Infinite Binary Words Containing Repetitions of Odd Period

Golnaz Badkobeh[a,c] and Maxime Crochemore[a,b]
[a]King's College London, UK
[b]Université Paris-Est, France
[c]University of Sheffield, UK

January 15, 2015

Abstract

A square is the concatenation of a nonempty word with itself. A word has period p if its letters at distance p match. The exponent of a nonempty word is its length divided by its smallest period. In this article, we give some new results on the trade-off between the number of squares and the number of cubes in infinite binary words whose square factors have odd periods.

Keywords: combinatorics on words, repetitions, repetitive threshold, word morphisms.

1 Introduction

Enumerating the repetitions in infinite words is a classic problem in combinatorics on words that has been studied in depth over the last 100 years (see for example, [10, 3] and references therein).

A square is the concatenation of a nonempty word with itself. Let $g(n)$ be the length of a longest binary word containing at most n distinct squares. Then $g(0) = 3$ (e.g., `010`), $g(1) = 7$ (e.g., `0001000`) and $g(2) = 18$ (e.g., `010011000111001101`).

In 1974, Entringer, Jackson, and Schatz [5] showed that there exists an infinite word with 5 distinct squares. Therefore, they proved that $g(5) = \infty$. Later, Fraenkel and Simpson [6] showed that there exists an infinite binary word that contains only three squares, 00, 11, and 0101, and thus $g(3) = \infty$. A somewhat simplified proof of this result was given by Rampersad, Shallit and Wang [9]. Later, in 2006, Harju and Nowotka [7] gave another simpler proof of this result and, eventually, Badkobeh [2] gave yet another proof exploiting two simple morphisms.

Instead of avoiding all squares, one interesting variation on the problem is to avoid larger repetitions. Entringer, Jackson, and Schatz [5] showed that there exist infinite binary words avoiding squares of period at least three. Later, avoiding large squares was studied by Dekking [4], Rampersad et al. [9], Shallit [11], Ochem [8], and many others.

In this article, we provide some new results as an outcome of studying pattern avoidance from a different point of view. We analyse the possibility of avoiding repetitions of even and odd periods, and further impose a constraint on their maximal exponent.

We show that there exists no infinite 3^+-free binary word avoiding all squares of odd period. We also show that there exists no infinite binary word simultaneously avoiding cubes and squares of even period. Moreover, we show that there exists an infinite 3^+-free binary word avoiding squares of even period.

The trade-off between the maximal period length and the number of repetitions follows a similar trade-off between the number of cubes and the number of distinct squares. A similar study was comprehensively carried out by the first author in [1].

The article is structured as follows. We provide some definitions in Section 2. In Section 3, we present the proof technique that will be used throughout this article. In addition, we prove that there exists no infinite 3^+-free binary word avoiding all squares of odd period and there exists no infinite binary word simultaneously avoiding cubes and squares of even period. In Section 4, we show that in fact there exists an infinite 3^+-free binary word avoiding squares of even period. In Section 5, we reduce the number of repetitions contained in infinite binary words without compromising the constraint on the parity of the periods of squares. We conclude that the minimal number of squares in such words is 7 when only 1 cube occurs. The number reduces to 4 when 2 cubes are allowed in the word. In Section 6, we give a summary of our results.

2 Preliminaries

An alphabet is any non-empty set, the members of which are called letters. A word, or a string, is a sequence of letters drawn from the alphabet. The empty word ϵ is a string of length 0 that is considered to be a word over every alphabet. The length of the word w, denoted by $|w|$, is the number of occurrences of letters in w. For example, $|\texttt{abaca}| = 5$.

We consider the ternary alphabet $\texttt{A} = \{\texttt{a}, \texttt{b}, \texttt{c}\}$, the binary alphabet $\texttt{B} = \{\texttt{0}, \texttt{1}\}$, and the n-ary alphabet Σ_n for $n > 3$.

The word v is called a factor of x if there exist words u and w such that $x = uvw$. In the case $u = \epsilon$ (resp., $w = \epsilon$), v is a prefix (resp., a suffix) of x. A nonempty word x has period p if $x[i] = x[i + p]$ for all i for which the equation is meaningful. The exponent of x is its length divided by its smallest period.

The maximum exponent of a word w is the supremum of $E(x)$, where $E(x)$ is the set of exponents of all finite factors of x.

A *square* is a word of the form xx, where x is a non-empty word. *Cubes* and k-th *powers* are defined accordingly. A word is *overlap-free* if it does not contain any factor of the form $xyxyx$ for a non-empty x. In general, a word is said to be α-free if it contains no factor of the form u^β for any rational number $\beta \geq \alpha$. It is α^+-free if it contains no factor of the form u^β for any rational number $\beta > \alpha$.

A morphism is a map $h : \Sigma_n^* \to \Sigma_m^*$ such that $h(uv) = h(u)h(v)$ for all $u, v \in \Sigma_n^*$. This implies that $h(\epsilon) = \epsilon$. In addition, the morphism h is completely defined by the pairs $(a, h(a))$ for $a \in \Sigma_n$. We refer to images of letters as codewords. If $h(a) = ax$ for some letter $a \in \Sigma_n$, then we say that h is prolongable on a, and we can then iterate h infinitely often to get the fixed point $h^\infty(a) := axh(x)h^2(x)h^3(x)\cdots$. For $q \geq 2$ a morphism h is said to be q-uniform if $|h(a)| = q$ for all $a \in \Sigma_n$. A uniform morphism h is *synchronising* when $h(ab) = vh(c)w$ implies that either $v = \epsilon$ and $a = c$ or $w = \epsilon$ and $b = c$, for any $a, b, c \in \Sigma_n$ and $v, w \in \Sigma_m^*$. Notice that a synchronising morphism h is always injective (actually it is injective on the set Σ_n of monoid generators). Moreover, if it is q-uniform then, for each factor u of a word in $h(\Sigma_n^*)$ such that $|u| \geq 2q - 1$, there exists a unique factorisation $u = xh(u')y$ where $u' \in \Sigma_n^*$ and $0 \leq |x|, |y| < q$.

3 Words containing only repetitions of odd period

Here, we study further the infinite binary words and the squares they contain. Looking at the parity of the periods of the squares reveals interesting properties.

Note that the only infinite binary words omitting 00 and 11 are $(01)^\infty$ and $(10)^\infty$, both of which contain 3^+-powers. This proves the following proposition.

Proposition 1. *There exists no infinite 3^+-free binary word avoiding all squares of odd period.*

Proposition 2. *There exists no infinite binary word, simultaneously avoiding cubes and squares xx with $|x| = 2k$ for $k > 0$. The length of a cube-free binary word containing only squares of odd period does not exceed 23.*

Proof. Here, we try to build a binary word that avoids cubes and squares of even period. The following list contains all possible strings with prefix 00, avoiding cubes and squares of even period:

`00100100`	`00110010010`	`0011011001001100`
`001001100`	`0011001001100`	`0011011001001101100`
`00100110110010010`	`001100100110110010010`	`0011011001001101101`
`0010011011001001100`	`001100100110110010011`	`00110110011`
`0010011011001001101`	`001100100110110011`	`001101101`
`00100110110011`	`0011001001101101`	
`001001101101`	`00110110010010`	

The maximum length of these words is 21. This is also true for words starting with 11. Now the only binary words avoiding 00, 11, cubes, and squares of even period are: $\{0, 1, 01, 10, 010, 101\}$. Concatenating these two sets will not produce a word complying with the properties whose length exceeds 23.

□

The remainder of this section is dedicated to demonstrating that if the constraint on the maximal exponent is relaxed so that the word may contain cubes, then avoiding squares of even period becomes possible.

The same technique is used to prove each of the theorems in this article. The technique is stated below. To demonstrate how this technique works, a step-by-step proof is given for Proposition 3, as an example.

Proof Technique. Let g be a synchronising morphism $g : \mathtt{A}^* \to \mathtt{B}^*$, and let s be an infinite square-free word in $\mathtt{A}^*$. Notice that the only squares occurring in $g(s)$ also occur in the images of square-free factors of s of length 3. Therefore, to study the squares contained in $g(s)$ it is enough to look at all the images of triplets in $\mathtt{A}^*$ (a triplet is a word of length 3). This set is finite and therefore it is possible to count all the squares contained in the images of the set. In order to prove the theorems presented in this article, it is sufficient to show that the given morphisms are synchronising. To demonstrate this we look at the images of all the doublets (words of length 2) in $\mathtt{A}^*$ to investigate if they comply with the definition of synchronising morphisms.

In this section, s is any infinite square-free ternary word.

Theorem 1. *There exists an infinite 3^+-free binary word avoiding squares of even period.*

The proof relies on the following synchronising 8-uniform morphism g_1 from $\mathtt{A}^*$ to $\mathtt{B}^*$ defined by:

$$g_1(\mathtt{a}) = \mathtt{11011001},$$
$$g_1(\mathtt{b}) = \mathtt{11001001},$$
$$g_1(\mathtt{c}) = \mathtt{00011000}.$$

Let $Sq = \{\mathtt{00},\ \mathtt{11},\ (\mathtt{001})^2,\ (\mathtt{010})^2,\ (\mathtt{011})^2,\ (\mathtt{100})^2,\ (\mathtt{110})^2,\ (\mathtt{00011})^2,\ (\mathtt{00110})^2,\ (\mathtt{01100})^2,\ (\mathtt{10001})^2,\ (\mathtt{11001})^2\}$ and $C = \{\mathtt{000}, \mathtt{111}, (\mathtt{100})^3\}$.

Proposition 3. *The infinite word $\mathbf{g_1} = g_1(s)$ contains no repetition with exponent greater than 3 and no square uu with $|u| = 2k$ for $k > 0$. Furthermore, $\mathbf{g_1}$ contains only 12 squares, all of which are in Sq. And $\mathbf{g_1}$ contains only 3 cubes, all of which belong to set C.*

Here, for the interest of the reader, we demonstrate how the proof technique works for Proposition 3.

Proof. Let us assume that $g_1(s)$ contains a square $uu \notin Sq$. There are two possibilities: either $|u| > 16$ or $|u| \leq 16$ (uu is a factor of the image of $w \in s$ for $w \leq 3$).

- Case $|u| > 16$:
$$uu = \overbrace{u_1 \underbrace{\cdots} v_1}\overbrace{u_1 \underbrace{\cdots} v_1}$$
where v_1u_1 is a codeword. Then v_1 is not longer than the longest common prefix between two different codewords, that is, $|v_1| \leq 3$. Symmetrically, u_1 is not longer than the longest common suffix of two different codewords, that is, $|u_1| \leq 4$. But then $|v_1u_1| \leq 7$ and cannot be a complete codeword: a contradiction.

- Case $|u| \leq 16$: Here, it is enough to look at images of all $w \in s$ for $w \leq 3$. A simple computer check verifies the fact that all squares contained in these images are in Sq. A cube is an extension of a square, therefore one can easily verify that the number of cubes is 3.

$\square$

The proof of Theorem 1 is a direct consequence of Proposition 3, since the set Sq only contains squares of odd period (1, 3 and 5).

4 Avoiding long repetitions

Looking at the length of periods of squares contained in $\mathbf{g_1}$ (Section 3), one may ask if it is possible to reduce the length of the longest squares in an infinite word without compromising the other conditions imposed on the word. It is trivial to show that there exists no infinite binary word containing only squares of period 1. However, the next theorem shows that we can reduce the longest period to 3.

Theorem 2. *There exists an infinite* 3^+*-free binary word containing only squares of period either 1 or 3.*

The proof relies on the following synchronising 11-uniform morphism g_2 from $\mathtt{A}^*$ to $\mathtt{B}^*$ defined by:

$$g_2(\mathtt{a}) = \mathtt{11001001101},$$
$$g_2(\mathtt{b}) = \mathtt{11001001110},$$
$$g_2(\mathtt{c}) = \mathtt{11001001000}.$$

Proposition 4. *The infinite word* $\mathbf{g_2} = g_2(s)$ *contains no repetition with exponent greater than* 3 *and no square* uu *with* $|u| = 2k$ *for* $k > 0$*. Furthermore,* $\mathbf{g_2}$ *contains only 7 squares:* $\mathtt{00}$, $\mathtt{11}$, $(\mathtt{001})^2$, $(\mathtt{010})^2$, $(\mathtt{011})^2$, $(\mathtt{100})^2$ *and* $(\mathtt{110})^2$*, and 3 cubes:* $\mathtt{000}$, $\mathtt{111}$ *and* $(\mathtt{100})^3$.

The proof of Proposition 4 is very similar to the proof of Proposition 3, therefore we have omitted it from this paper. Theorem 2 follows.

The following synchronising morphism g_3 also generates an infinite 3^+-free binary word containing squares of period 1 and 3 only. Furthermore, the infinite binary word generated by g_3 omits the third cube in $\mathbf{g_2}$. Thus it contains fewer cubes.

The word with 7 squares and 2 cubes. To generate an infinite word with these properties we use the following synchronising 12-uniform morphism g_3 from $\mathtt{A}^*$ to $\mathtt{B}^*$ defined by:

$$\begin{aligned} g_3(\mathtt{a}) &= \mathtt{110110001110}, \\ g_3(\mathtt{b}) &= \mathtt{110111000100}, \\ g_3(\mathtt{c}) &= \mathtt{110111001000}. \end{aligned}$$

Proposition 5. *The infinite word $\mathbf{g_3} = g_3(s)$ contains no repetition with exponent greater than 3 and no square uu with $|u| = 2k$ for $k > 0$. Furthermore, $\mathbf{g_3}$ contains only 7 squares: 00, 11, $(001)^2$, $(011)^2$, $(100)^2$, $(101)^2$ and $(110)^2$, and 2 cubes: 000 and 111.*

As explained in the proof technique, in order to prove Proposition 5 it is sufficient to show that g_3 is synchronising.

5 Reducing the number of repetitions

It is natural to ask if there exists an infinite binary word avoiding squares of even period and containing less than 7 squares or 2 cubes.

Proposition 6. *A binary word avoiding squares of even period that contains at most 6 squares and only one cube has length at most 57.*

Proof. We can build a binary word complying with the desired property using backtracking, meaning that every time we cannot extend the word, we find the rightmost zero and change it to 1. Here is a word of length 57 containing one cube and 6 squares whose period is not even:
001101110011011001001101110010011101100100110110011101100

□

Although Proposition 6 shows that simultaneously reducing the number of squares and cubes is not possible, the following two theorems show that there exist infinite binary words avoiding squares of even periods which either contain only 1 cube and 7 squares, or 2 cubes and less than 7 squares.

Theorem 3. *There exists an infinite 3^+-free binary word with at most one cube, avoiding squares of even period and containing only 7 squares.*

The proof relies on the following synchronising 73-uniform morphism g_4

from $\mathtt{A}^*$ to $\mathtt{B}^*$ defined by:

$$\begin{aligned} g_4(\mathtt{a}) = \;& \mathtt{110110001001101100100011011000100100011} \\ & \mathtt{011001001100010010001100100110110010000}, \\ g_4(\mathtt{b}) = \;& \mathtt{110110001001101100100110001001000110010} \\ & \mathtt{011011000100100011011001001100010010000}, \\ g_4(\mathtt{c}) = \;& \mathtt{110110001001101100100110001001000110110} \\ & \mathtt{010011011000100100011001001100010010000}. \end{aligned}$$

Proposition 7. *The infinite word $\mathbf{g_4} = g_4(s)$ contains no repetition with exponent greater than* 3 *and no square uu with $|u| = 2k$, for $k > 0$. Furthermore, $\mathbf{g_4}$ contains only 7 squares: 00, 11, $(001)^2$, $(010)^2$, $(011)^2$, $(100)^2$, and $(110)^2$, and only one cube 000.*

Theorem 4. *There exists an infinite 3^+-free binary word with at most two cubes, avoiding squares of even period and containing only 4 squares.*

The proof relies on the following synchronising 39-uniform morphism g_5 from $\mathtt{A}^*$ to $\mathtt{B}^*$ defined by:

$$\begin{aligned} g_5(\mathtt{a}) &= \mathtt{111000100110001110010001100100111001000}, \\ g_5(\mathtt{b}) &= \mathtt{111000100111001000110010011100011001000}, \\ g_5(\mathtt{c}) &= \mathtt{111000100111001001100010011100011001000}. \end{aligned}$$

Proposition 8. *The infinite word $\mathbf{g_5} = g_5(s)$ contains no repetition with exponent greater than* 3 *and no square uu with $|u| = 2k$ for $k > 0$. Furthermore, $\mathbf{g_5}$ contains only 4 squares: 00, 11, $(001)^2$, and $(100)^2$, and only two cubes: 000 and 111.*

The following result is verified by computer check:

Fact 1. *A binary word avoiding squares of even period that contains at most 3 squares has length at most 29.*

Here, it is worth mentioning that if the constraint on the parity of the squares period is removed, then the following results were shown by the first author in [1]:

- There exists a 3^+-free infinite binary word with only one cube that contains no more than 4 squares.
- There exists a 3-free infinite binary word with at most 8 squares.

In [1] the first author also demonstrates that these numbers are minimal.

6 Conclusion

In this article, we studied the infinite binary words whose square factors have odd periods. The tables below summarise these results.

	Longest allowed period	Number of cubes	Number of squares	Length of the morphism
Proposition 3	5	3	12	8
Proposition 4	3	3	7	11
Proposition 5	3	2	7	12
Proposition 7	3	1	7	73
Proposition 8	3	2	4	39

Note that all the infinite binary words considered in these proofs are 3^+-free and avoid squares of even period. Therefore all of the propositions mentioned above can prove Theorem 1. The morphisms in Propositions 4, 5 and 7 generate binary words containing 7 squares of periods 1 or 3. The only differences between them are the number of cubes they contain and their codeword lengths. The morphism whose codeword length is longer contains fewer cubes.

A similar comparison was made between two morphisms used in Propositions 5 and 8. Both morphisms generate binary words with only 2 cubes, however, the one with longer codewords (of length 39) contains fewer squares than the one with shorter codewords (of length 12).

	Allowed number of cubes	Minimum number of squares
Theorem 4	2	4
Theorem 3	1	7

References

[1] G. Badkobeh. Fewest repetitions vs maximal-exponent powers in infinite binary words. *Theoretical Computer Science*, 412(48):6625–6633, 2011.

[2] G. Badkobeh. Infinite words containing the minimal number of repetitions. *Journal of Discrete Algorithms*, 20:38–42, 2013.

[3] J. D. Currie and N. Rampersad. Dejean's conjecture holds for $n \geq 30$. *Theoretical Computer Science*, 410(30-32):2885–2888, 2009.

[4] F. M. Dekking. On repetitions of blocks in binary sequences. *Journal of Combinatorial Theory, Series A*, 20(3):292–299, 1976.

[5] R. C. Entringer and D. E. Jackson. On nonrepetitive sequences. *Journal of Combinatorial Theory, Series A*, 16(2):159–164, 1974.

[6] A. S. Fraenkel and J. Simpson. How many squares must a binary sequence contain? *Electronic Journal of Combinatorics*, 2, 1995. `http://www.combinatorics.org/Volume_2/Abstracts/v2i1r2.html`.

[7] T. Harju and D. Nowotka. Binary words with few squares. *Bulletin of the EATCS*, 89:164–166, 2006.

[8] P. Ochem. A generator of morphisms for infinite words. *RAIRO - Theoretical Informatics and Applications*, 40(3):427–441, 2006.

[9] N. Rampersad, J. Shallit, and M.-w. Wang. Avoiding large squares in infinite binary words. *Theoretical Computer Science*, 339(1):19–34, 2005.

[10] M. Rao. Last cases of Dejean's conjecture. *Theoretical Computer Science*, 412(27):3010–3018, 2011.

[11] J. Shallit. Simultaneous avoidance of large squares and fractional powers in infinite binary words. *International Journal of Foundations of Computer Science*, 15(2):317–327, 2004.