

Cubefree words with many squares

James Currie[†] and Narad Rampersad[‡]

Department of Mathematics and Statistics, University of Winnipeg
515 Portage Avenue, Winnipeg, Manitoba R3B 2E9 (Canada)
{j.currie,n.rampersad}@uwinnipeg.ca

received November 26, 2008, revised April 25, 2010, accepted April 28, 2010.

We construct infinite cubefree binary words containing exponentially many distinct squares of length n . We also show that for every positive integer n , there is a cubefree binary square of length $2n$.

Keywords: cubefree word, square

1 Introduction

A *square* is a non-empty word of the form xx , and a *cube* is a non-empty word of the form xxx . An *overlap* is a word of the form $axaxa$, where a is a letter and x is a word (possibly empty). A word is *squarefree* (resp. *cubefree*, *overlap-free*) if none of its factors are squares (resp. cubes, overlaps). For further background material concerning combinatorics on words we refer the reader to [2].

It is well-known that there exist infinite squarefree words over a ternary alphabet and infinite overlap-free words over a binary alphabet. Clearly, any overlap-free word is also cubefree. Any infinite cubefree binary word must contain squares; however, Dekking [9] proved that there exists an infinite cubefree binary word containing no squares xx where the length of x is greater than 3 (see also [14, 15]). In this paper we consider instead the existence of infinite cubefree binary words with many distinct squares.

Most known constructions of infinite cubefree words involve the iteration of a morphism. In the early 80's, Berstel [3] revitalized the study of the construction of words avoiding repetitions by the iteration of morphisms. Words constructed in this manner are often referred to as *infinite D0L words*. Ehrenfeucht and Rozenberg [10, 11, 12] proved several results concerning the factor complexity of infinite D0L words. They showed that any squarefree or cubefree D0L word has $O(n \log n)$ factors of length n . Thus, an infinite cubefree D0L word cannot have many distinct square factors. By contrast, we show here how to construct infinite cubefree binary words containing exponentially many distinct squares of length n .

Other work related to the problems considered here include [1, 7, 8].

Let μ denote the *Thue–Morse morphism*: i.e., the morphism that maps $0 \rightarrow 01$ and $1 \rightarrow 10$. The *Thue–Morse word* is the infinite word

$$\mathbf{t} = 011010011001011010010110 \dots$$

[†]The first author is supported by an NSERC Discovery Grant.

[‡]The second author is supported by an NSERC Postdoctoral Fellowship.

obtained by iteratively applying μ to the word 0. The Thue–Morse word is well-known to be overlap-free, and hence, a fortiori, cubefree [17]. The squares occurring in the Thue–Morse word were characterized by Pansiot [13] and Brlek [5] as follows. Define sets $\mathcal{A} = \{00, 11, 010010, 101101\}$ and

$$\mathcal{A} = \bigcup_{k \geq 0} \mu^k(\mathcal{A}).$$

The set \mathcal{A} is the set of squares appearing in the Thue–Morse word.

Shelton and Soni [16] characterized the overlap-free squares (the result is also attributed to Thue by Berstel [4]), as being the conjugates of the words in \mathcal{A} . (A *conjugate* of x is a word y such that $x = uv$ and $y = vu$ for some u, v .) Currie and Rampersad [7] showed that the conjugates of the words in \mathcal{A} are also precisely the $7/3$ -power-free squares. Thus, there are only $7/3$ -power-free squares of length $2n$ when n is a power of 2, or 3 times a power of 2. By contrast, we show that there are cubefree binary squares of length $2n$ for every positive integer n . We use this result to construct infinite cubefree binary words containing exponentially many distinct squares.

2 Main results

The main results of this paper are the following two theorems.

Theorem 1 *Let n be a positive integer. There exists a cubefree binary square of length $2n$.*

Theorem 2 *There exists an infinite cubefree binary word containing exponentially many distinct squares of length n .*

We first establish some preliminary results.

Lemma 3 *The Thue–Morse word contains a factor of the form $x = 1001x'' = x'1001$ of every positive even length $n \neq 2, 6$.*

Proof: Aberkane and Currie [1, Lemma 4] proved that for every integer $m \geq 6$, the Thue–Morse word contains a factor of length m of the form $10y10$. Then the Thue–Morse word also contains the factor $\mu(10y10) = 1001\mu(y)1001$, which has length $2m$. Finally, we observe that 10011001 and 1001101001 are factors of the Thue–Morse word of lengths 8 and 10 respectively. \square

Lemma 4 *If y is overlap-free and ayb is a cube of period p , then $p \leq |ab|$.*

Proof: Otherwise deleting a and b removes less than a full period from ayb , leaving an overlap. \square

Lemma 5 *If z is a factor of yyy where $|y| = p$ and $|z| \leq p + 1$, then there are two occurrences of z in yyy .*

Proof: Certainly if z is a factor of yy it occurs twice in yyy . If z is a factor of yyy but not of yy , then z must span the central y of yyy and a bit more on both ends, giving z a length of $p + 2$ or more. \square

Theorem 6 *Let x be a factor of the Thue–Morse word of the form $x = 1001x'' = x'1001$. Then the word $x0x0$ is cubefree.*

Remark 1 Word 01010 occurs exactly once in $x0x0$. (Note that this word is an overlap, and hence not a factor of the Thue–Morse word.)

Proof of Theorem 6: Suppose yyy is a cube in $x0x0$ with $|y| = p > 0$.

Case 1: Period $p \geq 4$.

By Lemma 5 and Remark 1, word 01010 is not a factor of yyy . We have two possibilities:

- (a) Cube yyy is a factor of $x'100101$. This is impossible by Lemma 4, since $x'1001$ is overlap-free, $|01| = 2$, and $p \geq 4 > 2$.
- (b) Cube yyy is a factor of $101001x''0$. This is again impossible by Lemma 4, since $1001x''$ is overlap-free.

Case 2: Period $p \leq 3$.

If 01010 is a factor of yyy , then one of 001010 and 010100 is a factor. However, neither of these has period 1, 2 or 3; this is impossible. We conclude that 01010 is not a factor of yyy . This gives a similar case breakdown as in Case 1.

- (a) Cube yyy is a factor of $x'100101$.
 - (i) Cube yyy is a suffix of $x'100101$. In this case, $p \leq 2$ by Lemma 4, since $x'1001$ is overlap-free. However, the longest suffix of $x'100101$ of period 1 or 2 is 0101, which is cubefree.
 - (ii) Cube yyy is a suffix of $x'10010$. This forces $p = 1$, which is impossible.
- (b) Cube yyy is a factor of $101001x''0$.
 - (i) Cube yyy is a prefix of $101001x''0$ or of $01001x''0$. If x'' is the empty word, then $x0x0 = 1001010010$ is clearly cubefree, so let us assume that $|x''| \geq 4$. Since $|yyy| = 3p \leq 9 \leq |01001x''|$, yyy is a factor of $101001x''$. This is symmetrical to Case 2a.
 - (ii) Cube yyy is a factor of $1001x''0 = x0$. This is impossible by Case 2a. □

Theorem 7 Let x be a factor of the Thue–Morse word of the form $x = 1001x'' = x'1001$. Then the word $x101100x101100$ is cubefree.

Remark 2 Word 00100 occurs exactly once in $x101100x101100$. Word 11011 occurs exactly twice.

Proof of Theorem 7: Suppose yyy is a cube in $x101100x101100$ with $|y| = p > 0$.

Case 1: Period $p \geq 4$.

By Lemma 5 and Remark 2, word 00100 is not a factor of yyy . We have two possibilities:

- (a) Cube yyy is a factor of $x10110010$.
Word $x10110010$ contains 11011 as a factor exactly once. By Lemma 5 and Remark 2, there are two possibilities:
 - (i) Cube yyy is contained in $x101$.
In this case, $p \leq 3$ by Lemma 4, since x is overlap-free. This is a contradiction.

(ii) *Cube yyy is contained in 10110010 .*

This is clearly impossible.

(b) *Cube yyy is a factor of $0x101100$.*

Again, word $0x101100$ contains 11011 as a factor exactly once. Therefore, either yyy is contained in 101100 or in $0x101$. The first alternative evidently is impossible, while the second is ruled out by Lemma 4.

Case 2: *Period $p \leq 3$.*

If 00100 is a factor of yyy , then we must have $p = 3$, since 00100 does not have period 1 or 2. However, in $x101100x101100$, the maximal factor of period 3 containing 00100 is 1001001 , which is not a cube. We conclude that 00100 is not a factor of yyy . This gives a similar case breakdown to Case 1:

(a) *Cube yyy is a factor of $x10110010$.*

By Lemma 4 the word $x10$ must be cubefree. Therefore, yyy must be a suffix of one of these words:

$$\begin{aligned} w_8 &= x'100110110010 \\ w_7 &= x'10011011001 \\ w_6 &= x'1001101100 \\ w_5 &= x'100110110 \\ w_4 &= x'10011011 \\ w_3 &= x'1001101 \end{aligned}$$

None of the w_n ends in a cube of period 1, 2 or 3. (In the case of words w_4, w_3 , the longest suffixes of period 3 have lengths 6 and 5 respectively.) It follows that yyy is not a suffix of any of the w_n , and this case does not occur.

(b) *Cube yyy is a factor of $0x101100$.*

Since $|yyy| = 3p \leq 9 \leq |0x|$, yyy is a factor of $0x$ or of $x101100$. The first possibility was ruled out in the proof of Theorem 6, and the second in Case 2a. \square

Theorems 6 and 7 together establish Theorem 1. Next we show that the number of cubefree binary squares of length n grows exponentially.

Proposition 8 *There exist exponentially many cubefree binary squares of length n .*

Proof: Let m be a positive integer and let xx be a cubefree binary square of length $2m$ over $\{0, 1\}$. Suppose that 0 occurs at least as often as 1 in x . Construct a new cubefree square yy over $\{0, 1, 2\}$, where y is obtained from x by arbitrarily replacing some of the 0's in x by 2's. There are at least $2^{m/2}$ such squares yy of length $2m$.

Let h be the morphism

$$\begin{aligned} 0 &\rightarrow 001011 \\ 1 &\rightarrow 001101 \\ 2 &\rightarrow 011001. \end{aligned}$$

Brandenburg [6, Theorem 6] showed that h maps cubefree words to cubefree words. Moreover, since h is uniform and injective, the set of words $h(yy)$ consists of at least $2^{m/2}$ cubefree squares of length $12m$. Asymptotically, we thus have exponentially many cubefree binary squares of length n , as required. \square

We now prove Theorem 2.

Proof of Theorem 2: In the proof of Proposition 8 we showed that there are at least $2^{m/2}$ cubefree binary squares of length $12m$ for every positive integer m . Let S therefore be any set of cubefree squares over $\{0, 1\}$ where S contains at least $2^{m/2}$ words of length $12m$ for every positive integer m . Let $\mathbf{x} = x_1x_2\cdots$ be any infinite cubefree binary word over $\{2, 3\}$. Construct a word

$$\mathbf{w} = x_1S_1x_2S_2\cdots,$$

where the set of S_i 's is equal to the set S , so that \mathbf{w} is cubefree and contains exponentially many distinct squares of length n . Let g be the morphism

$$\begin{aligned} 0 &\rightarrow 001001101 \\ 1 &\rightarrow 001010011 \\ 2 &\rightarrow 001101011 \\ 3 &\rightarrow 011001011. \end{aligned}$$

Brandenburg [6, Theorem 6] showed that g maps cubefree words to cubefree words. Thus, $g(\mathbf{w})$ is cubefree and, by the uniformity and injectivity of g , contains exponentially many distinct squares of length n . \square

Note that Theorem 2 implies that existence of an infinite cubefree binary word with exponential *factor complexity*—i.e., with exponentially many factors of length n . Similarly, one can easily construct an infinite squarefree word over $\{0, 1, 2\}$ with exponential factor complexity.

Proposition 9 *There exists an infinite squarefree word over $\{0, 1, 2\}$ with exponential factor complexity.*

Proof: Let \mathbf{w} be any infinite squarefree word over $\{0, 1, 2\}$ and let \mathbf{x} be any infinite word over $\{3, 4\}$ with 2^n factors of length n for every positive n . Let \mathbf{y} be the word obtained by forming the *perfect shuffle* of \mathbf{w} and \mathbf{x} : that is, if $\mathbf{w} = w_0w_1w_2\cdots$ and $\mathbf{x} = x_0x_1x_2\cdots$, then define $\mathbf{y} = w_0x_0w_1x_1w_2x_2\cdots$. Clearly, \mathbf{y} is a squarefree word with exponential factor complexity. Let f be the morphism

$$\begin{aligned} 0 &\rightarrow 010201202101210212 \\ 1 &\rightarrow 010201202102010212 \\ 2 &\rightarrow 010201202120121012 \\ 3 &\rightarrow 010201210201021012 \\ 4 &\rightarrow 010201210212021012. \end{aligned}$$

Brandenburg [6, Theorem 4] showed that f maps squarefree words to squarefree words. The uniformity and injectivity of f implies that $f(\mathbf{y})$ is a squarefree word with exponential factor complexity, as required. \square

References

- [1] A. Aberkane, J. Currie, “There exist binary circular $5/2^+$ power free words of every length”, *Electron. J. Combinatorics* **11** (2004), #R10.
- [2] J.-P. Allouche, J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge, 2003.
- [3] J. Berstel, “Mots sans carré et morphismes itérés”, *Discrete Math.* **29** (1980), 235–244.
- [4] J. Berstel, “Axel Thue’s work on repetitions in words”. In P. Leroux, C. Reutenauer, eds., *Séries formelles et combinatoire algébrique*, Publications du LaCIM, pp 65–80, UQAM, 1992.
- [5] S. Brlek, “Enumeration of factors in the Thue–Morse word”, *Discrete Appl. Math.* **24** (1989), 83–96.
- [6] F.-J. Brandenburg, “Uniformly growing k th power-free homomorphisms”, *Theoret. Comput. Sci.* **23** (1983), 69–82.
- [7] J. Currie, N. Rampersad, “Infinite words containing squares at every position”, *Theor. Inform. Appl.* **44** (2010), 113–124.
- [8] J. Currie, N. Rampersad, J. Shallit, “Binary words containing infinitely many overlaps”, *Electron. J. Combinatorics* **13** (2006), #R82.
- [9] F. M. Dekking, “On repetitions of blocks in binary sequences”, *J. Combin. Theory. Ser. A* **20** (1976), 292–299.
- [10] A. Ehrenfeucht, G. Rozenberg, “On the subword complexity of square-free DOL languages”, *Theoret. Comput. Sci.* **16** (1981), 25–32.
- [11] A. Ehrenfeucht, G. Rozenberg, “On the subword complexity of m -free DOL languages”, *Inform. Process. Lett.* **17** (1983), 121–124.
- [12] A. Ehrenfeucht, G. Rozenberg, “On the size of the alphabet and the subword complexity of square-free DOL languages”, *Semigroup Forum* **26** (1983), 215–223.
- [13] J.-J. Pansiot, “The Morse sequence and iterated morphisms”, *Inform. Process. Lett.* **12** (1981), 68–70.
- [14] N. Rampersad, J. Shallit, M.-w. Wang, “Avoiding large squares in infinite binary words”, *Theoret. Comput. Sci.* **339** (2005), 19–34.
- [15] J. Shallit, “Simultaneous avoidance of large squares and fractional powers in infinite binary words”, *Int’l. J. Found. Comput. Sci.* **15** (2004), 317–327.
- [16] R. Shelton, R. Soni, “Chains and fixing blocks in irreducible sequences”, *Discrete Math.* **54** (1985), 93–99.
- [17] A. Thue, “Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen”, *Kra. Vidensk. Selsk. Skrifter. I. Math. Nat. Kl.* **1** (1912), 1–67.