Constructing Premaximal Binary Cube-free Words of Any Level

Elena A. Petrova
Ural Federal University
Ekaterinburg, Russia
captain@akado-ural.ru

Arseny M. Shur
Ural Federal University
Ekaterinburg, Russia
arseny.shur@usu.ru

We study the structure of the language of binary cube-free words. Namely, we are interested in the cube-free words that cannot be infinitely extended preserving cube-freeness. We show the existence of such words with arbitrarily long finite extensions, both to one side and to both sides.

1 Introduction

The study of repetition-free words and languages remains quite popular in combinatorics of words: lots of interesting and challenging problems are still open. The most popular repetition-free binary languages are the *cube-free* language CF and the *overlap-free* language OF. The language CF is much bigger and has much more complicated structure. For example, the number of overlap-free binary words grows only polynomially with the length [8], while the language of cube-free words has exponential growth [3]. The most accurate bounds for the growth of OF is given in [6] and for the growth of CF in [13]. Further, there is essentially unique nontrivial morphism preserving OF [10], while there are uniform morphisms of any length preserving CF [5]. The sets of two-sided infinite overlap-free and cube-free binary words also have quite different structure, see [12].

Any repetition-free language can be viewed as a poset with respect to prefix, suffix, or factor order. In case of prefix [suffix] order, the diagram of such a poset is a tree; each node generates a subtree and is a common prefix [respectively, suffix] of its descendants. The following questions arise naturally. *Does a given word generate finite or infinite subtree? Are the subtrees generated by two given words isomorphic? Can words generate arbitrarily large finite subtrees?* For some power-free languages, the decidability of the first question was proved in [4] as a corollary of interesting structural properties. The third question for ternary square-free words constitutes Problem 1.10.9 of [1]. For all kth power-free languages, it was shown in [2] that the subtree generated by any word has at least one leaf. Note that considering the factor order instead of the prefix or the suffix one, we get a more general acyclic graph instead of a tree, but still can ask the same questions about the structure of this graph. For the language OF, all these questions were answered in [11, 14], but almost nothing is known about the same questions for CF.

In this paper, we answer the third question for the language CF in the affirmative. Namely, we construct cube-free words that generate subtrees of any prescribed depth and then extend this result for the subgraphs of the diagram of factor order.

2 Preliminaries

Let us recall necessary notation and definitions. We consider finite and infinite words over the binary alphabet $\Sigma = \{a, b\}$. If x is a letter, then $\bar{x}$ denotes the other letter. By default, “word” means a finite word.

P. Ambrož, Š. Holub and Z. Masáková (Eds.):
8th International Conference WORDS 2011
EPTCS 63, 2011, pp. 168–178, doi:10.4204/EPTCS.63.23

Words are denoted by uppercase characters (to denote one-sided infinite words, we add the subcsript $_\infty$ at the corresponding side). We write λ for the *empty word*, and $|W|$ for the length of the word W. The letters of nonempty finite and right-infinite words are numbered from 1; thus, $W = W(1)W(2)\cdots W(|W|)$. The letters of left-infinite words are numbered by *all nonnegative integers*, starting from the right.

We use standard definitions of factors, prefixes, and suffixes of a word. The factor $W(i)\cdots W(j)$ is written as $W(i\dots j)$. A positive integer $p \le |W|$ is a *period* of a word W if $W(i) = W(i+p)$ for all $i \in \{1,\dots,|W|-p\}$. The minimal period of W is denoted by $\mathrm{per}(W)$. The *exponent* of a word is the ratio between its length and its minimal period: $\exp(W) = |W|/\mathrm{per}(W)$. Words of exponent 2 and 3 are called squares and cubes, respectively. The *local exponent* of a word is the number $\mathrm{lexp}(W) = \sup\{\exp(V) | V \text{ is a factor of } W\}$. Periodic words possess the *interaction property* expressed by the textbook Fine and Wilf theorem: if a word U has periods p and q, and $|U| \ge p+q-\gcd(p,q)$, then U has the period $\gcd(p,q)$.

A word W is β-free [β^+-free] if $\mathrm{lexp}(W) < \beta$ [respectively, $\mathrm{lexp}(W) \le \beta$]. The 3-free words are called *cube-free*, and the 2^+-free words are *overlap-free*. The language of all cube-free [overlap-free] words over Σ is denoted by CF [respectively, OF]. A morphism $f : \Sigma^+ \to \Sigma^+$ *avoids an exponent* β if the condition $\mathrm{lexp}(U) < \beta$ implies $\mathrm{lexp}(f(U)) < \beta$ for any word U. The following theorem allowes one to check cube-freeness of a morphism over the binary alphabet.

Theorem 1 ([9]). *A morphism $f : \Sigma^+ \to \Sigma^+$ is cube-free if and only if the word $f(aabbababbabbaabaababaabb)$ is cube-free.*

The *Thue–Morse morphism* θ is defined over Σ^+ by the rules $\theta(a) = ab$, $\theta(b) = ba$. The words

$$T_n^a = \theta^n(a),\ T_n^b = \theta^n(b)\ (n \ge 0)$$

are called *Thue–Morse blocks* or simply *n*-blocks. From the definition it follows that $T_{n+1}^x = T_n^x T_n^{\bar{x}}$. Hence, the sequences $\{T_n^a\}$ and $\{T_n^b\}$ have "limits", which are right-infinite *Thue-Morse words* T_∞^a and T_∞^b, respectively. We also consider the reversal ${}_\infty^a T$ of T_∞^a. The factors of Thue-Morse words are *Thue-Morse factors*; the set of all these factors is denoted by TM. Note that any word in TM can be written as $W = xQ_1\cdots Q_n y$, where $x,y \in \Sigma \cup \{\lambda\}$, $Q_1,\dots,Q_n \in \{ba,ab\}$. It is known since Thue [15] that $\mathsf{TM} \subset \mathsf{OF}$.

Let $L \subset \Sigma^*$ and $W \in L$. Any word $U \in \Sigma^*$ such that $UW \in L$ is called a *left context* of W in L. The word W is *left maximal* [*left premaximal*] if it has no nonempty left contexts [respectively, finitely many left contexts]. The *level* of the left premaximal word W is the length of its longest left context; thus, left maximal words are of level 0. The right counterparts of the above notions are defined in a symmetric way. We say that a word is *maximal [premaximal]* if it is both left and right maximal [respectively, premaximal]. The *level* of a premaximal word W is the pair $(n,k) \in \mathbb{N}$ such that n and k are the length of the longest left context of W and the length of its longest right context, respectively.

In particular, a word $W \in \mathsf{CF}$ is maximal if by adding any of the two letters on the left or on the right we obtain a cube. The word *aabaabaa* is an example of such a word.

The aim of this paper is to prove the following theorems:

Theorem 2. *In CF, there exist left premaximal words of any level $n \in \mathbb{N}_0$.*

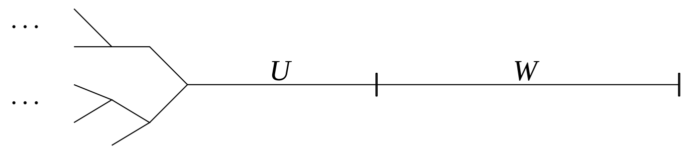
Theorem 3. *In CF, there exist premaximal words of any level $(n,k) \in \mathbb{N}_0^2$.*

3 Construction of premaximal words

Theorem 2 is proved by exhibiting a series of left premaximal words, containing words of any level. The series is constructed in two steps:

1. building an auxiliary series $\{W_n\}_0^\infty$ such that each word W_n has, up to one easily handled exception, a unique left context of any length $\leq n$;
2. completing the word W_n to a left premaximal word $\overline{W}_n$.

If a word $W \in \mathsf{CF}$ has a unique left context of length n, say U, and two left contexts of length $n{+}1$, then we say that U is the *fixed* left context of W (see the picture below).



Example 1. *Let* $W = aabaaba$*. Since* $aW = aaa\cdots$, $abW = (aba)^3$*, but* $aabbW, babbW \in \mathsf{CF}$*, we see that the fixed left context of the word* W *equals* abb*.*

Now let us explain step 1. We build the series $\{W_n\}_0^\infty$ inductively, one word per iteration, in a way that the fixed left context X_n of the word W_n is of length $\geq n$ (we will discuss the mentioned exception at the moment of its appearance). We put $W_0 = aabaaba$ and note that the left-infinite word

$$ {}^{a}_{\infty}T\, abaaba = \cdots abba\, baab\, baab\, abbW_0 $$

is cube-free. So, we require that each word W_n satisfies the following properties:

(W1) W_n starts with W_0;

(W2) any word ${}^{a}_{\infty}T(k\ldots1)$ is a left context of W_n;

(W3) some word ${}^{a}_{\infty}T(k\ldots1)$ with $k \geq n$ is the fixed left context of W_n, denoted by X_n;

(W4) if $|X_n| > n$, then $W_{n+1} = W_n$ (*trivial* iterations).

The basic idea for obtaining W_{n+1} from W_n at nontrivial iterations is to let

$$ W_{n+1} = \underbrace{W_n}\underbrace{xX_nW_n}\underbrace{xX_nW_n}, \tag{1} $$

where x is the letter "prohibited" at the $(n{+}1)$th iteration, i.e. xX_n certainly is not a left context of W_{n+1}. Thus, the fixed left context of W_{n+1} is longer than the one of W_n by definition.

Remark 1. *An attempt to build the series* $\{W_n\}_0^\infty$ *directly by (1) fails because cubes will occur at the border of some words* W_n *and* xX_n*. For instance, let us construct the word* W_4*. We have* $W_3 = W_0$ *in view of (W4) and Example 1,* $X_3 = abb$*, and the context* $aabb$ *should be forbidden in view of (W2), because* ${}^{a}_{\infty}T(4\ldots1) = babb$*. So,* $x = a$ *and the word* W_3xX_3 *has the factor* aaa*.*

A way out from this situation is the following idea: we insert a special "buffer" word after each of three occurrences of W_n in (1). This insertion allows us to avoid local cubes at the border. Below we use the following notation:

- $P'_n = xX_n$, $P_n = \bar{x}X_n$, where x is the letter, prohibited at the $(n{+}1)$th iteration; thus, $P_n \in \mathsf{TM}$;
- S_n is the word inserted after W_n at the $(n{+}1)$th iteration;
- $S'_n = S_0S_1\cdots S_n$ is the factor of W_{n+1} between W_0 and the nearest occurrence of P'_n;
- $W'_n = P'_nW_nS_n$.

In these terms, we have the following expressions for W_{n+1} for any nontrivial iteration:

$$W_{n+1} = \underbrace{W_nS_n}\underbrace{xX_nW_nS_n}\underbrace{xX_nW_nS_n} \tag{2a}$$

$$W_{n+1} = \underbrace{W_nS_n}\underbrace{P'_nW_nS_n}\underbrace{P'_nW_nS_n} \tag{2b}$$

The structure of the word W_{n+1} imposes the following restrictions on the words S_n and S_{n+1}:

(S1) Since the word $X_{n+1}W_{n+1}S_{n+1}$ is a factor of W_{n+2}, X_{n+1} ends with X_n, and $X_nW_{n+1}x = (X_nW_nS_nx)^3$ by (2a), the word S_{n+1} must start with $\bar{x}$, which is the first letter of P_n;

(S2) Since the word S_nxX_n is a factor of W_{n+1}, if X_n starts with x [$\bar{x}x\bar{x}x$], then S_n ends with $\bar{x}$ [respectively, x]. (Recall that $X_n \in \mathsf{TM}$ is an overlap-free word, whence any other prefix of X_n does not restrict the last letter of S_n.)

Thus, our first goal is to find the words S_n satisfying (S1) and (S2) such that all words S'_n are cube-free. In other words, we have to construct a cube-free right-infinite word $S'_\infty = S_0S_1 \cdots S_n \cdots$. The following lemma is easy.

Lemma 1. *The letters ${}^{a}_{\infty}T(n)$ and ${}^{a}_{\infty}T(n-1)$ coincide if and only if $n = m \cdot 2^k$ for some odd integers m and k.*

Remark 2. *If the only left context of length n of the word W_n begins with xx, then $|X_n| > n$, because the letter before xx is also fixed. Thus, by* (W4) *we have $W_{n+1} = W_n$ (and then $S_n = \lambda$) for all values of n mentioned in Lemma 1. For all other values of n ($n > 3$), the iterations will be nontrivial.*

While constructing the word S'_∞ we follow the next four rules:

1. For all nontrivial iterations, $S_n \in \{T_2^x, T_2^xT_2^x, T_4^x, T_2^xT_2^{\bar{x}}T_1^x, T_1^x, T_1^xT_2^{\bar{x}} \,|\, x \in \Sigma\}$; hence, $S_n \in \mathsf{TM}$.
2. Whenever possible, we choose S_n to be a 2-block or a product of 2-blocks.
3. Otherwise, if S_n ends with the block T_1^x, we put $S_{n+1} = T_1^{\bar{x}}$ or $S_{n+1} = T_1^{\bar{x}}T_2^x$ (or the same possibilities for S_{n+2} if $S_{n+1} = \lambda$).
4. If $S_n \neq \lambda$ and there is no restriction (S2) on the last letter of S_n, we add this restriction artificially. Namely, we fix the last letter of S_n to be $\bar{x}$ if S_{n-1} ends with x (or if S_{n-2} ends with x while $S_{n-1} = \lambda$).

Taking rules 1–4 into account, we can prove, by case examination, the following lemma about the first and the last letters of the words S_n.

Lemma 2. (1) *If S_n ends with x, then either S_{n+1} ends with $\bar{x}$, or $S_{n+1} = \lambda$ and S_{n+2} ends with $\bar{x}$.*
(2) *The first letter of a nonempty word S_n coincides with the last one for all n, except for the cases when $P_n = x\bar{x}x\bar{x}\cdots$ or $P_n = xx\bar{x}\cdots$.*

The construction of the word S'_∞, the correctness of which we will prove, is given by Table 1. According to this table, rule 3 applies to S_n if and only if P_n starts with $x\bar{x}x\bar{x}$. Hence if the word P_n has such a prefix, then P_{n-1} (or P_{n-2} if the $(n-1)$th iteration is trivial) has no such prefix; as a result, the word S_{n-1} (respectively, S_{n-2}) ends with a 2-block.

Now consider the case $P_n = x\bar{x}x\bar{x}\cdots$ in more details. Without loss of generality, let P_n start with b. Then $P_n = babaab\cdots$. Since $P'_n = aabaab\ldots$, the word S_n cannot end with a or with $baab$; thus, it cannot end with a 2-block and we should use rule 3.

Table 1: the suffixes S_n for 32 successive iterations starting from some number k divisible by 32. The righthand [lefthand] part of the table applies if the current letter of T^b_∞ is equal [resp., not equal] to the previous one. Trivial iterations are omitted.

Iteration no.	Prohibitions		
(n)	Start	End	S_{n-1}
k	$\overline{x}$	$\overline{x}$	T_2^x
$k+1$			
$k+2$	x	x	$T_2^{\overline{x}}T_2^{\overline{x}}$
$k+4$	$\overline{x}$	$\overline{x}$	T_2^x
$k+5$	$\overline{x}$	$x, T_2^{\overline{x}}$	$T_2^xT_2^{\overline{x}}T_1^x$
$k+6$	x	$\overline{x}$	$T_1^{\overline{x}}$
$k+8$	x	x	$T_2^{\overline{x}}$
$k+10$	$\overline{x}$	$\overline{x}$	$T_2^xT_2^x$
$k+12$	x	x	$T_2^{\overline{x}}$
$k+13$	x	$\overline{x}, T_2^x$	$T_2^{\overline{x}}T_2^xT_1^{\overline{x}}$
$k+14$	$\overline{x}$	x	T_1^x
$k+16$	$\overline{x}$	$\overline{x}$	T_2^x
$k+17$	$\overline{x}$	x	T_1^x
$k+18$	$xx\overline{x}$	$\overline{x}$	$T_1^{\overline{x}}$
$k+20$	$\overline{xx}x$	x	$T_2^{\overline{x}}$
$k+21$	x	$\overline{x}$	$T_1^{\overline{x}}$
$k+22$	$\overline{xx}x$	x	T_1^x
$k+24$	$xx\overline{x}$	$\overline{x}$	T_2^x
$k+26$	x	x	$T_2^{\overline{x}}$
$k+28$	$\overline{x}$	$\overline{x}$	T_4^x
$k+29$	$\overline{x}$	$x, T_2^{\overline{x}}$	$T_2^xT_2^{\overline{x}}T_1^x$
$k+30$	x	$\overline{x}$	$T_1^{\overline{x}}(T_1^{\overline{x}}T_2^x)$

Iteration no.	Prohibitions		
(n)	Start	End	S_{n-1}
k	x	x	$T_2^{\overline{x}}$
$k+1$	x	$\overline{x}$	$T_1^{\overline{x}}$
$k+2$	$\overline{xx}x$	x	T_1^x
$k+4$	$xx\overline{x}$	$\overline{x}$	T_2^x
$k+5$	$\overline{x}$	x	T_1^x
$k+6$	$xx\overline{x}$	$\overline{x}$	$T_1^{\overline{x}}$
$k+8$	$\overline{xx}x$	x	$T_2^{\overline{x}}$
$k+10$	$\overline{x}$	$\overline{x}$	T_2^x
$k+12$	x	x	$T_4^{\overline{x}}$
$k+13$	x	$\overline{x}, T_2^x$	$T_2^{\overline{x}}T_2^xT_1^{\overline{x}}$
$k+14$	$\overline{x}$	x	T_1^x
$k+16$	$\overline{x}$	$\overline{x}$	T_2^x
$k+17$	$\overline{x}$	x	T_1^x
$k+18$	$xx\overline{x}$	$\overline{x}$	$T_1^{\overline{x}}$
$k+20$	$\overline{xx}x$	x	$T_2^{\overline{x}}$
$k+21$	x	$\overline{x}$	$T_1^{\overline{x}}$
$k+22$	$\overline{xx}x$	x	T_1^x
$k+24$	$xx\overline{x}$	$\overline{x}$	T_2^x
$k+26$	x	x	$T_2^{\overline{x}}$
$k+28$	$\overline{x}$	$\overline{x}$	T_4^x
$k+29$	$\overline{x}$	$x, T_2^{\overline{x}}$	$T_2^xT_2^{\overline{x}}T_1^x$
$k+30$	x	$\overline{x}$	$T_1^{\overline{x}}$

Since P_n is a factor of ${}^{a}_{\infty}T$ while ${}^{a}_{\infty}T$ is an infinite product of the blocks $T_2^a = abba$ and $T_2^b = baab$, one of the blocks T_2^a ends in the second position of P_n. First consider the following occurrence of P_n in ${}^{a}_{\infty}T$:

$$ {}^{a}_{\infty}T = \cdots \overbrace{abba}^{T_2^a}\overbrace{ab}^{T_2^a}\underbrace{ba\,\overbrace{baab}^{T_2^b}\,\overbrace{baab}^{T_2^b}\cdots}_{P_n} \qquad (3) $$

Since $P'_{n-1} = bbaab\cdots$, the word S_{n-1} ends with $abba$. Therefore, we cannot put $S_n = ab$ (otherwise S_n will have the suffix $baab$). Further, P_{n-1} starts with $abaab$, whence the first letter of S_n is a by (S1). Hence, according to rule 1, the only possibility for S_n is $T_2^aT_2^bT_1^a = abbabaabab$. It is easy to see that $S_{n+1} = ba$ satisfies both (S1) and (S2).

If the last embraced 2-block of (3) is T_2^a, not T_2^b, then we have, up to renaming the letters, the same case as below:

$$ {}_{\infty}^{\;a}T = \cdots \overbrace{baab}^{T_2^b}\, \overbrace{ab\,ba}^{T_2^a}\, \overbrace{baab}^{T_2^b}\cdots \quad \text{where } P_n = \underbrace{ba\,baab\cdots} $$

We assign, as above, $S_n = T_2^a T_2^b T_1^a$ and $S_{n+1} = T_1^b$. The problem appears on the $(n{+}5)$th iteration, because

$$P'_{n+4} = \underbrace{b}\,\underbrace{bab}\,\underbrace{bab}\,aab\cdots ,$$

i.e., S_{n+4} cannot end with *ba* or *ab*. Here we have an exclusion from the general method. We use the following trick. At the next three iterations ($(n{+}5)$th to $(n{+}7)$th, the last of them being trivial) we have to add the prefix *baa* to the fixed context. We will do this prohibiting 3-letter contexts instead of single letters. The word $P_{n+3} = babbaba\cdots$ has three left contexts of length 3: *aab*, *baa*, and *bba*. We will prohibit *bba* on the $(n{+}5)$th iteration and *aab* on the $(n{+}6)$th one. To do this, we deliberately put $P'_{n+4} = bba\,babbabaab\cdots$, $P'_{n+5} = aab\,babbabaab\cdots$. This allows us to choose $S_{n+4} = ba, S_{n+5} = ab$.

Remark 3. *The above trick leads to one local violation of the general rule on X_n. Namely, $|X_{n+5}| = n{+}4$ (this word coincides with X_{n+4}). The situation is corrected on the next iteration, when we get $|X_{n+6}| = n{+}7$ (and the $(n{+}7)$th iteration is trivial).*

Remark 4. *The word $T_2^a T_2^a T_2^b T_2^a T_2^a = \theta^2(aabaa)$ is not a factor of ${}_{\infty}^{\;a}T$. Hence, the factor $T_2^a T_2^b T_2^a$ occurs in ${}_{\infty}^{\;a}T$ inside the factor $T_2^b T_2^a T_2^b T_2^a$ or $T_2^a T_2^b T_2^a T_2^b$. Each such factor requires two uses of the above trick with 3-letter contexts.*

Let us consider the 108-uniform morphism $\psi : \Sigma^* \to \Sigma^*$, defined by the rules

$$\psi(a) = T_4^a T_2^a T_2^b T_2^a T_4^b T_2^b T_2^a T_4^b T_2^b T_2^a T_2^b T_4^a T_2^a T_2^b T_2^a, \tag{4a}$$

$$\psi(b) = T_4^b T_2^b T_2^a T_2^b T_4^a T_2^a T_2^b T_4^a T_2^a T_2^b T_2^a T_4^b T_2^b T_2^a T_2^b. \tag{4b}$$

Note that the words $\psi(b)$ and $\psi(a)$ coincide up to renaming the letters. A computer check shows that the word $\psi(aabbababbabbaabaababaabb)$ is cube-free. Hence by Theorem 1, ψ is a cube-free morphism and the word $\psi(T_\infty^b)$ is cube-free. So we put $S'_\infty = \psi(T_\infty^b)$. The ψ-image of one letter equals the product $S_{n-1}S_n \cdots S_{n+30}$ for some number n divisible by 32, see Table 1. The only exception is described below. Thus, such a ψ-image corresponds to 32 successive iterations, during which a 5-block is added to the fixed left context X_{n-1} to get X_{n+31}.

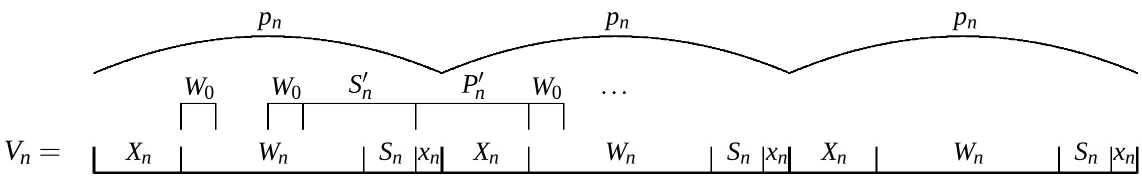
There are two different factorizations of the ψ-image of a letter, depending on the positions of the factors $T_2^b T_2^a T_2^b T_2^a$ and $T_2^a T_2^b T_2^a T_2^b$ inside and on the borders of the current 5-block of ${}_{\infty}^{\;a}T$. These factorizations are presented in the two parts of Table 1. The mentioned factors occur in the middle of $(2k{+}1)$-blocks for each $k \geq 2$. Thus, these factors occur in the middle of each 5-block, and also at the border of two equal 5-blocks. For the latter case, the factorization of the ψ-image of the second of two equal letters is given in the righthand part of Table 1. In the lefthand part of Table 1, there are two possibilities for S_{n+29}: the longer [shorter] one should be used if the next 5-block is equal [respectively, not equal] to the current one. In the first case, S_{n+29} consists of the last two letters of the ψ-image of the current letter and first four letters of the ψ-image of the next letter. In the second case, S_{n+29} consists exactly of the two last letters of the ψ-image.

The first several iterations are special. Namely, for the regularity of general scheme, we artificially put $W_3 = W_0 S_{-1} S_1$ (the 1st and the 3rd iterations are trivial by the general condition).

Thus, we defined the words S_n and then the words W_n for all positive integers n. The correctness of the construction is based on the following lemma.

Lemma 3. *The word X_nW_n is cube-free for all $n \in \mathbb{N}_0$.*

Proof. We prove by induction that all the words $V_n = (X_nW_nS_nx_n)^3$, where x_n is the letter forbidden on $(n{+}1)$th iteration, have no proper factors that are cubes. This fact immediately implies the statement of the lemma. The inductive base $n \le 4$ can be easily checked by hand or by computer. Let us prove the inductive step.The structure of the word V_n is illustrated by the following picture.



Assume to the contrary that the word V_n, $n \ge 5$, contains some cube U^3. Of course, it is enough to consider the case when the $(n{+}1)$th iteration is nontrivial. The factor U^3 of V_n has periods $q = |U|$ and $p_n = |V_n|/3$, but obviously does not satisfy the interaction property. Hence, $|U^3| = 3q \le q + p_n - 2$ by the Fine and Wilf theorem, yielding $q \le p_n/2 - 1$. On the other hand, by definition of W_n, the longest proper suffix of the word X_nW_n coincides with the longest proper prefix of V_{n-1}. If U^3 contains this prefix, then the latter has periods q and $p_{n-1} = |V_{n-1}|/3$. Applying the Fine and Wilf theorem again, we get $p_{n-1} \le q/2 - 1$. Excluding q from the two obtained inequalities, we get $p_n \ge 4p_{n-1} + 3$. But $p_n = |V_{n-1}| + |S_n| + 1 \le 3p_{n-1} + 17$. Thus, $p_{n-1} \le 14$. For $n \ge 5$, this is not the case. So, we conclude that U^3 does not contain the word X_nW_n.

Claim 1. The word S'_n occurs in V_n only three times.

Proof. Recall that S'_n is a product of 2-blocks (possibly except the last "odd" 1-block), and if $n \ge 5$, then S'_n begins with a 4-block. Hence, S'_n has no factor W_0 and, moreover, cannot begin inside W_0. Furthermore, it can be checked by hand or by computer that S'_∞ has no Thue-Morse factors of length >48. Now looking at the structure of S'_n and of V_n one can conclude that any "irregular" occurrence of S'_n in V_n should be a prefix of some word $S'_kP'_kW_0$, where $k < n$. The word S'_k is a proper prefix of S'_n. The word P'_k is obtained from a Thue-Morse factor by changing the first letter, and hence never begins with a 2-block. Hence, the only possibility is $k = n-1$, and S_n should be the 1-block coinciding with the prefix of P'_k. By Table 1, in all cases when S_n is a 1-block, P'_{n-1} begins with the square of letter, so this possibility cannot take place. ☐

Claim 2. The word $X_nW_nS_nx_n$ is cube-free.

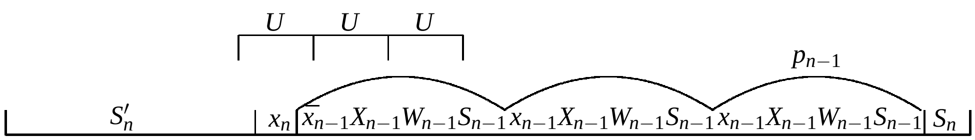
Proof. The word X_nW_n is a factor of V_{n-1} and hence is cube-free by the inductive assumption. Using again the fact that S'_n is "almost" a product of 2-blocks, we conclude that S'_nx_n is also cube-free. So, a cube in $X_nW_nS_nx_n$, if any, contains inside the suffix S'_{n-1} of the word W_n. This suffix is preceded by $W_0 = aabaaba$; the latter word breaks all periods of S'_{n-1} and does not produce a cube. Hence, the cube should contain more than one occurrence of the factor S'_{n-1}. Applying Claim 1 to the words S'_{n-1} and V_{n-1}, we see that the cube has the period $p_{n-1} = (|X_nW_n|{+}1)/3$. But this is impossible by condition (S1). The claim is proved. ☐

Combining Claim 2 with the fact that U^3 has no factor X_nW_n, we get that U^3 is contained inside the word $X_nW_nS_nx_nX_nW_n$. Furthermore, if S'_n is a factor of U^3, then the middle occurrence of U is inside S'_n (otherwise, U^3 contains one more occurrence of S'_n, contradicting Claim 1). In this case, the positions of all factors aa and bb in U have the same parity. But the rightmost occurrence of U in U^3 contains a suffix of S'_n followed by a prefix of the word $x_nX_n = P'_n$. The letter x_n breaks this parity of positions, which is impossible. The cases in which all the positions of *aa* and *bb* in the rightmost occurrence of U are on the same side of the letter x_n, can be easily checked by hand. Thus, we obtain that S'_n is not a factor of U^3. Thus, U^3 begins inside the factor S'_nx_n.

Where the word U^3 ends? It is easy to see that the word

$$X_nW_n = \bar{x}_{n-1}X_{n-1}W_{n-1}S_{n-1}x_{n-1}X_{n-1}W_{n-1}S_{n-1}x_{n-1}X_{n-1}W_{n-1}S_{n-1}$$

has the same three occurrences of the factor S'_{n-1} as V_{n-1}. So, if U^3 contains S'_{n-1}, then the middle occurrence of U is inside S'_{n-1}. But this is impossible because S'_{n-1} is a rather short suffix of W_{n-1} and the whole word X_nW_n is cube-free. Therefore, U^3 should end inside the prefix $\bar{x}_{n-1}X_{n-1}W_{n-1}S_{n-1}$ of X_nW_n, like in the following picture.



Using the same parity argument as above, we conclude that the word $S'_nx_nX_n = S'_nP'_n$ is cube-free and, moreover, U^3 should contain the prefix *aabaa* of the word W_{n-1}. Two cases are to be considered: either *aabaa* is a factor of U or *aabaa* occurs in U^3 only twice, on the borders of consecutive U's. The second case is impossible, because two closest occurrences of *aabaa* in W_{n-1} are separated by the factor *babaababbaabbabaabaabb* which does not contain P'_n as a suffix. For the first case, we get that some (not the leftmost) occurrence of *aabaa* in U^3 is preceded by the concatenation of some suffix of S'_n and the word P'_n. If this occurrence of *aabaa* is a prefix of some W_0, then it is preceded by some P'_k, $k < n$. But P'_k is not a suffix of P'_n, a contradiction. The remaining position for this occurrence of *aabaa* is the border of some words S'_k and P'_k. But then S'_k contains the factor which is on the border between S'_n and P'_n, and the parity argument shows that S'_k cannot be partitioned into 2-blocks. This final contradiction shows that U^3 cannot be a factor of V_n. The lemma is proved. □

By construction, the word X_n is the fixed left extension of W_n. Now we consider the second step, that is, the completion of such "almost uniquely" extendable word W_n to a premaximal word. The main idea is the same as at the first step. In order to obtain a premaximal word of level n, we build the word W_{n+1} in $n+1$ iterations by scheme (2a) and then prohibit the extension of W_{n+1} by the first letter of the word P_n. We denote the obtained premaximal word of level n by $\overline{W}_n$. Then

$$\overline{W}_n = \underbrace{W_{n+1}\overline{S}_n}\underbrace{P_nW_{n+1}\overline{S}_n}\underbrace{P_nW_{n+1}\overline{S}_n}, \tag{5}$$

where $\overline{S}_n$ is a "buffer" inserted similarly to S_n in order to avoid cubes at the border of the occurrences of W_{n+1} and P_n. In contrast to the first step, we do not need to build a cube-free right-infinite word, because the construction (5) is used only once. The form of the word $\overline{S}_n$ depends on the last iteration according to Table 1; this dependence is described in Table 2. We choose $\overline{S}_n$ to be the left extension of the word P_n within ${}^{a}_{\infty}T$ (recall that $P_n = {}^{a}_{\infty}T(n+1\ldots1)$).

The above idea works without additional gadgets in all cases when $|X_n| = n$. Due to the following obvious remark, it is enough to construct left premaximal words of level n for all n such that $|X_n| = n$; hence, we do not consider constructing the words $\overline{W}_n$ for other values of n.

Table 2: the “final” suffixes $\overline{S}_n$ for the corresponding iterations from Table 1. The first column contains the number of the last iteration.

Iteration no. (n)	Prohibitions (Start)	$\overline{S}_{n-1}$
k		
$k+1$	$\overline{x}$	$x\overline{x}$
$k+3$	x	$\overline{x}$
$k+4$	x	λ
$k+5$	$\overline{x}$	$x\overline{x}\overline{x}x$
$k+7$	$\overline{x}$	$x\overline{x}$
$k+9$	x	$\overline{x}x$
$k+11$	$\overline{x}$	x
$k+12$	$\overline{x}$	λ
$k+13$	x	λ
$k+15$	x	$\overline{x}$
$k+16$	x	λ
$k+18$	$xx\overline{x}$	$x\overline{x}$
$k+19$		
$k+20$	$\overline{x}$	λ
$k+23$	$\overline{x}\overline{x}x$	$\overline{x}x$
$k+25$	$\overline{x}$	$x\overline{x}$
$k+27$	x	$\overline{x}$
$k+28$	x	λ
$k+29$	$\overline{x}$	λ
$k+31$	$\overline{x}$	x

Iteration no. (n)	Prohibitions (Start)	$\overline{S}_{n-1}$
k	$\overline{x}$	λ
$k+1$		
$k+3$	$\overline{x}\overline{x}x$	$\overline{x}$
$k+4$	x	λ
$k+5$		
$k+7$	$xx\overline{x}$	$x\overline{x}$
$k+9$	x	$\overline{x}x$
$k+11$	$\overline{x}$	x
$k+12$	$\overline{x}$	λ
$k+13$	x	λ
$k+15$	x	$\overline{x}$
$k+16$	x	λ
$k+18$		
$k+19$	$xx\overline{x}$	x
$k+20$	$\overline{x}$	λ
$k+23$	$\overline{x}\overline{x}x$	$\overline{x}x$
$k+25$	$\overline{x}$	$x\overline{x}$
$k+27$	x	$\overline{x}$
$k+28$	x	λ
$k+29$	$\overline{x}$	λ
$k+31$	$\overline{x}$	$x\overline{x}$

Remark 5. *In order to prove the Theorem 2, it is sufficient to show the existence of left premaximal words of level n for infinitely many different values of n. Indeed, if a word W is left premaximal of level n and* $a_1 \cdots a_n W$ *is a left maximal word, then the word* $a_n W$ *is left premaximal of level* $n-1$.

Using the facts that $W_{n+1} \in \mathsf{CF}$, $\overline{S}_n P_n \in \mathsf{TM}$, and the suffix S'_n of W_{n+1} has no long Thue-Morse factors (this is the property of any ψ-image), we prove the following lemma. The proof resembles the one of Lemma 3.

Lemma 4. *The word* $X_n\overline{W}_n$ *is cube-free for all* $n \in \mathbb{N}_0$.

Since the word $P_n\overline{W}_n$ is a cube by (5) and at the same time $P_n = X_{n+1}$ is the fixed left context of W_{n+1}, we conclude that X_n is the longest left context of the word $\overline{W}_n$. Theorem 2 is proved.

Remark 6. *For any n, the word* $\mathrm{rev}(\overline{W}_n) = \overline{W}_n(|\overline{W}_n|)\cdots\overline{W}_n(1)$ *is right premaximal of level n.*

Remark 7. *Our construction provides an upper bound for the length of the shortest left premaximal word of any given level n. The results of [4] suggest that this length is exponential in n. Let* $l(n) = |W_n|$*. For nontrivial iterations, we have* $l(n) = 3l(n-1) + O(n)$*. It is well known that two successive letters in the Thue-Morse word are equal with probability* $1/3$*. Thus, to obtain* W_n*, we make approximately* $2n/3$ *nontrivial iterations. So,* $l(n)$ *is exponential at base* $3^{2/3} \approx 2.08$*. The same property holds for* $|\overline{W}_n| = 3l(n+1) + O(n)$*. It is interesting whether this asymptotics is the best possible.*

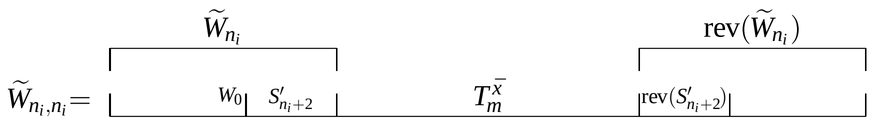
Sketch of the proof of Theorem 3. Similar to Remark 5, it is enough to build premaximal words of level (n_i, n_i) for some infinite sequence $n_1 < n_2 < \ldots < n_i < \ldots$ of positive integers. We take $n_i = 32i + 3$ (Table 2 indicates that $\overline{S}_{n_i} = \lambda$, which makes the construction easier). The natural idea is to concatenate left premaximal and right premaximal words through some "buffer" word. But we cannot use the words $\overline{W}_n$ for this purpose, because all words $X_n\overline{W}_n$ appear to be right maximal.

So, we modify the last step in constructing left premaximal words as follows. The proof of Lemma 3 implies that the word $X_nW_nS_n\cdots S_{n+l}$ is cube-free for any l. So, we put

$$\widetilde{W}_{n_i} = \underbrace{W_{n_i+1}S_{n_i+1}S_{n_i+2}}\,\underbrace{P_{n_i}W_{n_i+1}S_{n_i+1}S_{n_i+2}}\,\underbrace{P_{n_i}W_{n_i+1}S_{n_i+1}S_{n_i+2}}.$$

By Table 1, $S_{n_i+3} = \lambda$ and $S_{n_i+4}(1) \neq S_{n_i+1}(1) = x$. The proof of the fact that $X_{n_i}\widetilde{W}_{n_i} \in \mathsf{CF}$ reproduces the proof of Lemma 4. Recall that $S_{n_i+1}(1) = P_{n_i}(1)$ by (S1), yielding that this letter breaks the period of W_{n_i+1} (see (2b)). On the other hand, the letter $\bar{x}$ breaks the global period of the word $\widetilde{W}_{n_i}$. Hence, the condition $X_{n_i+1}W_{n_i+1}S_{n_i+1}\cdots S_{n_i+l} \in \mathsf{CF}$ implies $X_{n_i}\widetilde{W}_{n_i}S_{n_i+3}\cdots S_{n_i+l} \in \mathsf{CF}$ for any l. Thus, $\widetilde{W}_{n_i}$ is infinitely extendable to the right, left premaximal word of level n_i.

Choose an even m such that $|X_{n_i}\widetilde{W}_{n_i}| < 2^{m-2}$ and consider the word $\widetilde{W}_{n_i,n_i} = \widetilde{W}_n T_m^{\bar{x}} \operatorname{rev}(\widetilde{W}_n)$:



It remains to prove that the word $X_{n_i}\widetilde{W}_{n_i,n_i}\operatorname{rev}(X_{n_i})$ is cube-free. By the choice of m and overlap-freeness of $T_m^{\bar{x}}$, no cube can contain the factor $T_m^{\bar{x}}$. So, by symmetry, it is enough to check that the word $U = X_{n_i}\widetilde{W}_{n_i}T_m^{\bar{x}}$ is cube-free. Assume to the contrary that it contains a cube YYY. Recall that the word $X_{n_i}\widetilde{W}_{n_i}$ is cube-free. Since the first letter of $T_m^{\bar{x}}$ breaks the period of $X_{n_i}\widetilde{W}_n$, one has $|Y| < \operatorname{per}(\widetilde{W}_{n_i})$. Consider the rightmost factor *aabaa* in U; it is inside the factor W_0 immediately before the suffix S'_{n_i+2} of $\widetilde{W}_n$. If this factor belongs to YYY, then $|Y|$ symbols to the left we have another *aabaa*, followed by S'_{n_i+2}. Then $|Y| = \operatorname{per}(\widetilde{W}_{n_i})$, a contradiction. Hence, YYY has no factors *aabaa*, i.e., is a factor of $abaaba\,S'_{n_i+2}T_m^{\bar{x}}$. One can check that the word S'_{n_i+2} contains no Thue-Morse factors of length > 48. The shorter factors can be checked by brute force.

Thus, the word $\widetilde{W}_{n_i,n_i}$ is premaximal of level (n_i, n_i). The theorem is proved. □

References

[1] J.-P. Allouche, J. Shallit (2003): *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge Univ. Press, doi:10.1017/CBO9780511546563.

[2] D. R. Bean, A. Ehrenfeucht, G. McNulty (1979): *Avoidable patterns in strings of symbols*, Pacific J. Math. **85**, 261–294.

[3] F.-J. Brandenburg (1983): *Uniformly growing k-th power free homomorphisms*, Theor. Comput. Sci. **23**, 69–82, doi:10.1016/0304-3975(88)90009-6.

[4] J. D. Currie (1995): *On the structure and extendability of k-power free words*, European J. Comb. **16**, 111–124, doi:10.1016/0195-6698(95)90051-9.

[5] J. D. Currie, N. Rampersad (2009): *There are k-uniform cubefree binary morphisms for all* $k \geq 0$, Discrete Appl. Math. **157**, 2548–2551, doi:10.1016/j.dam.2009.02.010. Available at http://arxiv.org/abs/0812.4470v1.

[6] R. M. Jungers, V. Y. Protasov, V. D. Blondel (2009): *Overlap-free words and spectra of matrices*, Theor. Comput. Sci. **410**, 3670–3684, doi:10.1016/j.tcs.2009.04.022. Available at http://arxiv.org/abs/0709.1794.

[7] M. Lothaire (1983): *Combinatorics on words*, Addison-Wesley, Reading, doi:10.1017/CBO9780511566097.

[8] A. Restivo, S. Salemi (2002): *Words and Patterns*, Proc. 5th Int. Conf. Developments in Language Theory. Springer, Heidelberg, 117–129. (LNCS Vol. **2295**), doi:10.1007/3-540-46011-X_9.

[9] G. Richomme, F. Wlazinski (2000): *About cube-free morphisms*, Proc. STACS'2000. Springer, Berlin, 99–109. (LNCS Vol. **1770**), doi:10.1007/3-540-46541-3_8.

[10] P. Séébold (1984): *Overlap-free sequences*, Automata on Infinite Words. Ecole de Printemps d'Informatique Theorique, Le Mont Dore. Springer, Heidelberg, 207–215. (LNCS Vol. **192**).

[11] A. M. Shur (1998): *Syntactic semigroups of avoidable languages*, Siberian Math. J. **39** (1998), 594–610.

[12] A. M. Shur (2000): *The structure of the set of cube-free Z-words over a two-letter alphabet*, Izv. Math. **64**(4), 847–871, doi:10.1070/IM2000v064n04ABEH000301.

[13] A. M. Shur (2009): *Two-sided bounds for the growth rates of power-free languages*, Proc. 13th Int. Conf. on Developments in Language Theory. Springer, Berlin, 466–477. (LNCS Vol. **5583**), doi:10.1007/978-3-642-02737-6_38.

[14] A. M. Shur (2011): *Deciding context equivalence of binary overlap-free words in linear time*, Semigroup Forum. (Submitted)

[15] A. Thue (1912): *Über die gegenseitige Lage gleicher Teile gewisser Zeichentreihen*, Norske Vid. Selsk. Skr. I, Mat. Nat. Kl. **1**. Christiana, 1–67.