_w=(S, \delta, s, (F_0, F_1), \{0, 1\})$ *be the sorting automaton over* $\{0, 1\}$, *where:*

$$S=\{\langle \alpha \rangle_0, \langle \alpha \rangle_1 \mid \alpha \text{ is a proper initial factor of } w\},$$

$$\delta(\langle \alpha \rangle_i, x)=\begin{cases} \langle \alpha x \rangle_i & \text{if } \langle \alpha x \rangle_i \in S,\\ \langle \alpha' \rangle_{1-i} & \text{if } \alpha x = w,\\ \langle \alpha' \rangle_i & \text{otherwise,}\end{cases}$$

where $i\in\{0, 1\}$, $x\in\{0, 1\}$, α' *is the longest proper terminal factor of* αx, *being a proper initial factor of* w.

$s=0^k$, *where* 0^k, $k\geqq 0$ *is the longest initial factor of* w *not containing* 1.

$F_i=\{\langle \alpha \rangle_i\}$, $i=0, 1$.

Then A_w *is minimal among the sorting automata with the sorting sequence* a_w.

Proof: By induction on $|z|$ one can easily show for $z\in\{0, 1\}^*$, $i=0, 1$: $\delta(s_0, z)=\langle \alpha \rangle_i$ iff $\#_w(0^{|w|}z)\equiv i \pmod 2$ and α is the longest terminal factor of $0^{|w|}z$, being the proper initial factor of w.

The nonequivalence of each pair of distinct states is evident. (Two states s_1, s_2 are equivalent iff for each $x\in\{0, 1\}^*$ $\delta(s_1, x)\in F_0$ iff $\delta(s_2, x)\in F_0$.)

It is sufficient to show the accessibility of each state (from the initial state). Obviously, each state is accessible from $\langle \varepsilon \rangle_0$ or $\langle \varepsilon \rangle_1$. On the other hand, the states $\langle \alpha_0 \rangle_0$, $\langle \alpha_0 \rangle_1$, where α_0 is the longest proper initial factor of w, are accessible. Thus it is sufficient to show that $\langle \varepsilon \rangle_i$ is accessible from $\langle \alpha_0 \rangle_i$, $i=0, 1$. The proof is based on induction. For each $\alpha\in\{0, 1\}^*$, $\alpha\neq\varepsilon$ a word $\gamma\in\{0, 1\}^*$ is given such that:

$$\delta(\langle \alpha \rangle_i, \gamma)=\langle \alpha' \rangle_i \quad \text{and} \quad |\alpha'|<|\alpha|.$$

Let $w=\alpha\beta$. We consider several cases.

1. $w=1^k$, $k\geqq 1$. Then $\gamma=0$.

2. $w=x^k \bar{x}^m$, $k\geqq 1$, $m\geqq 1$, x, $\bar{x}\in\{0, 1\}$, $\bar{x}\neq x$. Then $\gamma=\beta w$.

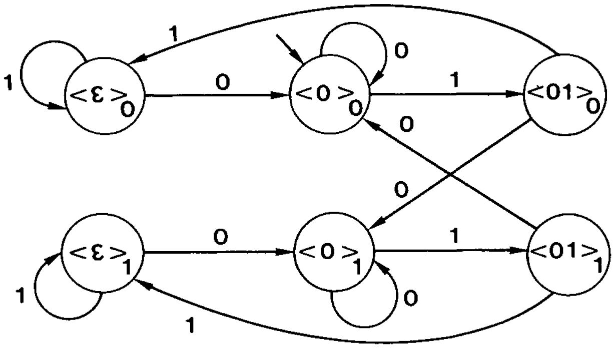
3. $w=x^k \bar{x}^m xs$, k, m, x, $\bar{x}$ like in case 2., $s\in\{0, 1\}^*$.

3.1. $|\alpha|\leqq k$. Then $\gamma=x^{k-|\alpha|}\bar{x}^{m+1}$.

3.2. $k+1\leqq|\alpha|\leqq k+m$. Then $\gamma=\bar{x}^{m+k-|\alpha|+1}$.

3.3. $k+m+1\leqq|\alpha|$, *then* $\gamma=y$ where y is the inverse of the first letter of β. ■

Example 1: Let us consider the word $w=010$. The transition diagram of the corresponding minimal sorting automaton A_{010} is depicted in the following figure.



In the following let $T=(S, s, \sigma, \{0, 1\}, \tau)$ be the 2-uniform tag system corresponding to the automaton A from lemma 1 according to proposition 1. Hence $\mathrm{seq}_T=\mathrm{sort}_A=\underline{a}_w$. Denote:

$$\underline{b}_w=b(0)\,b(1)\,b(2)\ldots=\mathrm{intseq}_T=\mathrm{state}_A.$$

LEMMA 2: *σ is an injective mapping.*

Proof: Let $s_1, s_2 \in S$, $s_1 \neq s_2$, $\sigma(s_1)=\sigma(s_2)$. From minimality of the automaton A we get $\tau(s_1)\neq\tau(s_2)$. Let now d be the rightmost symbol of w and $\bar{d}$ its inverse. Since no occurrence of w in the word scanned by A can be terminated by $\bar{d}$, we get:

$$\begin{aligned}\tau(s_1)&=\tau(\delta(s_1, \bar{d})),\\ &=\tau(\delta(s_2, \bar{d}))\ \text{since}\ \sigma(s_1)=\sigma(s_2),\\ &=\tau(s_2)\ -\ \text{a contradiction.}\ \blacksquare\end{aligned}$$

To obtain our main result we will first investigate the structure of the word $\underline{b}_w$; the results for $\underline{a}_w$ will follow directly as can be seen from the following lemma 4.

Let $x \in S$, $x=\langle\alpha\rangle_i$. Denote $\bar{x}=\langle\alpha\rangle_{1-i}$. Elements $x, y \in S$ will be called *associated* $(x \sim y)$ iff $x=y$ or $x=\bar{y}$.

REMARK 1: If for some $x, y \in S$ we have $x \sim y$ and $\tau(x)=\tau(y)$ then $x=y$.

LEMMA 3: *For each* $i \geq 0$ $b(i) \sim b(i+2^{|w|-1})$.

Proof: Let $s \in S$. The terminal factor of length $|w|-1$ of the words $i_{[2]}$ and $(i+2^{|w|-1})_{[2]}$ is the same. Hence $\delta(s, i_{[2]})=\langle\alpha\rangle_j$ for some $\langle\alpha\rangle_j \in S$ if and only if $\delta(s, (i+2^{|w|-1})_{[2]})=\langle\alpha\rangle_k$ for some $k \in \{0, 1\}$. $\blacksquare$

LEMMA 4: *If* $\underline{a}_w = \alpha u^{2^{|w|}} \ldots$ *for some* $\alpha, u \in \{0, 1\}^*$, *then* $\underline{b}_w = \alpha' (u')^2 \ldots$ *for some* $\alpha', u' \in S^*$ *such that* $|\alpha'| = |\alpha|$, $|u'| = |u^{2^{|w|-1}}|$.

Proof: Since $|u^{2^{|w|-1}}|$ is a multiple of $2^{|w|-1}$, the assertion follows from lemma 3 and remark 1. ■

LEMMA 5: (i) *Let* $i \geqq 0$. *Then:*

$$[\tau(b(2i)) + \tau(b(2i+1))] \equiv [\tau(b(2i+2^{|w|})) + \tau(b(2i+1+2^{|w|}))] \pmod 2.$$

(ii) *For each* $i \geqq 0$ *there is exactly one* $0 \leqq j < 2^{|w|-2}$ *such that for each* $0 \leqq k < 2^{|w|-2}$:

$$[\tau(b(2i+2k)) + \tau(b(2i+2k+1))] \not\equiv [\tau(b(2i+2k+2^{|w|-1})) + \tau(b(2i+2k+1+2^{|w|-1}))] \pmod 2$$

is valid if and only if $k = j$. *Moreover*, j *depends only on the value* $i \bmod 2^{|w|-1}$.

Proof: The assertions follow from proposition 2. ■

A word $x \in S^*$ will be called *m-block* ($m \geqq 0$) iff $x = \sigma^m(d)$ for some $d \in S$. A word $x \in S^*$ is *m-factorizable* iff it is a (possible empty) concatenation of m-blocks. The set of all m-blocks will be denoted $\mathscr{B}_m$, the set of all m-factorizable words will be denoted $\mathscr{F}_m$.

REMARK 2: Each m-block is of length 2^m.

Eeach initial factor of $\underline{b}_w$ is of length divisible by 2^m iff it is m-factorizable.

An m-block x will be called *even/odd/* iff for some $i \geqq 0$ $x = \sigma^m(b(2i))$ $/x = \sigma^m(b(2i+1))/$.

REMARK 3: For $m \geqq 1$ each m-block is a concatenation of some even $(m-1)$-block with some odd $(m$-$1)$-block.

LEMMA 6: *For* $m \geqq 0$ *no m-block can be both even and odd.*

Proof: Let for some m, i, $k \geqq 0$ $\sigma^m(b(2i)) = \sigma^m(b(2k+1))$.

1. Let $m = |w| - 1$. Let j be as in (ii) of lemma 5. Since $\tau(\sigma^m(b(2i))) = \tau(\sigma^m(b(2k+1)))$, using (i) of lemma 5 we get:

$$\begin{aligned}[\tau(b(2^{m+1}k+2j)) + \tau(b(2^{m+1}k+2j+1))] &\equiv [\tau(b(2^{m+1}i+2j)) + \tau(b(2^{m+1}i+2j+1))] \\ &\equiv [\tau(b(2^{m+1}k+2j+2^m)) + \tau(b(2^{m+1}k+2j+1+2^m))] \pmod 2\end{aligned}$$

— a contradiction to (ii) of lemma 5.

2. If $m \neq |w|-1$, then by several applications of σ or σ^{-1} (lemma 2) one obtains case 1 (σ is injective). ■

LEMMA 7: *If* $\underline{b}_w = x B \ldots$, *where* $B \in \mathscr{B}_m$ *then* $x \in \mathscr{F}_m$.

Proof: Induction on m. The case $m=0$ is evident.

Let $m>0$. Then $B=B_0 B_1$, where B_0, $B_1 \in \mathscr{B}_{m-1}$, B_0 is even. By induction hypothesis, $x \in \mathscr{F}_{m-1}$. If $x \notin \mathscr{F}_m$ then B_0 is odd — a contradiction to lemma 6. ■

LEMMA 8: *If* $\underline{b}_w = x_1 u \ldots = x_2 u \ldots$, *where* $x_1 \in \mathscr{F}_m - \mathscr{F}_{m+1}$ $x_2 \in \mathscr{F}_{m+1}$, *then* u *is a proper initial factor of both some even and some odd m-block.*

Proof: From the first factorization of $\underline{b}_w$ we obtain the fact that u is an initial factor of an infinite word beginning by an m-block (since $x_1 \in \mathscr{F}_m$) which is odd (since $x_1 \notin \mathscr{F}_{m+1}$). The second factorization implies that u is an initial factor of an infinite word beginning by an $(m+1)$-block, hence also by an even m-block. If $|u| \geqq 2^m$ then the first m-block of u is simultaneously even and odd — a contradiction to lemma 6. ■

LEMMA 9: *If* $\underline{b}_w = xuBu \ldots$, B *being a word of length divisible by* 2^m, *and* $x \in \mathscr{F}_m$, *then* $u \in \mathscr{F}_m$.

Proof: Let $u \in \mathscr{F}_{m'} - \mathscr{F}_{m'+1}$, $m' \geqq 0$. If $m'<m$ then $x \in \mathscr{F}_{m'+1}$, $xuB \in \mathscr{F}_{m'+1}$. From lemma 8 and the fact that $u \in \mathscr{F}_{m'}$ follows $u=\varepsilon$ hence $u \in \mathscr{F}_{m'+1}$ — a contradiction. ■

LEMMA 10: $\underline{b}_w$ *contains no factors of the form* $uBuBu$ *where* $u \in S^*$, $B \in \mathscr{B}_m$, $m \geqq 0$.

Proof: Let $\underline{b}_w = xuBuBu \ldots$ We use induction on $|u|$.

1. $|u|=0$. In this case $\underline{b}_w = xBB \ldots$, from lemma 7 we obtain $x \in \mathscr{F}_m$ thus B is both even and odd — a contradiction.

2. $|u|>0$. Lemma 7 implies $xu \in \mathscr{F}_m$, $xuBu \in \mathscr{F}_m$ hence $u \in \mathscr{F}_m$.

2.1. If B is an even m-block then u can be factorized as $u=Cv$, $c \in \mathscr{B}_m$, $|v|<|u|$ and $\underline{b}_w = xCvBCvBCv \ldots$, $BC \in \mathscr{B}_{m+1}$ — a contradiction to induction hypothesis.

2.2. If B is an odd m-block then a similar contradiction can be obtained using the factorization $u=vC$, $C \in \mathscr{B}_m$. ■

COROLLARY 1: $\underline{b}_w$ *contains no cubes.*

Proof: If $\underline{b}_w = xv^3 \ldots$ then (if $v \neq \varepsilon$) $v=uB$ for some $B \in \mathscr{B}_0 = S$ and $\underline{b}_w = xuBuBuB \ldots$: — a contradiction to lemma 10. ■

LEMMA 11: *$\underline{b}_w$ contains no factor of the form $xyBzxyBzx\ldots$ where $x \in S$, $B \in \mathscr{B}_m$, $zxy \in \mathscr{B}_m$, $m = |w| - 2$.*

Proof: If $\underline{b}_w$ contains such a factor then it is of length $2^{|w|}+1$. Let the first occurrence of x in this factor appears in $\underline{b}_w$ at the place with the index q. Then for $q \leqq p \leqq q + 2^{|w|-1}$ we have $b(p) = b(p + 2^{|w|-1})$, hence for no $0 \leqq k < 2^{|w|-2}$ the congruence from (ii) of lemma 5 is valid for $i = [q/2]$ — a contradiction. ■

LEMMA 12: *$\underline{b}_w$ contains no factor of the form $xyBzxyBzx$ where $x \in S$, $B \in \mathscr{B}_m$, $zxy \in \mathscr{B}_m$, $m \geqq 0$.*

Proof: If $B \in \mathscr{B}_m$ then $\sigma(B) \in \mathscr{B}_{m+1}$. Thus if $\underline{b}_w$ contains a factor $xyBzxyBzx$ for some m, then it contains a similar factor for $m+1$. Lemma 11 now directly implies that $\underline{b}_w$ does not contain a factor $xyBzxyBzx$ for $m \leqq |w| - 2$. For $m > |w| - 2$ we proceed by induction.

Let $\underline{b}_w = \alpha xyBzxyBzx\ldots$ for some $m > |w| - 2$.

1. Let $|\alpha|$ be even, i. e. x is an even 0-block. Then $y \neq \varepsilon$ otherwise the m-block zxy would be terminated by an even 0-block. Thus $y = dv$, $d \in S$, $v \in S^*$, and:

$$\underline{b}_w = \alpha xdvBzxdvBzx\ldots$$

From lemma 3 we obtain:

$$\underline{b}_w = \alpha xdvBzxdvBzxd'\ldots,$$

where $d \sim d'$. From (i) of lemma 5 and from remark 1 we get:

$$\tau(x) + \tau(d) = \tau(x) + \tau(d')$$
$$\tau(d) = \tau(d')$$
$$d = d'.$$

Since σ is injective, $\underline{b}_w$ can be factorized as follows:

$$\underline{b}_w = \alpha' x' y' B' z' x' y' B' z' x' \ldots,$$

where $\sigma(\alpha') = \alpha$, $\sigma(x') = xd$, $\sigma(y') = v$, $\sigma(B') = B$ $\sigma(z') = z$ — a contradiction to induction hypothesis.

2. If $|\alpha|$ is odd then by factorization $z = vd$ one can obtain a contradiction analogically to case 1. ■

LEMMA 13: $\underline{b}_w$ *contains no factor of the form* $xyBzxyBzx$, *where* $x \in S$, $B \in \mathscr{B}_m$, $z, y \in S^*$.

Proof: Induction on $|zxy|$.

1. $|zxy|=1$ i. e. $z=y=\varepsilon$. The assertion follows from lemma 10.

2. $|zxy|>1$. Let $\underline{b}_w=\alpha xyBzxyBzx\ldots$ From lemma 7 we obtain $zxy \in \mathscr{F}_m$. Moreover, $|zxy|$ is an odd multiple of 2^m, otherwise B would be simultaneously odd and even.

2.1. $y=vA$, A being an even m-block. Then:

$$\underline{b}_w=\alpha xvABzxvABzx\ldots, \qquad AB \in \mathscr{B}_{m+1}, \quad |zxv|<|zxy|,$$

— a contradiction to induction hypothesis.

2.2. $z=Av$, $A \in \mathscr{B}_{m+1}$. Then:

$$\underline{b}_w=\alpha xyBAvxyBAvx\ldots, \qquad |vxyB|<|Avxy|=|zxy|$$

— a contradiction to induction hypothesis.

2.3. $z=Av$, A being an odd m-block. Then:

$$\underline{b}_w=\alpha xyBAvxyBAvx\ldots, \qquad BA \in \mathscr{B}_{m+1}, \quad |vxy|<|zxy|$$

— a contradiction to induction hypothesis.

2.4. $y=vA$, $A \in \mathscr{B}_{m+1}$. Then:

$$\underline{b}_w=\alpha xvABzxvABzx\ldots, \qquad |Bzxv|<|zxvA|=|zxy|$$

— a contradiction to induction hypothesis.

2.5. $zxy \in \mathscr{B}_m$ — a contradiction to lemma 12. ■

We have now proved the following properties of $\underline{b}_w$ and $\underline{a}_w$:

THEOREM 2: $\underline{b}_w$ *does not contain a factor of the form* $xvxvx$, $x \in S$, $v \in S^*$.

Proof: The case $v=\varepsilon$ follows from corollary 1.

In the case $v \neq \varepsilon$ using the factorization $v=dz$, $d \in \mathscr{B}_0=S$, $z \in S^*$, one obtains $xvxvx=xdzxdzx$ and lemma 13 implies that such a factor cannot be contained in $\underline{b}_w$. ■

Now we are able to prove theorem 1. Let us suppose that $\underline{a}_w$ contains a factor of the form $(xv)^{2^{|w|}}x$, $x \in \{0, 1\}$, $v \in \{0, 1\}^*$. Applying lemma 4 twice (for $u=xv$ and $u=vx$) we obtain that the corresponding factor of $\underline{b}_w$ has the form $x'v'x'v'x'$ — a contradiction to theorem 2. ■

Theorem 1 does not exclude the possibility that $\underline{a}_w$ contains a factor of the form $u^{2|w|}$. Our next goal is to find some necessary conditions for appearing of such a factor in $\underline{a}_w$. First we shall describe how do the squares in $\underline{b}_w$ look like.

LEMMA 14: *Let* $\underline{b}_w = \alpha u B u \ldots$, $B \in \mathscr{B}_m$, $u \neq \varepsilon$.

Then $\alpha, u \in \mathscr{F}_m$.

Proof: Let $\alpha \in \mathscr{F}_{m'} - \mathscr{F}_{m'+1}$. Let $m' < m$. Lemma 7 implies $u \in \mathscr{F}_{m'}$. Since $\alpha \notin \mathscr{F}_{m'+1}$, the first m'-block of u is odd. Since $B \in \mathscr{F}_{m'+1}$, the same block is even — a contradiction.

Thus $m' \geqq m$, $\alpha \in \mathscr{F}_m$. Lemma 7 implies $u \in \mathscr{F}_m$. ■

LEMMA 15: *Let* $\underline{b}_w = \alpha u B u \ldots$, $B \in \mathscr{B}_m$. *Then:*

$$|u| = 2^q - 2^m \quad \textit{for some} \quad q \geqq m.$$

Proof: Induction on $|u|$.

1. If $|u| = 0$, then $|u| = 2^m - 2^m$.

2. Let $|u| > 0$. Lemma 14 implies that $|u|$ is divisible by 2^m.

2.1. If $|u| = 2^m$ then $|u| = 2^{m+1} - 2^m$.

2.2. Let $|u| > 2^m$. Then $u = A v C$, $A, C \in \mathscr{B}_m$, $v \in S^*$, and $\underline{b}_w = \alpha A v C B A v C \ldots$ Either CB or BA is an $(m+1)$-block. By induction hypothesis, for some $q \geqq m+1$ $|A v| = 2^q - 2^{m+1}$ or $|v C| = 2^q - 2^{m+1}$. In both cases $|u| = 2^q - 2^{m+1} + 2^m = 2^q - 2^m$. ■

LEMMA 16: *Let* $\underline{b}_w = \alpha u u \ldots$, $u \neq \varepsilon$, $\alpha \in \mathscr{F}_m$. *Then*

$$|u| = 2^q \quad \textit{for some} \quad q \geqq m + |w| - 1.$$

Proof: $u = v B$ for some $v \in S^*$, $B \in S = \mathscr{B}_0$. Lemma 15 implies $|v| = 2^q - 1$ for some $q \geqq 0$, hence $|u| = 2^q$. Let $q < |w| - 1 + m$. Since σ is injective (lemma 2) and $\underline{b}_w = \sigma(\underline{b}_w)$, for $k = |w| - 1 - q$ (satisfying $-m < k \leqq |w| - 1$) one obtains $|\sigma^k(u)| = 2^{|w|-1}$, $\sigma^k(\alpha) \in \mathscr{F}_1$, and $\underline{b}_w = \sigma^k(\alpha)\, \sigma^k(u)\, \sigma^k(u) \ldots$ — a contradiction to (ii) of lemma 5. ■

LEMMA 17: *Let* $\underline{b}_w = \alpha u u \ldots$, $u \neq \varepsilon$, $\alpha \in \mathscr{F}_m - \mathscr{F}_{m+1}$. *Then:*

$$|u| = 2^{m+|w|-1} \quad \textit{and} \quad u \in \mathscr{F}_m - \mathscr{F}_{m+1}.$$

Proof: According to lemma 16 $|u|$ is divisible by 2^m. $\alpha \in \mathscr{F}_m$ implies $u \in \mathscr{F}_m$. Since $\alpha \notin \mathscr{F}_{m+1}$ the initial m-block of u is odd, thus $u \notin \mathscr{F}_{m+1}$.

Considering lemma 16 it is enough to prove $|u| \leqq 2^{m+|w|-1}$. Let $u=2^{r+m+|w|-1}$, $r \geqq 0$. Then $\underline{b}_w=\sigma^{-m}(\underline{b}_w)=\alpha' u' u' \ldots$, where $|u'|=2^{r+|w|-1}$, $\alpha' \in \mathscr{F}_0-\mathscr{F}_1$. Let $u'=xv$, $x \in S$, $v \in S^*$. If $r \geqq 1$, then from lemma 3 and (i) of lemma 5 follows (since the rightmost letter of $u' u'$ has in $\underline{b}_w$ an even index):

$\underline{b}_w=\alpha' xvxv \ldots = \alpha' xvxvx \ldots$ —a contradiction to theorem 1. ■

Corollary 2: *If $\underline{b}_w=\alpha uu \ldots$, $u \neq \varepsilon$, $\alpha \in \mathscr{F}_m-\mathscr{F}_{m+1}$ for some $m \geqq 0$, then the same is true for $m=0$.*

Lemma 18: *Let $\underline{b}_w=\alpha uu \ldots$, $u \neq \varepsilon$, $\alpha \in \mathscr{F}_0-\mathscr{F}_1$. Then either for $y=\alpha$ or for $y=\alpha u$:*

$$(*) \qquad |y|=v(w)+\bar{d} \pmod{2^{|w|}},$$

where $\bar{d}$ is the inverse of the rightmost digit of w and $v(w)$ is the integer whose binary notation is w.

Proof: Lemma 17 implies $|u|=2^{|w|-1}$. It is easy to see that (ii) of lemma 5 is satisfied only if $(*)$ is valid. ■

Our knowledge of the powers in $\underline{b}_w$ and $\underline{a}_w$ can now be summarized in the following theorems:

Theorem 3: *If $\underline{b}_w=\alpha uu \ldots$, $u \neq \varepsilon$, $\alpha \in \mathscr{F}_m-\mathscr{F}_{m+1}$, then:*

(i) $|u|=2^{m+|w|+1}$ *and* $u \in \mathscr{F}_m-\mathscr{F}_{m+1}$.

(ii) $\underline{b}_w \doteq \alpha' u' u' \ldots$ *for some* $u' \neq \varepsilon$, $\alpha' \in \mathscr{F}_0-\mathscr{F}_1$ *and either for* $y=\alpha'$ *or for* $y=\alpha' u'$:

$$|y| \equiv v(w)+\bar{d} \pmod{2^{|w|}}.$$

Theorem 4: *If $\underline{a}_w=\alpha u^{2^{|w|}} \ldots$, $u \notin \varepsilon$, $|\alpha|$ divisible by 2^m and not divisible by 2^{m+1}, then:*

(i) $|u|=2^m$.

(ii) *either for* $z=|\alpha|/2^m$ *or for* $z=|\alpha|/2^m+2^{|w|-1}$:

$$z \equiv v(w)+\bar{d} \pmod{2^{|w|}}.$$

Using corollary 2 one can show that $\underline{b}_{1101}$ does not contain squares, and consequently that $\underline{a}_{1101}$ does not contain a factor of the form u^{16}. On the other hand for each w of the form 1^k, $k>1$, $\underline{a}_w$ contains the subword 0^{2^k} beginning in $\underline{a}_w$ at the place with the index (in binary notation) $1\,w=1^{k+1}$.

Each $\underline{a}_w$ contains the factor $0^{2^{|w|}-1}$, as shown in the following table where n is a binary notation of such an index in $\underline{a}_w$ that:

$$a(n+1)\,a(n+2)\ldots a(n+2^{|w|}-1)=0^{2^{|w|}-1}$$

and $x \in \{0, 1\}^*$, $k \geqq 1$.

w	n	Remark
$00x$.	$1^{\|x\|}01w$	$w \notin 0^*$
$10x$.	$1^{\|w\|}w$	$x \notin 0^*$
$11x$.	$1w$	
$01x$.	$01^{\|w\|}w$	$x \notin 0^* \cup 1^*$
10^k.	$ww\,1^{k-1}w$	
010^k.	$011011\,w$	
011^k.	$w\,01\,ww$	
01.	010000	

ACKNOWLEDGMENT

The author is deeply indebted to prof. Jean Berstel for introduction to the problem and for many stimulating discussions.

REFERENCES

[CKM-FR 80] G. Christol, T. Kamae, M. Mendes-France et G. Rauzy, *Suites algébriques, automates et substitutions*, Bull. Soc. math. Fr., Vol. 108, 1980, pp. 401-419.

[Co 72] A. Cobham, *Uniform Tag Sequences*, Math. Syst. Theory, Vol. 6, 1972, pp. 164-192.

[Mo 21] H. M. Morse, *Recurrent Geodesics on a Surface of Negative Curvature*, Trans. Amer. Math. Soc., Vol. 22, 1921, pp. 84-100.

[Pa 81] J. J. Pansiot, *The Morse Sequence and Iterated Morphisms*, Inf. Proc. Letters, Vol. 12, No. 2, 1981, pp. 68-70.

[Th 06] A. Thue, *Über unendliche Zeichenreihen*, *Videnskabs-Selskabets Skifter*, Math. Naturv. Klasse, Kristiania, No. 7, 1906, pp. 1-22.

[Th 12] A. Thue, *Über di gegenseitige Lage gleichen Teile gewisser Zeichenreihen*, *Videnskapsselskapets Skifter*, *I.* Mat. - naturv. Klasse, Kristiania, No. 1, 1912, pp. 1-67.