

# Nonoverlap Property of the Thue-Morse Sequence

T.W. CUSICK<sup>a</sup>, P. STĂNICĂ<sup>b</sup>

<sup>a</sup>Department of Mathematics, The State University of New York  
Buffalo, NY 14260-2900, USA; [cusick@buffalo.edu](mailto:cusick@buffalo.edu)

<sup>b</sup>Department of Applied Mathematics, Naval Postgraduate School  
Monterey, CA 93943, USA; [pstanica@nps.edu](mailto:pstanica@nps.edu)

April 20, 2010

## Abstract

In this note, we provide a new proof for the nonoverlap property of the Thue-Morse sequence using a Boolean functions approach and investigate other patterns that occur in a generalization of the Thue-Morse sequence.

## 1 Introduction

The *Thue-Morse (TM) sequence*  $T = (t_n)_{n \geq 0} = 011010011001011010010 \cdots$  is defined as the limit of iterates  $\varphi^n(0)$ , where the map  $\varphi$  is defined by  $\varphi(0) = 01$ ,  $\varphi(1) = 10$ . We denote the  $2^n$ -length initial segment of the TM sequence by  $T_{2^n}$ . It can be seen that the TM sequence can also be generated by setting  $T_1 = 0$  and

$$T_{2^n} = T_{2^{n-1}} \overline{T_{2^{n-1}}}, \quad n \geq 1. \quad (1)$$

or

$$\begin{aligned} T_{2^n} &= T_{2^{n-1}} r(\overline{T_{2^{n-1}}}), \quad \text{for } n \text{ odd.} \\ T_{2^n} &= T_{2^{n-1}} r(T_{2^{n-1}}), \quad \text{for } n \text{ even,} \end{aligned}$$

where  $r(\cdot)$  is the map that reverses the bits of the argument, and  $\overline{B}$  is the complement of  $B$ . Moreover, the TM sequence can also be generated by using the bit expansion of the position, that is, if  $i = \sum_j b_j 2^j$ , then  $t_i = \sum_j b_j \pmod{2}$ . So,  $T = (t_n)_{n \geq 0}$  counts the number of 1's (mod 2) in the base-2 representation of  $n$ . A sequence has the *nonoverlap* property (also known as the *BBb* property) if the subsequence *BBb*, where  $B$  is a block of bits of any  $> 0$  length, and  $b$  is the first bit of  $B$ , does not appear in that sequence. The nonoverlap property was originally proved by Thue in his seminal papers from 1906 and 1912 [11, 12] for the TM sequence, and later reproved in [8] and other places (see

[2, 3, 9, 10] for more on the TM sequence). It is said [2] that the Thue-Morse sequence was the start of what we now call *combinatorics on words*.

Let  $\mathbb{F}_2^n$  be the vector space of dimension  $n$  over the two element field  $\mathbb{F}_2$ . Let us denote the addition operator over  $\mathbb{F}_2$  by  $\oplus$ , and the direct product by “.”. The vectors consisting of all 1, respectively, all 0 (of some length) are denoted by  $\mathbf{1}$ , respectively,  $\mathbf{0}$ . By abuse of notation, when there is no danger of confusion, we sometimes use  $\mathbf{1}, \mathbf{0}$  to denote a binary string consisting of all 1, respectively, all 0. A Boolean function on  $n$  variables may be viewed as a mapping from  $\mathbb{F}_2^n$  into  $\mathbb{F}_2$ . We order  $\mathbb{F}_2^n$  lexicographically, and denote  $\mathbf{v}_0 = (0, \dots, 0, 0)$ ,  $\mathbf{v}_1 = (0, \dots, 0, 1)$ ,  $\mathbf{v}_{2^n-1} = (1, \dots, 1, 1)$ . We interpret a Boolean function  $f(x_1, \dots, x_n)$  as the output column of its *truth table*, i.e., a binary string of length  $2^n$ ,  $f = [f(\mathbf{v}_0), f(\mathbf{v}_1), f(\mathbf{v}_2), \dots, f(\mathbf{v}_{2^n-1})]$ .

Let  $\epsilon := \epsilon_1 \epsilon_2 \dots$  be a sequence of  $\epsilon_i \in \{0, 1\}$  bits (possibly infinite). Define a function  $r_{\epsilon_i}$  on arbitrary bit-blocks  $B$ , in the following way:

$$r_{\epsilon_i}(B) = \begin{cases} B & \text{if } \epsilon_i = 0 \\ \bar{B} & \text{if } \epsilon_i = 1. \end{cases} \quad (2)$$

In [4] we introduced the *generalized Thue-Morse* sequence  $T^\epsilon = (t_n^\epsilon)_{n \geq 0}$  (we called it the  $\epsilon$ -TM sequence) by the following algorithm ( $T_{2^i}^\epsilon$  is the binary string made up of the first  $2^i$  bits of  $T^\epsilon$ ):

$$\begin{aligned} T_1^\epsilon &= t_0 \in \{0, 1\} \\ T_{2^i}^\epsilon &= T_{2^{i-1}}^\epsilon r_{\epsilon_i}(T_{2^{i-1}}^\epsilon) \end{aligned} \quad (3)$$

The classical Thue-Morse sequence is  $T^\epsilon$ , where  $\epsilon = 11\dots$ . Since [4] was published, we learned that Keane [7] also studied this generalization.

In [4] we proved

**Theorem 1.** *The initial segment of length  $2^n$ ,  $n \geq 2$ , of the TM sequence is the truth table of the Boolean function*

$$f(x_1, x_2, \dots, x_n) = x_1 \oplus x_2 \oplus \dots \oplus x_n,$$

defined on  $\mathbb{F}_2^n$  (ordered lexicographically). Moreover, given an initial segment  $T_{2^n}$  of length  $2^n$  of a generalized Thue-Morse sequence, there exists an affine Boolean function  $f$  (if  $t_0 = 0$ , then  $f$  is linear) on  $n$  variables, such that  $T_{2^n}$  is the truth table of  $f$ .

We also define the following set  $\mathcal{B}$  of 4-bit strings:

$$\begin{aligned} \mathcal{B} = \{ & A = 0, 0, 1, 1; \bar{A} = 1, 1, 0, 0; B = 0, 1, 0, 1; \bar{B} = 1, 0, 1, 0; \\ & C = 0, 1, 1, 0; \bar{C} = 1, 0, 0, 1; D = 0, 0, 0, 0; \bar{D} = 1, 1, 1, 1\}. \end{aligned} \quad (4)$$

As a consequence of Theorem 1 we have the next corollary.

**Corollary 2.** *The Thue-Morse sequence can be written as*

$$T = C\bar{C}\bar{C}C\bar{C}C\bar{C}\dots \quad (5)$$

In [4], we showed that if  $\epsilon \neq 1$ , then the  $\epsilon$ -TM sequence does not have the nonoverlap property, and we raised the question of investigating the occurrence of other patterns in this generalization. In this short note, we will provide yet another proof (arguably the simplest known proof; we use a Boolean functions approach) of Thue's nonoverlap property, and find other patterns in the  $\epsilon$ -TM sequence.

## 2 The nonoverlap property

**Theorem 3.** *The Thue-Morse sequence satisfies the nonoverlap property.*

*Proof.* Assume that the TM sequence  $T$  does not satisfy the nonoverlap property, and so there exist blocks  $B$  (of length  $> 0$ ) such that  $BBb$  occurs in  $T$ . Take  $n$  to be the smallest integer such that  $T_{2^n}$  contains such a pattern  $BBb$ . We assume that  $n \geq 8$ , since for  $n < 8$ , one can check easily that there is no occurrence of the pattern  $BBb$ . Write  $T_{2^n} = T_{2^{n-1}}\overline{T_{2^{n-1}}}$ . If there exists  $B$  such that  $BBb$  occurs in the second half of  $T_{2^n}$ , then, Theorem 1 or (1) implies that  $\overline{B}\overline{B}\overline{b}$  must occur in the first half, and so, there exists an overlap pattern occurring in  $T_{2^{n-1}}$ , which contradicts the minimality of  $n$ . Further,  $BBb$  cannot occur in the first half of  $T_{2^n}$ , since  $n$  is minimal. Therefore,  $BBb$  must intersect the first and second halves of  $T_{2^n}$ , as pictured in Figure 1 ( $B = B_1B_2$ , where either  $B_i$  could be empty), where the pattern  $BBb$  is shown in the center of  $T_{2^n}$  and is split by the “dividing line” between  $T_{2^{n-1}}$  on the left and  $\overline{T_{2^{n-1}}}$  on the right. If this central occurrence of  $BBb$  does not extend beyond the blocks  $T_{2^{n-3}}$  and  $\overline{T_{2^{n-3}}}$  on either side of the dividing line, then Theorem 1 or (1) implies that the pattern  $BBb$  also occurs in the leftmost block  $T_{2^{n-2}}$  inside  $T_{2^n}$  (see Figure 1), which contradicts the minimality of  $n$ .

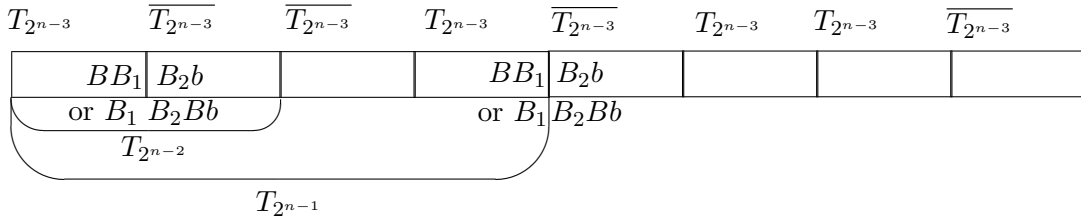


Figure 1: Assuming that an overlap occurs in  $T_{2^n}$

Now, we consider the case when the pattern  $BBb$  extends beyond the blocks  $T_{2^{n-3}}$  and  $\overline{T_{2^{n-3}}}$ . Let the length of  $B$  be denoted by  $r$ , and assume that such an  $r$  is minimal for any pattern  $BBb$ . Let  $f$  be the Boolean function generating  $T_{2^n}$ , and let the initial bit of  $B$  be  $b = f(\mathbf{v}_i)$ .

If  $r$  is odd, then, since  $f(\mathbf{v}_i) = f(\mathbf{v}_{i+r}) = f(\mathbf{v}_{i+2r})$ , and since the length of the block  $B$  is  $r \geq 2^{n-4} \geq 2^4$ , then, there must be two blocks  $C$  (or  $\overline{C}$ ) at distance exactly  $r$  apart in the two blocks  $B$ . This means that the middle 2-bit blocks  $S = 11$  (or  $S = 00$ ) must have an odd distance ( $r - 2$  bits) between them, which is impossible, since we know that if  $t_i = t_{i+1}$ , then  $i$  must be odd. Thus assuming  $r$  odd gives a contradiction.

If  $r$  is even, say  $r = 2s$ , we consider two cases depending on the parity of  $i$ . If  $i = 2j$ , then  $f(\mathbf{v}_{2j}) = f(\mathbf{v}_{2j+2s}) = f(\mathbf{v}_{2j+4s})$ , which implies that  $f(\mathbf{v}_j) = f(\mathbf{v}_{j+s}) = f(\mathbf{v}_{j+2s})$ . Moreover, replacing  $k$  by  $2k_0$  in  $f(\mathbf{v}_i) = f(\mathbf{v}_{i+r+k})$  ( $0 \leq k \leq r$ ,  $k$  even), we obtain  $f(\mathbf{v}_{2j}) = f(\mathbf{v}_{2j+2s+2k_0})$  ( $0 \leq k_0 \leq s$ ), and so  $f(\mathbf{v}_j) = f(\mathbf{v}_{j+s+k_0})$  ( $0 \leq k_0 \leq s$ ). This implies the existence of a pattern  $BBb$  with  $B$  of size  $s < r$ , contradicting the minimality of  $r$ . If  $i = 2j + 1$ , then  $f(\mathbf{v}_{2j+1}) = f(\mathbf{v}_{2j+2s+1}) = f(\mathbf{v}_{2j+4s+1})$ , which implies  $f(\mathbf{v}_j) \oplus 1 = f(\mathbf{v}_{j+s}) \oplus 1 = f(\mathbf{v}_{j+2s}) \oplus 1$ , and so,  $f(\mathbf{v}_j) = f(\mathbf{v}_{j+s}) = f(\mathbf{v}_{j+2s})$ . Further, as before, replacing  $k$  by  $2k_0 - 1$  in  $f(\mathbf{v}_i) = f(\mathbf{v}_{i+r+k})$  ( $0 \leq k \leq r$ ,  $k$  odd), we obtain  $f(\mathbf{v}_{2j+2k_0}) = f(\mathbf{v}_{2j+2s+2k_0})$  ( $0 \leq k_0 \leq s$ ), which is equivalent to  $f(\mathbf{v}_{j+k_0}) = f(\mathbf{v}_{j+s+k_0})$  ( $0 \leq k_0 \leq s$ ), which again gives a pattern  $BBb$  that contradicts the minimality of  $r$ .  $\square$

Next, we consider the generalized Thue-Morse sequence. For simplicity, we label the patterns:  $\alpha = BBb$ ,  $\beta = BB\bar{b}$ ,  $\gamma = B\bar{B}b$ ,  $\delta = B\bar{B}\bar{b}$ . Using our method from [4] we show the next result.

**Theorem 4.** *If  $\epsilon = \mathbf{1}$ , the Thue-Morse sequence avoids  $\alpha$  and cannot avoid  $\beta, \gamma, \delta$ . If  $\epsilon = \mathbf{0}$ , the Thue-Morse sequence avoids  $\beta, \gamma, \delta$  and cannot avoid  $\alpha$ . If  $\epsilon \neq \mathbf{1}, \mathbf{0}$ , then the  $\epsilon$ -TM sequence cannot avoid any of the patterns  $\alpha, \beta, \gamma, \delta$ .*

*Proof.* If  $\epsilon = \mathbf{1}$ , by Theorem 3, we know that  $\alpha$  is avoided. From (5), we see that  $C\bar{C}\bar{C}$ ,  $C\bar{C}C$ ,  $C\bar{C}\bar{C}$  occur in  $T$ , and so,  $\beta, \gamma, \delta$  are not avoided in  $T$ . If  $\epsilon = \mathbf{0}$ , then the  $\epsilon$ -TM sequence is simply  $0000\dots$ , and so, the second claim is true.

If  $\epsilon \neq \mathbf{0}, \mathbf{1}$ , we showed in [4] that the  $\epsilon$ -TM sequence does not have the nonoverlap property, and so,  $\alpha = BBb$  must occur in  $T$ . Further, since  $\epsilon \neq \mathbf{1}, \mathbf{0}$ , then  $\epsilon$  must contain at least one of the patterns  $010, 0110, 0111$ . We shall consider these cases separately. For easy writing, we let  $B_0 = \{t_0\}$  and  $B_i := T_{2^i}^\epsilon$ ,  $i \geq 1$ .

*Case (i)* Let  $\epsilon_i \epsilon_{i+1} \epsilon_{i+2} = 010$ , for some  $i \geq 1$ . Then

$$\begin{aligned} B_{i+2} &= B_{i+1} r_{\epsilon_{i+2}}(B_{i+1}) = B_{i+1} B_{i+1} \\ &= B_i r_{\epsilon_{i+1}}(B_i) B_i r_{\epsilon_{i+1}}(B_i) = B_i \bar{B}_i \bar{B}_i \bar{B}_i \\ &= B_{i-1} r_{\epsilon_i}(B_{i-1}) \overline{B_{i-1} r_{\epsilon_i}(B_{i-1})} B_{i-1} r_{\epsilon_i}(B_{i-1}) \overline{B_{i-1} r_{\epsilon_i}(B_{i-1})} \\ &= B_{i-1} B_{i-1} \overline{B_{i-1} B_{i-1}} B_{i-1} B_{i-1} \overline{B_{i-1} B_{i-1}}, \end{aligned}$$

which contains  $B_{i-1} B_{i-1} \overline{B_{i-1} B_{i-1}}$ ,  $B_i \bar{B}_i B_i$ ,  $B_{i-1} \overline{B_{i-1} B_{i-1}}$ , and so,  $\beta, \gamma, \delta$  are not avoided in  $T$ .

*Case (ii)* Let  $\epsilon_i \epsilon_{i+1} \epsilon_{i+2} \epsilon_{i+3} = 0110$ , for some  $i \geq 1$ . Then

$$\begin{aligned} B_{i+3} &= B_{i+2} B_{i+2} = B_{i+1} \overline{B_{i+1} B_{i+1}} B_{i+1} \overline{B_{i+1} B_{i+1}} \\ &= B_i \bar{B}_i \bar{B}_i B_i B_i \bar{B}_i \bar{B}_i B_i, \end{aligned}$$

which contains  $B_i B_i \bar{B}_i$ ,  $B_i \bar{B}_i B_i$ ,  $B_i \bar{B}_i \bar{B}_i$ , and so,  $\beta, \gamma, \delta$  are not avoided in  $T$ .

*Case (iii)* Let  $\epsilon_i \epsilon_{i+1} \epsilon_{i+2} \epsilon_{i+3} = 0111$ , for some  $i \geq 1$ . Then

$$\begin{aligned} B_{i+3} &= B_{i+2} \overline{B_{i+2} B_{i+2}} = B_{i+1} \overline{B_{i+1} B_{i+1}} \overline{B_{i+1} B_{i+1}} B_{i+1} \\ &= B_i \bar{B}_i \bar{B}_i B_i \bar{B}_i B_i \bar{B}_i B_i, \end{aligned}$$

which contains  $B_i B_i \bar{B}_i$ ,  $B_i \bar{B}_i B_i$ ,  $B_i \bar{B}_i \bar{B}_i$ , which implies that  $\beta, \gamma, \delta$  are not avoided in  $T$ , and the theorem is proved.  $\square$

## References

- [1] J.-P. Allouche, A. Arnold, J. Berstel, S. Brlek, W. Jockusch, S. Plouffe, B.E. Sagan, A relative of the Thue-Morse sequence, in *Formal power series and algebraic combinatorics* (Montreal, PQ, 1992), *Discrete Math.* 139 (1995), 455–461.
- [2] J.-P. Allouche, J. Shallit, The ubiquitous Prouhet-Thue-Morse sequence, In C. Ding, T. Helleseth, and H. Niederreiter, editors, *Sequences and Their Applications, Proceedings of SETA' 98* (1999), Springer-Verlag, 1–16.
- [3] J.-P. Allouche, J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge University Press, 387-391, 2003.
- [4] T.W. Cusick, H. Fredricksen, P. Stănică, On the delta sequence of the Thue-Morse sequence, *Australasian J. Combinatorics* 39 (2007), 293–300.
- [5] H. Fredricksen, Gray codes and the Thue-Morse-Hedlund sequence, *J. Combin. Math. Combin. Comput.* 11 (1992), 3–11.
- [6] G.A. Hedlund, Remarks on the work of Axel Thue on sequences, *Nordisk Mat. Tidskr.* 15 (1967), 148–150.
- [7] M. Keane, Generalized Morse sequences, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 10 (1968), 335–353.
- [8] M. Morse, A one-to-one representation of geodesics on a surface of negative curvature, *Amer. J. Math.* 43 (1921), 35–51.
- [9] A.M. Shur, Binary words avoided by the Thue-Morse sequence, *Semigroup Forum* 53 (1996), 212–219.
- [10] A. M. Shur, Overlap-free words and Thue-Morse sequences, *International Journal of Algebra and Computation* 6 (1996), 353–367.
- [11] A. Thue, Über unendliche Zeichenreihen, *Norske Vid Selsk. Skr. I. Mat. Nat. Kl. Christiania* 7 (1906), 1–22. Reprinted in “Selected mathematical papers of Axel Thue”, T. Nagell ed. Universitetsforlaget, Oslo, 139–158, 1977.
- [12] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Norske Vid Selsk. Skr. I. Mat. Nat. Kl. Christiania* 7 (1912), 1–67. Reprinted in “Selected mathematical papers of Axel Thue”, T. Nagell ed. Universitetsforlaget, Oslo, 413–478, 1977.