There are ternary circular square-free words of length n for $n \geq 18$.

James D. Currie*
Department of Mathematics and Statistics
University of Winnipeg
Winnipeg, Manitoba
Canada R3B 2E9
currie@uwpg02.uwinnipeg.ca

Submitted: March 18, 2002 Accepted: October 11, 2002.
MR Subject Classifications: 05B45, 05B30, 11B99

Abstract

There are circular square-free words of length n on three symbols for $n \geq 18$. This proves a conjecture of R. J. Simpson.

Keywords: Combinatorics on words, square-free words

1 Introduction

The word *hotshots* can be written as $(hots)^2$. We thus call *hotshots* a **square**[1]. On the alphabet $\{0, 1\}$ the longest words not containing squares are 010 and 101. On the other hand, at the beginning of the last century, Thue [9] proved that over $\{0, 1, 2\}$ there is an infinite squarefree word, i.e. an infinite sequence not containing any squares.

Variations on the problem of finding squarefree words have included finding infinite squarefree tilings [2], or finding infinite squarefree walks on graphs and digraphs [3, 4]. The problem of finding an infinite squarefree tiling can be viewed as that of finding a coloring of the lattice graph of the tiling on which certain walks give squarefree colour sequences. For example one may ask for colourings of the infinite checkerboard with finitely many colours, such that rook or bishop moves always trace squarefree words [5].

A recent paper of Alon et al. [1] looks for colourings of finite graphs such that all cycle-free walks give squarefree sequences of colours. Let C_n be the cycle on n vertices. They offer the following conjecture.

*This work was supported by an NSERC operating grant.
[1]Thanks to J. Shallit for this interesting natural square.

Conjecture 1.1 For $n \geq 18$, there is a colouring of C_n with 3 colours such that every cycle-free walk gives a square-free word.

The conjecture is several years old, and appears to be due to R. Jamie Simpson. One checks by computer that the result of the conjecture holds for $1 \leq n \leq 179$, with exceptions at $n = 5, 7, 9, 10, 14$ and 17. We establish the conjecture by proving the existence of circular squarefree words for $n \geq 180$.

2 Preliminaries

A word w is **squarefree** if it is impossible to write $w = xyyz$ with y a non-empty word. Word v is a **conjugate** of word w if there are words x and y such that $w = xy$ and $v = yx$. We say that w is a **circular squarefree word** if all of its conjugates are squarefree.

Example 2.1 The set of conjugates of word 123 is $\{123, 231, 312\}$. Each of these is squarefree, so 123 is a circular squarefree word. On the other hand, 1231213 is squarefree, but its conjugate 3123121 starts with the square 312312. Thus 1231213 is not a circular squarefree word.

Our main result is the following:

Main Theorem *For each $n \geq 18$ there is a word over $\{0, 1, 2\}$ of length n which is circular square free.*

If u and v are words, we write $u \leq v$ (equivalently, $v \geq u$) if u is a subword of v, that is, if $v = xuy$ for words x and y. We write $u \leq_p v$ (equivalently, $v \geq_p u$) if u is a prefix of v, that is, if $v = uy$ for some word y. Again, we write $u \leq_s v$ (equivalently, $v \geq_s u$) if u is a suffix of v, that is, if $v = xu$ for some word x.

Remark 2.2 If $uavbw \leq xx$ then $a \leq x$ or $b \leq x$.

Our constructions deal with binary words, i.e. strings over $\{0, 1\}$. If w is a binary word we denote by $\overline{w}$ the binary complement of w, obtained from w by replacing 0's with 1's and vice versa. For example, $\overline{01101001} = 10010110$.

If w is a word, we denote by $|w|$ its length, that is, the number of letters in w. Thus $|01101001| = 8$. If a is a letter, we denote by $|w|_a$ the number of occurrences of a in w. Thus $|01101001|_0 = 4$.

3 A few properties of the Thue-Morse word

The Thue-Morse [8, 9] 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. Let us write

$$t = t_0t_1t_2t_3t_4\cdots$$

where the $t_i \in \{0,1\}$. It follows from the definition that $t_m \neq t_{m+1}$ if m is even.

A ternary square-free sequence s results from counting 1's between subsequent 0's in t:

$$s = 210121020120210\cdots$$

Sequence s has the property that 010 and 212 are not among its subwords. Also, word s inherits from t the property that any subword of s appears in s infinitely often.

Remark 3.1 Suppose that u is a subword of t such that u begins and ends with 0. Counting 1's between subsequent 0's in u gives a subword v of s. We find that $|v| = |u|_0 - 1$.

Lemma 3.2 *Let n be an integer, $n \geq 2$. Word t contains a subword of the form $a\overline{a}vb\overline{b}$ where $a, b \in \{0,1\}$ and $|a\overline{a}v| = n$.*

Proof: If n is even, the subword $t_nt_{n+1}\ldots t_{2n}t_{2n+1}$ will do, since $t_n \neq t_{n+1}$, $t_{2n} \neq t_{2n+1}$. If n is odd, the subword $t_3t_4\ldots t_{n+3}t_{n+4}$ will do. $\Box$

Since the set of subwords of t is closed under binary complementation, we have the following corollary:

Corollary 3.3 *Let n be an integer, $n \geq 2$. Either t contains a subword of the form $10v10$ with $|10v| = n$, or a subword of the form $10v01$ with $|10v| = n$.*

4 Proof of Main Theorem

Remark 4.1 Suppose word $10v_210$ is a subword of t, and consider the word s_2 obtained by counting 1's in $h^3(10v_210) = 1001011001101001h^3(v_2)1001011001101001$. Thus s_2 has the form $0120210u_20120210$, and contains a subword of s of the form $0210u_20120$ of length $4|v_2| + 9$.

Similarly, suppose word $10v_101$ is a subword of t. Consider the word s_1 obtained by counting 1's in $h^3(10v_101) = 1001011001101001h^3(v_1)0110100110010110$. Thus s_1 has the form $0120210u_12102012$, and contains a subword of s the form $0210u_12102$ of length $4|v_1| + 9$.

Combining this remark with Corollary 3.3, we have the following Theorem:

Theorem 4.2 *For every $n \equiv 1 \pmod 4$, $n \geq 9$, either s has a subword of the form $u = 0210w2102$, $|u| = n$ or a subword of the form $u = 0210w0120$, $|u| = n$.*

Claim 4.3 *If $u = 0210w2102$ is a subword of s, then s will contain the subword $012u012$. In fact, the only subword of s of the form vuw with $|v| = |w| = 3$ is $012u012$. In particular, the word $12u01$ is square-free.*

Sketch of Proof: Evidently, $0u$ commences with the square 00, so that $0u$ doesn't appear in s. On the other hand, if the word $1u$ appears in s, then either $21u$ or $01u$ is a subword of s; the first of these starts with the repetition 210210, while the second contains 010, which is not a subword of s. Since neither $0u$ nor $1u$ can appear in s, $2u$ must appear in s.

An easy (but lengthy) continuation of this argument establishes the claim.$\square$

Similarly, one verifies the following:

Claim 4.4 *If $u = 0210w0120$ is a subword of s, then s will have contain the subword $012u210$. In fact, the only subword of s of the form vuw with $|v| = |w| = 3$ is $012u210$. In particular, the word $12u21$ is square-free.*

Remark 4.5 Let ν_0, ν_1, ν_2, ν_3 be the words

$$\nu_0 = 010212010201202101201020120212010212$$

$$\nu_1 = 0102120102012021012010212$$

$$\nu_2 = 010212010201202120121012010212$$

$$\nu_3 = 0102120121012010212.$$

One checks that for each i, $2102\nu_i0210$ is squarefree. Each ν_i contains 010212 exactly twice, but none of the ν_i contains 2102 or 2120210. The shortest suffix of one of the ν_i to contain the word 0210 is suffix 021012010212 of ν_1, of length 12. Whenever word 0210 occurs in one of the ν_i it is in the context 021012010. The longest of the ν_i is ν_0, of length 36.

Here $|\nu_0| = 36 \equiv 0 \pmod 4$, $|\nu_1| = 25 \equiv 1 \pmod 4$, $|\nu_2| = 30 \equiv 2 \pmod 4$, $|\nu_3| = 19 \equiv 3 \pmod 4$.

Theorem 4.6 *Let u be a subword of s of the form $u = 0210w2102$, $|u| \geq 4|\nu_0|$. Then for each i, $u\nu_i$ is a circular squarefree word.*

Proof: Suppose not. We form 4 cases based on which conjugate of uv contains a square:

1. $xx = v_2uv_1$, $v_2 \leq_s \nu_i$, $v_1 \leq_p \nu_i$, $|v_1v_2| \leq |\nu_i|$.

2. $xx = u_2\nu_iu_1$, $u_2 \leq_s u$, $u_1 \leq_p u$, $|u_1u_2| \leq |u|$.

3. $xx = u_2v_1$, $u_2 \leq_s u$, $v_1 \leq_p \nu_i$, $|v_1v_2| \leq |v|$.

4. $xx = v_2u_1$, $v_2 \leq_s \nu_i$, $u_1 \leq_p u$.

Case 1: We have $xx = v_2uv_1$. Since $012u012$ is a subword of s, word $12u01$ is squarefree. We must therefore have one of $|v_1|$,$|v_2| \geq 3$. Also, $|x| \geq |u|/2 \geq 2|\nu_i| \geq 2|v_i|$, $i = 1, 2$. We can thus write $x = v_2u_1 = u_2v_1$, where $u = u_1u_2$, $|u_i| \geq |x|/2 \geq |v_j|$, $i = 1, 2; j = 1, 2$. If $|v_1| \geq 3$ then $v_2u_1 = u_2v_1$ implies that $010 \leq_p v_1 \leq u_1 \leq s$. This is impossible, as 010 is not a subword of s. If $|v_2| \geq 3$ then 212 is a subword of s. Thus $|v_1|$,$|v_2| \leq 2$, which is a contradiction.
Case 2: We have $xx = u_2\nu_iu_1$. However, ν_i, and hence xx, contains subword 010212 exactly twice. By Remark 2.2, 010212 will be a subword of x.

By Remark 4.5 $2102\nu_i0210$ is square-free, so that one of $|u_1|, |u_2| \geq 5$. If $|u_2| \geq 5$, then 2102010212 appears in the first x of xx. Since the second 010212 in xx is a suffix of ν_i, the second occurrence of 2102010212 in xx will be a subword of ν_i, so that ν_i contains 2102, contradicting Remark refnu 0 to 3. We get a similar contradiction if $|u_1| \geq 5$, when $2120210 \leq \nu_i$.
Case 3: We have $xx = u_2v_1$. We may assume that $|v_1| \geq 3$; otherwise xx would be a subword of the square-free word $u01 \leq 12u01$. Similarly, $|u_2| \geq 5$, since $0120\nu_i$ is non-repetiitive. Finally, $|u_2| < |v_1|$. Otherwise, $010 \leq v_1 \leq x \leq u_2 \leq s$, which is impossible.

Now, however, $2102 \leq u_2 \leq x \leq v_1 \leq \nu_i$, which is impossible by Remark 4.5.
Case 4: We have $xx = v_2u_1$. Here $5 \leq |u_1| < |v_2|$. Since $|u_1| \geq 5$, $0210 \leq u_1 \leq x \leq v_2$. Write $x = v_3 = v_4u_1 = v_40210u_2$. By Remark 4.5, $12 \leq |0210v_6| = |x|$. If $|v_4| \geq 3$, then v_3 contains 2120210, contradicting Remark 4.5. On the other hand, if $|v_4| \leq 2$, then $|u_1| = |x| - |v_4| \geq 10$. It follows that some word $0210z$, $|z| \geq 6$ is a subword of u_1,and hence of v_3. The prefix of length 9 of $0210z$ appears in ν_i, and must thus be 021012010. This means that $010 \leq 0210z \leq u_1$, which is impossible. $\square$

Remark 4.7 Let μ_0, μ_1, μ_2, μ_3 be the words

$$\mu_0 = 2120102012021012010210120212012 10212$$

$$\mu_1 = 212010201202101210201 0212$$

$$\mu_2 = 21201020120210120102120121 0212$$

$$\mu_3 = 2120102012102010212.$$

One checks that for each i, $0120\mu_i0210$ is squarefree. Each μ_i contains 212 either two or three times. None of the μ_i contains 2120210 as a subword. The only appearance of 0120212 in one of the μ_i is in the context 01021 0120212 012 in μ_0.

Each μ_i has a prefix 212010, but contains no other 212010. Words μ_1 and μ_3 contain 010212 only as a suffix. The shortest prefix of one of the μ_i to contain 212 twice is μ_3, of length 19. No suffix of μ_i of length 14 or less contains 0210 as a subword. Every suffix of μ_i of length 11 or more contains a word of form $010a212$ or $212a212$ for some a.

Finally, for each i, $|\mu_i| = |\nu_i|$.

Theorem 4.8 *Let u be a subword of s of the form $u = 0210w0120$, $|u| \geq 4|\mu_0|$. Then for each i, $u\mu_i$ is a circular squarefree word.*

Proof: Suppose not. Again we form 4 cases:
Case 1: If $xx = v_2uv_1$ then 212 is a subword of s.
Case 2: We have $xx = u_2\mu_iu_1$.

Here, x must contain 212 exactly once, while one of $|u_1|, |u_2| \geq 5$. If $|u_1| \geq 5$, then 2120210 appears in the second x of xx. The first occurrence of 2120210 in xx will be a subword of μ_i, contradicting Remark 4.7.

If $|u_2| \geq 5$, then 0120212 is a subword of x, and hence of μ_i. The second occurrence of 212 in μ_0 lies in the second x of xx. However, the third 212 of μ_0 will also lie in this second x of xx, contradicting the fact that x contains 212 exactly once.
Case 3: We have $xx = u_2v_1$. We may assume that $5 \leq |u_2| < |v_1|$.

We have $212010 \leq_p v_1$. Write $x = u_2v_3 = v_4$. We must have $|v_3| > 3$; otherwise $u_2 \geq 010$, which is impossible. This means that $212 \leq v_3$, and v_1 contains 212 twice. Thus $|v_1| \geq 19$.

On the other hand, $|v_3| \leq 5$, or else 212010 appears twice in μ_i. This means that $|v_4| \geq 19 - 5 = 14$. This implies $|u_2| = |v_4| - |v_3| \geq 14 - 5 = 9$.

Since $|u_2| \geq 9$, word v_4 contains a subword $z0120212$ where $|z| = 5$.

By Remark 4.7, z is completely determined here; $z = 01021$. Now, however, $010 \leq z \leq u_2$, which is impossible.
Case 4: We have $xx = v_2u_1$. Again $|v_2| > |u_1| \geq 5$.

We have $0210 \leq_p u_1 \leq v_2$. From Remark 4.7 $|v_2| > 14$. Since $|v_2| \geq 11$, word v_2 contains a subword $010a212$ or $212a212$. It follows from Remark 2.2 that x must contain 010 or 212. In particular, x is not a subword of s.

Write $x = v_4 = v_3u_1$. We must have $|v_3| \leq 2$. Otherwise $\mu_i \geq v_4 = v_3u_1 \geq_s 2120210$, contradicting Remark 4.7. Since $|v_3| \leq 2$, $x = v_3u_1 \leq s$. This is a contradiction. □

Theorem 4.9 *For any length $n \geq 180$ there is a circular square-free word of length n over $\{0, 1, 2\}$*

Proof: Let $n \geq 180$ be given. Let r be the least residue of $n-1 \pmod 4$. Let $m = n-|\nu_r|$. Then $m \equiv n-r \equiv n-(n-1) \equiv 1 \pmod 4$. Also, $m \geq n - |\nu_0| \geq 180 - 36 = 144 = 4|\nu_0|$.

By Theorem 4.2, we can find a subword u of s such that $|u| = m$, and of the form $u = 0210w2102$ or the form $u = 0210w0120$. If $u = 0210w2102$, let $W = u\nu_r$. If $u = 0210w0120$, let $W = u\mu_r$. By the preceding two theorems, W will be a circular square-free word, and $|W| = |u| + |\nu_r| = m + |\nu_r| = n - |\nu_r| + |\nu_r| = n$.□

Combining this theorem with a computer search for $n < 180$ gives the **Main Theorem**.

References

[1] Noga Alon, Jaroslaw Grytczuk, Mariusz Haluszczak & Oliver Riordan, Non-repetitive colorings of graphs, Random Structures and Algorithms **21**, Issue 3-4, (2002).

[2] Dwight R. Bean, Andrzej Ehrenfeucht & George McNulty, Avoidable Patterns in Strings of Symbols, Pacific J. Math. **85** (1979), 261–294.

[3] James D. Currie, *Non-repetitive walks in graphs and digraphs*, PhD thesis, University of Calgary (1987).

[4] James D. Currie, Which graphs allow infinite non-repetitive walks? Discrete Math. **87** (1991), 249–260; MR **92a**:05124.

[5] James D. Currie & R. Jamie Simpson, Non-repetitive tilings, Electronic Journal of Combinatorics **9(1)**, R28 (2002).

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

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

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

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