There are binary circular cube-free words of length n contained within the Thue-Morse word for all positive integers n.

D. Sean Fitzpatrick*
Department of Mathematics and Statistics
University of Winnipeg
Winnipeg, Manitoba R3B 2E9, Canada
e-mail: dfitzpat@io.uwinnipeg.ca

March 4, 2003

Abstract

We extend the work of Currie and Fitzpatrick [1] on circular words avoiding patterns by showing that, for any positive integer n, the Thue-Morse word contains a subword of length n which is circular cube-free. This proves a conjecture of V. Linek.

Keywords: Combinatorics on words, cube-free words, Thue-Morse word

1 Introduction

A word such as *ratatat* is said to contain a **cube**, as it can be written as $xyyy$, where $x = r$ and $y = at$; i.e. at occurs three times consecutively.

Early in the twentieth century, it was shown by Thue [6] that it is possible to construct an infinite cube-free word using the alphabet $\{0, 1\}$. In fact, Thue proved a somewhat stronger result, that there was an infinite binary sequence not containing a pattern of the form BBb, where B is a word, and b is the first letter of B. Such a pattern is known as an *overlap*. The sequence used by Thue is known as the Thue-Morse sequence t, which is defined below. It was also shown by Thue that one can construct circular words which are overlap-free, but the only lengths for which this is possible

*The author was supported by an NSERC USRA.

are of the form 2^n and $3 \cdot 2^n$. A recent paper of Currie and Fitzpatrick [1] showed that it is possible to construct circular cube-free words of arbitrary length. Using a related approach, we show that such words exist within the Thue-Morse word:

Main Theorem: *Let n be a natural number. Then there is a binary circular cube-free word of length n contained in t.*

2 Preliminaries

Let u and v be two words. We say u is a **subword** of v, denoted $u \leq v$, or $v \geq u$, if we can write v as $v = xuy$ for some words x and y. It may occur that u is a subword of v in more than one way; that is, if $v = x_1uy_1 = x_2uy_2$, where $x_1 \neq x_2$. In this case we say that there is more than one occurrence of u in v. We say u is a **prefix** of v, denoted $u \leq_p v$, or $v \geq_p u$, if we can write $v = uy$ for some word y. The analogous definition applies when u is a **suffix** of v, denoted $u \leq_s v$ or $v \geq_s u$. If w is a word, let $|w|$ denote its length; that is, the number of letters in w. Thus $|01101001| = 8$, for example. Let w be a word. If one cannot write $w = xyyyz$ with y a non-empty word, then w is said to be **cube-free**. We call a word v a **conjugate** of w if there are words x and y such that $w = xy$ and $v = yx$. If all of the conjugates of w are cube-free, then w is a **circular cube-free word**.

The results below are concerned with binary words; namely, strings over $\{0, 1\}$. If w is a binary word, denote by $\overline{w}$ the binary complement of w, obtained from w by replacing 0's with 1's and vice versa. For example, $\overline{01101001} = 10010110$. Write w as $w = w_1w_2 \dots w_n$, where the w_i are letters. We say that w is **periodic** if for some k we have $w_i = w_{i+k}$, $i = 1, 2, \dots, n-k$. We call k a **period** of w. The **exponent** of w is $|w|/k$, where k is the shortest period of w. For example, the exponent of 01010 is $5/2$. Squares, overlaps and cubes are periodic.

3 The Thue-Morse word: Some useful facts

The Thue-Morse [5, 6] word is the basis of the construction used in the proof of our main theorem. Readers who are unfamiliar with the word have many sources available to them; some examples are [3, 5, 6].

The Thue-Morse sequence t is defined to be $t = h^\omega(0) = \lim_{n\to\infty} h^n(0)$, where $h : \{0, 1\}^* \to \{0, 1\}^*$ is the substitution generated by $h(0) = 01$, $h(1) = 10$. Thus

$$t = 01101001100101101001011001101001\cdots$$

Every subword of t is a subword of $h^n(0)$ for some n. Therefore, every subword of t appears in t infinitely often.

The following are well-known facts regarding the Thue-Morse word which will be useful for our proof:

1. Word t is overlap-free.

2. If w is a subword of t then so is $\overline{w}$. (The set of subwords of t is closed under binary complementation.)

3. If $w = t_0 t_1 \ldots t_n$ is a subword of t, then so is $t_n t_{n-1} \ldots t_1$ (The set of subwords of t is closed under reversal)

4. None of 00100, 01010 or 11011 is a subword of t.

5. If w is a overlap (cube)-free binary word, then $h(w)$ is also overlap (cube)-free.

Lemma 3.1 *Let n be an integer, $n \geq 4$. Word t contains a subword of the form $0\nu00$, where $|0\nu00| = n$.*

Proof: Suppose $0\nu00 \leq t$, $|0\nu00| = n$. Since t is cube-free, we must have $0\mu1001 \leq t$, where $\mu1 = \nu$. Then $h(0\mu1001) = 01h(\mu)10010110 \leq t$, and t contains the subwords $01h(\mu)100$ of length $2n - 3$, and $1h(\mu)1001011$ of length $2n$. Since the set of subwords of t is closed under binary complementation, the existence of $1h(\mu)1001011$ implies that t contains a subword of the form $0\nu00$ of length $2n$. The words 0010 and 001100 are of the desired form and of length 4 and 6, respectively; thus the result is seen to follow by induction. □

Corollary 3.2 *Let n be an integer, $n \geq 4$. Word t contains a subword of the form $00\nu0$, where $|00\nu0| = n$.*

Proof: This follows from the previous lemma, and the fact that the set of subwords of t is closed under reversal. □

By an argument similar to that of Lemma 3.1, one proves the following:

Lemma 3.3 *Let n be an integer, $n \geq 3$. Word t contains a subword of the form $0\nu10$ with $|0\nu10| = n$.*

The following lemma is left as an easy exercise. See [2], for example.

Lemma 3.4 *Let $w = h(\omega)$ for some binary string u. Suppose w contains a cube xxx. Then $|x|$ is even, and ω contains a cube yyy with $|y| = |x|/2$.*

Lemma 3.5 *Let ν be a circular cube-free word of length n. Then $h(\nu)$ is a circular cube-free word of length $2n$.*

Proof: Let ν be a cube-free word, and suppose $v = h(\nu)$ contains a circular cube. Then some conjugate w of v contains a subword xxx. This is equivalent to saying that $xxx \leq vv$, with $|x| \leq |v|/3$, since vv contains all conjugates of v. But $vv = h(\nu)h(\nu) = h(\nu\nu)$; thus by Lemma 3.4, $yyy \leq \nu\nu$ with $|y| = |x|/2 \leq (|v|/3)/2 = (2|\nu|/3)/2 = |\nu|/3$, which implies a circular cube in ν, a contradiction. □

Lemma 3.6 *Suppose that $u \leq xxx$ and $|u| \leq |x|$. Then there is more than one occurrence of u in xxx.*

Proof: Since $u \leq xxx$ and $|u| \leq |x|$, we must have $u \leq xx$. Since there is more than one occurrence of xx in xxx, there is more than one occurrence of u in xxx. □

Lemma 3.7 *Let $v \leq t$. Let $u \in \{0,1\}^*$. Suppose that vu contains a cube. Then for some prefix w of u, word vw ends in a cube xxx with $|x| \leq |w|$.*

Proof: Let w be the shortest prefix of u such that vw contains a cube, possibly $w = u$. It follows that vw ends in xxx for some non-empty word x. We claim that $|x| \leq |w|$; if $|x| > |w|$, write $x = x_0yw$, where x_0 is the first letter of x. Then $t \geq v \geq xxx_0$. an overlap. This is impossible, since t is overlap-free. □

Similarly, one proves the following two lemmas:

Lemma 3.8 *Let $v \leq t$. Let $u \in \{0,1\}^*$. Suppose that uv contains a cube. For some suffix w of u, word wv begins in a cube xxx with $|x| \leq |w|$.*

Lemma 3.9 *Let $v \leq t$. Let $u', u'' \in \{0,1\}^*$. Suppose that $u''vu' = xxx$. Then $|x| \leq |u'| + |u''|$.*

4 Proof of the Theorem

We begin by restating our main theorem, the proof of which depends on a number of technical results which are included after the proof.

Theorem 4.1 *Let n be a natural number. Then there is a binary circular cube-free word of length n contained in t.*

Proof: We prove the existence of circular cube-free words of all lengths within t by considering the lengths falling into the four congruence classes modulo 4. The existence of the words of odd length will be proved below.

The even lengths can then be obtained by applying the morphism h to the words of odd length, using Lemma 3.5.

Suppose $n \equiv 1(mod 4)$, and $n \geq 9$. Then we may write $n = 4k - 3$, for $k \geq 3$. By Lemma 3.3, t contains a subword $0\nu 10$ of length k. Thus t contains the subword $0110h^2(\nu)10010 \leq 0110h^2(\nu)10010110$, where $|0110h^2(\nu)10010| = n$, and $0110h^2(\nu)10010$ is circular cube-free by Theorem 4.4.

Suppose $n \equiv 3(mod 4)$, and $n \geq 15$. Then we may write $n = 4k - 1$, for $k \geq 4$. By Corollary 3.2, t contains a subword $00\nu 0$ of length k. Thus t contains the subword $1100110h^2(\nu)0110 \leq 01100110h^2(\nu)0110$, where $|1100110h^2(\nu)0110| = n$, and $1100110h^2(\nu)0110$ is circular cube-free by Theorem 4.7.

One can easily provide the words of lengths 1, 3, 5, 7 and 11 to complete the proof. □

Lemma 4.2 *Let* $u_a = 010$. *Let* $v = h^2(0\nu)$, *where* $0\nu \leq t$. *Then* $u_a v$ *is cube-free.*

Proof: Word v commences 0110. Also, $v \leq t$. For each non-empty suffix w of u_a, let p_w be the prefix of wv with period at most $|w|$ and maximal exponent. Here are the w and p_w:

w	v	p_w	exponent of p_w
0	0110...	00	2
10	0110...	1	1
010	0110...	01001	5/3

Each p_w has exponent less than 3; thus no wv has a prefix $p = xxx$ where $|x| \leq |w|$. It follows from Lemma 3.9 that $u_a v$ is cube-free.□

The following lemma is proved similarly:

Lemma 4.3 *Let* $v_a = 0$. *Let* $u = h^2(\mu 1)$ *where* $\mu 1 \leq t$. *Then* uv_a *is cube-free.*

Theorem 4.4 *Let* $w = h^2(0\omega 10)$ *where* $0\omega 10 \leq t$. *Let* $g_a = 0$ *and let* $u = h^2(0\omega 1) = 0110h^2(\omega)1001$. *Then word* $ug_a \leq w$ *is circular cube-free.*

Proof: Word $h^2(0\omega 1)$ begins 0110 and ends 1001. Suppose some conjugate of ug_a contains a cube xxx. By Lemma 4.2, $g_a u$ is cube-free, and by Lemma 4.3, ug_a is cube-free.

Thus we must have $xxx = u'' g_a u'$, where $u' \leq_p u$, $u'' \leq_s u$, $u', u'' \neq \epsilon$. Since $0 \leq_p u$ and $001 \leq_s u$, $f_a = 00100 \leq xxx$, and f_a must occur at most once in xxx, since f_a does not occur in t. Thus, $|x| < 5$. The cases where $|x| < 5$ are eliminated by finite checking. We conclude that ug_a is circular cube-free.□

Lemma 4.5 *Let* $u_b = 10110$. *Let* $v = h^2(01\nu)$, *where* $01\nu \leq t$. *Then* $u_b v$ *is cube-free.*

Proof: Word v commences 0110. Also, $v \leq t$. For each non-empty suffix w of u_b, let p_w be the prefix of wv with period at most $|w|$ and maximal exponent. Here are the w and p_w:

w	v	p_w	exponent of p_w
0	01101...	00	2
10	01101...	1001	4/3
110	01101...	11	2
0110	01101...	01100110	2
10110	01101...	10110	5/3

Each p_w has exponent less than 3; thus, no wv has prefix $p = xxx$ where $|x| \leq |w|$. It follows from Lemma 3.9 that $u_b v$ is cube-free.□

The following lemma is proved similarly:

Lemma 4.6 *Let* $v_b = 110$. *Let* $u = h^2(\mu 0)$, *where* $\mu 0 \leq t$. *Then* uv_b *is cube-free.*

Theorem 4.7 *Let* $w = h^2(00\omega 0)$ *where* $00\omega 0 \leq t$. *Let* $g_b = 110$ *and let* $u = h^2(0\omega 0) = 0110h^2(\omega)0110$. *Then word* $g_b u \leq w$ *is circular cube-free.*

Proof: Word $h^2(0\omega 0)$ begins 0110 and ends 0110. Suppose some conjugate of $g_b u$ contains a cube xxx. $10g_b u$ and ug_b are cube-free by Lemmas 4.5 and 4.6, respectively. This leaves the following two cases:
Case 1: Suppose that $xxx = g_b'' u g_b'$, where $g_b = g'gg''$, and g may be the empty word. By Lemma 3.9, $|x| \leq |g''| + |g'|$. Thus, $|x| \leq 3$. But $3|x| = |xxx| = |g''ug'| \geq 1 + 2|0110| + 1 = 10$, which implies that $|x| > 3$, a contradiction.
Case 2: Suppose $xxx = u'' g_b u'$, where $u' \leq_p u$, $u'' \leq_s u$, $u' \neq \epsilon$ and $u'' \geq_s 110$. Since $0 \leq_p u$ and $110 \leq_s u$, $f_b = 11011 \leq xxx$, and f_b must occur at most once in xxx, since f_b does not occur in t. Thus, $|x| \leq 5$. The cases where $|x| \leq 5$ are eliminated by finite checking. We conclude that $g_a u$ is circular cube-free.□

References

[1] James D. Currie & D. Sean Fitzpatrick, Circular Words Avoiding Patterns, *Proceedings of Developments in Language Theory, Sixth International Conference* (2002), 114-121.

[2] James D. Currie, A Direct Proof of a Result of Thue, *Utili tas Mathematica* **25** (1984), 299-302.

[3] Earl D. Fife, Binary sequences which contain no BBb, *Trans. Amer.Math. Soc.* **261** (1980), 115-136; MR **82a**:05034.

[4] M. Lothaire, *Combinatorics on Words,* Encyclopedia of Mathematics and its Applications vol. 17, Addison-Wesley, Reading Mass., 1983.

[5] Marston Morse & Gustav A. Hedlund, Symbolic dynamics I, II, *Amer.J. Math.* **60** (1938), 815–866; **62** (1940) 142; MR **1**, 123d.

[6] Axel Thue, Über unendliche Zeichenreihen, *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl.* Christiana (1906), 1-22.