Cyclically Repetition-free Words on Small Alphabets

Tero Harju

Department of Mathematics, University of Turku and Turku Centre for Computer Science (TUCS), Finland

Dirk Nowotka[1]

Institute for Formal Methods in Computer Science (FMI), Universität Stuttgart, Germany

---

Abstract

All sufficiently long binary words contain a square but there are infinite binary words having only the short squares 00, 11 and 0101. Recently it was shown by J. Currie that there exist cyclically square-free words in a ternary alphabet except for lengths 5, 7, 9, 10, 14, and 17. We consider binary words all conjugates of which contain only short squares. We show that the number $c(n)$ of these binary words of length n grows unboundedly. In order for this, we show that there are morphisms that preserve circular square-free words in the ternary alphabet.

Key words: combinatorics on words, repetitions, conjugates, square-free words, cyclically square-free, almost square-free words
2000 MSC: 68R15

---

1. Introduction

We shall consider binary ($w \in \{0,1\}^*$) and ternary ($w \in \{0,1,2\}^*$) words. A word u is a *factor* of a word w if there are words w_1 and w_2 such that $w = w_1uw_2$. In this case, u *occurs* in w. Two words u and v are *conjugates* if $u = xy$ and $v = yx$ for some words x and y. The *conjugacy class* of a word w consists of the words that are conjugates of w. For a given lexicographic order on words, the conjugacy class of any primitive word has a minimal element, which is called a *Lyndon word*. A nonempty factor u^2 $(= uu)$ of a word w is a *square* in w. The word w is *square-free* if it has no squares. Moreover, w is *cyclically square-free* if all of its conjugates are square-free.

While each binary word $w \in \{0,1\}^*$ of length at least four contains a square, R. Entringer, D. Jackson, and J. Schatz [3] showed that there exists an infinite word with only 5 different squares. Later A. Fraenkel and J. Simpson [4] showed

Email addresses: `harju@utu.fi` (Tero Harju), `nowotka@fmi.uni-stuttgart.de` (Dirk Nowotka)

[1]This work has been supported by the DAAD project D/08/08853 PPP Finland.

n	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$c(n)$	3	2	2	2	1	0	0	0	3	0	1	0	0	0

n	18	19	20	21	22	23	24	25	26	27	28	29
$c(n)$	0	2	1	0	0	0	3	0	0	0	1	0

n	30	31	32	33	34	35	36
$c(n)$	1	0	0	0	0	0	2

Table 1: Curious sequence of numbers of cyclically almost square-free binary words.

that there exists an infinite binary word having only the three squares 00, 11, and 0101. We say that a binary word w is *almost square-free* if its squares belong to the set $\{00, 11, 0101\}$ – but we do not allow the square 1010.

Theorem 1 (Fraenkel–Simpson)**.** *For each $n \geq 1$, there exists an almost square-free binary word of length n.*

A simplified proof of Theorem 1 was given by N. Rampersad, J. Shallit, and M.-w. Wang [7] which was further shortened by the present authors in [5]. In this paper we consider cyclically words with short squares. The problem was motivated by the following result due to J. Currie [2].

Theorem 2 (Currie)**.** *There exists a cyclically square-free ternary word w of length n if and only if $n \notin \{5, 7, 9, 10, 14, 17\}$.*

A word w is *cyclically almost square-free* if its conjugates are all almost square-free. We shall show in Theorem 8 that there are unboundedly long cyclically almost square-free binary words.

The exception list of lengths for cyclically almost square-free binary words is much more extensive than the list for cyclically square-free ternary words given by Currie. Indeed, it is an open problem to characterize the set L_{cyc} of lengths n for which there exists a cyclically almost square-free binary word of length n. Also, even for each length $n \in L_{\text{cyc}}$ there seems to be only a small number of examples as seen from Table 1.

Let $c(n)$ denote the number of conjugacy classes of cyclically almost square-free binary words of length n. Thus $c(n)$ equals the number of cyclically almost square-free binary Lyndon words having length n.

Remark 3. One can check that every almost square-free word w (not necessarily cyclic) that omits either 000 or 111 as factors is not longer than 21. The longest such words are of length 21:

$$001000110010110001101 \quad \text{and} \quad 001000110010110001011$$

and the variants obtained by renaming and reversal. Hence a Lyndon representative of a cyclically almost square-free binary word w of length at least 22 starts with 11100 when the order is given as $1 \prec 0$. Indeed, it cannot start with 11101 since it then has a conjugate starting with 0111011 which gives a contradiction at the next bit.

Example 4. Let us consider some examples of cyclically almost square-free binary words. We choose the ordering $1 \prec 0$ for the alphabet for our own convenience.

The Lyndon representatives of length $n = 12$ are the following three words:

$$111001011000\,,$$
$$111000101100\,,$$
$$111000110010\,.$$

The Lyndon representatives of length $n = 24$ are the following words:

$$111001011001110001011000\,,$$
$$111001011100011001011000\,,$$
$$111000110010111000101100\,.$$

There are, however, only two Lyndon representatives of length $n = 36$:

$$111001011001110001100101110001011000\,,$$
$$111001011100010110011100011001011000\,.$$

Despite of Table 1 suggesting that the number of cyclically almost square-free binary words decreases as the length grows, we will show

Theorem 5. *The function $c(n)$ is unbounded:*

$$\limsup_{n\to\infty} c(n) = \infty\ .$$

A mapping $\xi\colon X^* \to Y^*$ is called a *morphism* if $\xi(uv) = \xi(u)\xi(v)$, and ξ is called a *uniform morphism* if additionally we have for some k that $|\xi(a)| = k$ for all $a \in X$.

Now consider a uniform morphism $\xi\colon \{0,1,2\}^* \to \{0,1\}^*$ that takes cyclically square-free ternary words to cyclically almost square-free binary words. Such a morphism can be found by composing β from Section 3 with α from Section 2 below, that is, $\xi(w) = \alpha(\beta(w))$. Note that this morphism is uniform since $|\beta(0)|_i = |\beta(1)|_i = |\beta(2)|_i$ for every $i \in \{0,1,2\}$ where $|w|_a$ denotes the number of occurrences of a in w. Let u and v be two different cyclically square-free ternary words of the same length. Then $\xi(u)$ and $\xi(v)$ are two different cyclically almost square-free binary words of the same length. Hence, Theorem 5 follows from the next result. Let $c_3(n)$ denote the number of cyclically square-free ternary Lyndon words of length n w.r.t. some fixed order.

Theorem 6. *The function $c_3(n)$ is unbounded:*

$$\limsup_{n\to\infty} c_3(n) = \infty\ .$$

This result will be proved in Section 3. We also state the following conjecture.

Conjecture 7. *There exists an integer N such that $c(n) > 0$ for all $n \geq N$.*

2. On Cyclically Binary Words with Short Squares

The following theorem is proven in this section.

Theorem 8. *There are unboundedly long cyclically almost square-free binary words.*

Before we prove Theorem 8 let us recall a morphism from [5] that maps square-free ternary words to almost square-free binary words.

Let $\alpha\colon \{0,1,2\}^* \to \{0,1\}^*$ be the (nonuniform) morphism defined by

$$\begin{aligned}
\alpha(0) = A &:= 1^3 0^3 1^2 0^2 1 0 1^2 0^3 1^3 0^2 1 0\,,\\
\alpha(1) = B &:= 1^3 0^3 1 0 1^2 0^3 1^3 0^2 1 0 1^2 0^3 1 0\,,\\
\alpha(2) = C &:= 1^3 0^3 1^2 0^2 1 0 1^2 0^3 1 0 1^3 0^2 1 0 1^2 0^2\,.
\end{aligned}$$

We notice in passing that these words are almost square-free, and the words A and C are cyclically almost square-free, but B is not. Indeed, B has a conjugate 1000101110001011000111100101 with the long square $(10001011)^2$ as its prefix.

The following result was shown in [5].

Theorem 9. *Let $w \in \{0,1,2\}^*$. Then w is a square-free ternary word if and only if $\alpha(w)$ is an almost square-free binary word.*

We now turn to the proof of the announced result.

Proof of Theorem 8. Let w be a cyclically square-free ternary word provided by Theorem 2, and consider the binary word $\alpha(w)$. Assume that $|w| \geq 2$ w.l.o.g. By Theorem 9, $\alpha(w)$ is almost square-free. The claim follows when $\alpha(w)$ is shown to be cyclically almost square-free. Assume, on the contrary, that $\alpha(w)$ has a conjugate v that is not almost square-free. Without loss of generality, we can assume that v has a square as a suffix, say

$$v = su^2\,,$$

where u^2 is a shortest possible square in the conjugates of $\alpha(w)$ with $u \notin \{0,1,01\}$. One easily checks that $|u| \geq 8$ by considering the words $\alpha(r)$ for $|r| \leq 2$ (see also the comment above Theorem 9). Since w is cyclically square-free, it follows that $v \neq \alpha(w')$ for all conjugates w' of w.

Denote $\Delta = \{A, B, C\}$. We have the following *marking property* of $1^3 0^3$:

$$1^3 0^3 \text{ occurs in cyclic words from } \Delta^* \text{ only as a prefix of } A,\ B, \text{ or } C.$$

Let z be the shortest prefix of v, say $v = zt$, such that the conjugate tz is in Δ^*. In particular, there exists an $X \in \Delta$ such that $X = yz$ for some y.

Since u^2 is not a factor of the conjugate tz, we must have $|s| < |z|$, say $z = sz'$. Therefore, $u^2 = z't = z'x'y$ for some word x'. However, the marking property and $|u| \geq 8$ and $|w| \geq 2$ imply $|u| > |y|$ and, hence,

$$u = z'xy \quad \text{and} \quad X = ysz'$$

for some prefix x of a word in Δ^*. Now $tz = xyz'xyz \in \Delta^*$ which ends with the word $X = yz$. It follows that $xyz'x \in \Delta^*$, i.e., x occurs as a suffix and a prefix in Δ^*. This implies that $x \in \Delta^*$ by the marking property. Hence also for the middle part $yz' \in \Delta^*$. Since yz' is shorter than X, it follows that $yz' \in \Delta$. Now both yz' and ysz' are in Δ. This would imply that $|s| = 3$ or 6; however there is no solution for these parameters in Δ. (The length of the longest common prefix, rep. suffix, of two different words of Δ is 18, resp. 4.) $\square$

3. On the Number of Cyclically Square-Free Words

A morphism is called (cyclically) square-free whenever the image of any (cyclically) square-free word is itself (cyclically) square-free. In this section we will construct a set of uniform cyclically square-free morphisms on $\{0,1,2\}^*$ such that an arbitrary number of cyclically square-free words of the same length can be generated.

We start from certain square-free factors taken from an infinite square-free word in order to find substitutions that preserve square-freeness. Then we introduce several markers that allow us to both ensure cyclically square-freeness and the construction of arbitrarily many different substitutions without sacrificing the preservation of square-freeness.

Thue gave in [8] the following morphism ϑ on $\{0,1,2\}^*$ which generates the infinite *Thue word* $\mathbf{t}$ when iterated starting in 0. Consider

$$\vartheta(0) = 012\,, \qquad \vartheta(1) = 02\,, \qquad \vartheta(2) = 1$$

which gives

$$\mathbf{t} = \lim_{k\to\infty} \vartheta^k(0) = \underline{0120210}121020120210\underline{201210}12021\underline{012102}012\cdots \tag{1}$$

where we point out three underlined factors of $\mathbf{t}$ which will be used further below. It is well-known that $\mathbf{t}$ is square-free. The following morphism $\eta\colon \{0,1,2\}^* \to \{0,1\}^*$ maps $\mathbf{t}$ to an overlap-free binary word [6], the so called *Thue-Morse word*,

$$\eta(0) = 011\,, \quad \eta(1) = 01\,, \quad \eta(2) = 0\,.$$

A word is called *overlap-free* if it has no overlapping factors, i.e., if no factor of the form $awawa$ occurs where a is a letter and w is a (possibly empty) word. In particular the words in the following set do not occur in $\mathbf{t}$:

$$T_{\text{no}} = \{010\,, 212\,, 1021\,, 1201\}\,. \tag{2}$$

Indeed, $\eta(010) = 01101011$ which contains the overlap 10101. Assume that contrary to the claim 212 occurs in $\mathbf{t}$. Then it must be preceded and succeeded by 0 since $\mathbf{t}$ is square-free. But, $\eta(02120) = 0110010011$ contains the overlap 1001001; a contradiction. If 1021 occurs in $\mathbf{t}$, then it must be preceded by 2 and succeeded by 0 by the previous arguments. But, then $\mathbf{t}$ contains the square 210210; a contradiction. A similar argument holds for the word 1201.

So far, we have identified in $T_{\rm no}$ square-free words that do not occur in $\mathbf{t}$. They will serve as markers in the proof of Theorem 6 below.

Iterating ϑ gives

$$\begin{aligned}
\vartheta(0) &= 012\\
\vartheta^2(0) &= 012021\\
\vartheta^3(0) &= 012021012102\\
\vartheta^4(0) &= 012021012102012021020121\\
&\vdots
\end{aligned}$$

and

$$\begin{aligned}
\vartheta(1) &= 02\\
\vartheta^2(1) &= 0121\\
\vartheta^3(1) &= 01202102\\
\vartheta^4(1) &= 0120210121020121\\
&\vdots
\end{aligned}
\qquad\text{and}\qquad
\begin{aligned}
\vartheta(2) &= 1\\
\vartheta^2(2) &= 02\\
\vartheta^3(2) &= 0121\\
\vartheta^4(2) &= 01202102\\
&\vdots
\end{aligned}$$

Consider the words $\vartheta^4(0)$ and $\vartheta^4(1)$ and $\vartheta^4(2)$ that start with 012021 and that all have an occurrence in $\mathbf{t}$ followed by 0120. Indeed, $\vartheta^6(0)$ is a prefix of $\mathbf{t}$ and $\vartheta^6(0) = \vartheta^4(012021) = \vartheta^4(0)\vartheta^4(1)\vartheta^4(2)\vartheta^4(0)\vartheta^4(2)\vartheta^4(1)$.

Let δ be a morphism on $\{0,1,2\}^*$ defined by

$$\begin{aligned}
\delta(0) &= (012)^{-1}\vartheta^4(0)012 = 021012102012021020121012\,,\\
\delta(1) &= (012)^{-1}\vartheta^4(1)012 = 0210121020121012\,,\\
\delta(2) &= (012)^{-1}\vartheta^4(2)012 = 02102012\,.
\end{aligned}$$

We have

Claim 10. *The δ-image of each factor of $\mathbf{t}$ occurs itself in $\mathbf{t}$ followed by* 021.

Indeed, let w be a factor of $\mathbf{t}$, then $\vartheta(w)$, and hence, $\vartheta^4(w)$ is a factor of $\mathbf{t}$. Therefore, $(012)^{-1}\vartheta^4(w)$ is a factor of $\mathbf{t}$ which proves the claim since $(012)^{-1}\vartheta^4(wa)$ occurs in $\mathbf{t}$, for some letter a such that wa occurs in $\mathbf{t}$, and 012021 is a prefix of $\vartheta^4(a)$.

Consider the factors 0201210 and 0120210 and 0121020 of $\mathbf{t}$ as marked in (1). Note that these factors are of the same length and have the same number of occurrences of 0, 1, and 2, respectively.

Let us define the following uniform morphism β on $\{0,1,2\}^*$ where the length of the images of letters is $|\beta(i)| = 122$:

$$\begin{aligned}
\beta(0) &= \delta(0201210)01\,,\\
\beta(1) &= \delta(0120210)01\,,\\
\beta(2) &= \delta(0121020)01\,.
\end{aligned}$$

Remark 11. We note that for different letters a and b, the prefixes of length 61 and the suffixes of length 62 of $\beta(a)$ and $\beta(b)$ are different.

Claim 12. *The images $\beta(i)$ are cyclically square-free for all $i \in \{0, 1, 2\}$.*

Proof. The claim can be easily proven by inspection or a computer test. However, we give an alternative proof for illustrating some arguments also used later below.

By Claim 10 the prefix $\beta(i)1^{-1}$ of $\beta(i)$ is a factor of $\mathbf{t}$ for all $i \in \{0, 1, 2\}$. The words $\beta(i)$ end with 1201 which is in the set T_{no} of forbidden factors of $\mathbf{t}$. It follows that the words $\beta(i)$ are square-free. It is also straightforward to verify that $\beta(i)$ are cyclically square-free. Indeed, any cyclic square x^2 must contain the last letter 1 of $\beta(i)$. The case where $|x| < 6$ is easily checked by hand. Note that $1\beta(i)1^{-1}$ begins with 1021 and $\beta(i)$ ends with 1201. Hence, if $|x| \geq 6$ then x contains 1021 or 1201. But $1021, 1201 \in T_{\text{no}}$ and therefore they occur at most once in any conjugate of $\beta(i)$ which contradicts that x^2 occurs in a conjugate of $\beta(i)$. This concludes the proof of Claim 12. $\square$

Let π be any permutation on $\{0, 1, 2\}$. We define the following morphisms

$$\beta_\pi(i) = \beta(\pi(i))$$

for $i \in \{0, 1, 2\}$. Before we show that every β_π is cyclically square-free, we recall the following theorem by Thue [8]; see [1] for a slightly improved version.

Theorem 13. *A morphism α is square-free if the following two conditions are satisfied:*

(1) $\alpha(u)$ is square-free whenever u is square-free with $|u| \leq 3$, and

(2) $\alpha(a)$ is not a proper factor of $\alpha(b)$ for any letters a and b.

In order to show that the constructed morphisms are cyclically square-free we state the following result.

Proposition 14. *A morphism α is cyclically square-free if the following two conditions are satisfied:*

(1) α is square-free and

(2) $\alpha(a)$ is cyclically square-free for all letters a.

Proof. Let $w_{(i)}$ denote the ith letter of the word w. Consider a cyclically square-free word w of length n and suppose, contrary to the claim, that $\alpha(w)$ is not cyclically square-free. Let x^2 be a shortest square in a conjugate of $\alpha(w)$. Let $w' = w_{(i)}w_{(i+1)} \cdots w_{(n)}w_{(1)} \cdots w_{(i-1)}w_{(i)}$. Then x^2 occurs in $\alpha(w')$ for some i. Now, w' is square-free if w is cyclically square-free, except if $n = 1$; a contradiction of either (1) or (2) in any case. $\square$

It is now straightforward to establish the cyclically square-freeness of any β_π which implies Theorem 6.

Lemma 15. *Let π be any permutation on $\{0, 1, 2\}$. Then β_π is a cyclically square-free morphism.*

Proof. We begin by showing that β_π is square-free. By Theorem 13 the square-freeness of β_π can be checked by hand. However, this is cumbersome and therefore we give an alternative proof avoiding Theorem 13. Suppose contrary to the claim that $\beta_\pi(w)$ contains a square x^2 where w is square-free. Surely, x^2 does not occur in $\beta_\pi(a)$ for any letter a by Claim 12. Note that 1201021 occurs in $\beta_\pi(w)$ only at a point where two β_π images of letters are concatenated. Assume that $|x| \geq 6$; the smaller cases can be easily excluded. Then, as in the proof of Claim 12, x contains 1201 or 1021. Both 1021 and 1201 mark the beginnings and ends of the β_π images of letters, and hence, β_π is injective. Let $u \in \{1021, 1201\}$ be such that u occurs in x. Suppose $u = 1201$, the other case follows analogous reasons. Then either u occurs in the beginning or end of x. We have then

$$\begin{aligned} x &= yu\beta_\pi(w_{(j)})\beta_\pi(w_{(j+1)}) \cdots \beta_\pi(w_{(j+r)})z \\ &= yu\beta_\pi(w_{(j+r+2)})\beta_\pi(w_{(j+r+3)}) \cdots \beta_\pi(w_{(j+2r+2)})z \,, \end{aligned}$$

where $1 < j < |w| - r$ and $-1 \leq r < |w|/2$ and $zyu = \beta_\pi(w_{(j+r+1)})$. Here the word u is a marker and hence the β_π-images are aligned in the two occurrences of x, i.e., $w_{(j+\ell)} = w_{(j+r+2+\ell)}$ for $\ell = 0, \ldots, r$. Now, $w_{(j-1)} \neq w_{(j+r+1)}$ and $w_{(j+2r+3)} \neq w_{(j+r+1)}$, since w is square free. However, yu is a suffix of $\beta_\pi(w_{(j-1)})$ and, by Remark 11, $|yu| \leq 61$. Also, z is a prefix of $\beta_\pi(w_{(j+2r+3)})$ and thus $|z| \leq 60$. But now $|zyu| \leq 121$ gives a contradiction with $zyu = \beta_\pi(w_{(j+r+1)})$. Therefore, β_π is square-free. Claim 12 and Proposition 14 conclude the proof. $\square$

Now, Theorem 6 follows.

Theorem 6. *The function $c_3(n)$ is unbounded:*

$$\limsup_{n\to\infty} c_3(n) = \infty \ .$$

Proof. Indeed, the image of the cyclically square-free word 021 under β_π gives a different cyclically square-free word for any permutation π by Lemma 15. Each of these cyclically square-free words starts with 021, and hence, gives six new cyclically square-free words (one for each β_π). This process can be iterated arbitrarily many times. The uniformness of β_π ensures that the images of a word are of the same length for each π. The number of different cyclically square-free words after k iterations equals 6^k and they are of length $3 \cdot 122^k$. $\square$

Remark 16. We mention another approach to show Theorem 6 using substitutions instead of morphisms. Consider the following words of length 18:

$$\begin{aligned} u_1 &= 010201210201021012 \\ u_2 &= 010201210212021012 \\ v_1 &= 010201202120121012 \\ v_2 &= 010201202102010212 \\ w_1 &= 010201202101210212 \end{aligned}$$

The substitution $s\colon \{0,1,2\}^* \to 2^{\{0,1,2\}^*}$ defined by

$$s(0) = \{u_1, u_2\}, \quad s(1) = \{v_1, v_2\}, \quad s(2) = \{w_1\}$$

preserves cyclically square-freeness, i.e., if w is cyclically square-free, then so is each $u \in s(w)$. Now, the sequence $|s^n(0)|$ of elements in $s^n(0)$ is strictly increasing with increasing n, and thus proves our theorem.

This claim on s needs to be verified in order to make this approach work. This can be shown by similar techniques as the ones used above in the proof of Lemma 15.

Acknowledgement

We would like to thank the anonymous referees for their careful work. Their comments greatly helped to improve this article.

References

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

[2] J. D. Currie. There are ternary circular square-free words of length n for $n \geq 18$. *Electron. J. Combin.*, 9(1):Note 10, 7 pp. (electronic), 2002.

[3] R. C. Entringer, D. E. Jackson, and J. A. Schatz. On nonrepetitive sequences. *J. Combin. Theory, Ser. A*, 16:159–164, 1974.

[4] A. S. Fraenkel and J. Simpson. How many squares must a binary sequence contain? *Electronic J. Combin.*, 2(#R2 (electronic)), 1995.

[5] T. Harju and D. Nowotka. Binary words with few squares. *Bull. EATCS*, 89:164–166, 2006.

[6] M. Lothaire. *Combinatorics on Words*, volume 17 of *Encyclopedia of Mathematics*. Addison-Wesley, Reading, MA, 1983. Reprinted in the Cambridge Mathematical Library, Cambridge Univ. Press, 1997.

[7] N. Rampersad, J. Shallit, and M.-w. Wang. Avoiding large squares in infinite binary words. *Theoret. Comput. Sci.*, 339:19–34, 2005.

[8] A. Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske Vid. Skrifter I. Mat.-Nat. Kl., Christiania*, 1:1–67, 1912.