Journal of Integer Sequences, Vol. 17 (2014),
Article 14.10.2

The Least Self-Shuffle of the Thue-Morse Sequence

James D. Currie[1]
Department of Mathematics & Statistics
University of Winnipeg
Winnipeg, MB R3B 2E9
Canada
j.currie@uwinnipeg.ca

Abstract

We show that the self-shuffle of Thue-Morse given by Charlier et al. is optimal/canonical in the sense that among self-shuffles of Thue-Morse, it has the lexicographically least directive sequence starting with 1.

1 Introduction

Henshall et al. [3] initiated the topic of self-shuffles of finite words. They considered, in particular, closure properties of languages under self-shuffles, proving several results as well as posing open problems.

No non-empty finite word can be equal to one of its self-shuffles, but for infinite words, the question of whether a word can be written as a self-shuffle is interesting. Charlier et al. [1] exhibited a self-shuffle of the Thue-Morse word. The Thue-Morse word is the fixed point of a morphism, so that we can immediately get other shuffles; the image of any self-shuffle under the morphism gives a different self-shuffle. Endrullis and Hendriks [2] proved that there are in fact other self-shuffles; in particular, they showed that a shuffle distinct from that of Charlier et al. is *optimal* — it switches back and forth between shuffled copies as quickly as possible. The Thue-Morse word thus allows at least two distinct families of self-shuffles.

[1]The author is supported by an NSERC Discovery grant.

In this note, we show that the self-shuffle of Thue-Morse given by Charlier et al. is optimal/canonical in a different sense: among self-shuffles of Thue-Morse, it has the lexicographically least directive sequence starting with 1.

2 Notation

We follow Lothaire [4] as a standard notational reference for combinatorics on words. Thus $|x|$ is the length of word x, $|x|_0$ the number of 0's in x, etc. If x is a non-empty word, let x' denote the word obtained by deleting the last letter of x. Thus, $(12341234)' = 1234123$, for example. Let u, v, w be finite words, and let d be a word over $\{0, 1\}$ such that $|w| = |d| = |u| + |v|$. We define recursively what it means for w to be *the shuffle of u and v directed by d*, written $w = u \oplus_d v$:

1. If $d = \epsilon$, then $w = u \oplus_d v$
2. If the last letter of d is 0 then $w = u \oplus_d v$ if
   (a) $w' = u' \oplus_{d'} v$
   (b) The last letter of w is the same as the last letter of u
3. If the last letter of d is 1 then $w = u \oplus_d v$ if
   (a) $w' = u \oplus_{d'} v'$
   (b) The last letter of w is the same as the last letter of v

In other words, each letter of w is read from either u or v, and d determines whether we read it from u (0) or from v (1). We call d the *directive word* of the shuffle.

By ω-word we mean a 1-sided infinite word. For ω-words $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, $\mathbf{d}$, we extend the definition above and write $\mathbf{w} = \mathbf{u} \oplus_{\mathbf{d}} \mathbf{v}$ if there are arbitrarily long prefixes $\hat{u}$, $\hat{v}$, $\hat{w}$, $\hat{d}$ of $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, $\mathbf{d}$, respectively, such that $\hat{u} \oplus_{\hat{d}} \hat{v} = \hat{w}$.

Remark 1. Suppose that $d_0 \in \{0, 1\}^*$ is a finite prefix of $\mathbf{d}$ and write $\mathbf{d} = d_0\mathbf{d_1}$.

- Let w_0 be the prefix of $\mathbf{w}$ of length $|d_0|$ and write $\mathbf{w} = w_0\mathbf{w_1}$.
- Let u_0 be the prefix of $\mathbf{u}$ of length $|d_0|_0$ and write $\mathbf{u} = u_0\mathbf{u_1}$.
- Let v_0 be the prefix of $\mathbf{v}$ of length $|d_0|_1$ and write $\mathbf{v} = v_0\mathbf{v_1}$.

Then

$$\mathbf{w} = \mathbf{u} \oplus_{\mathbf{d}} \mathbf{v} \Leftrightarrow (w_0 = u_0 \oplus_{d_0} v_0 \text{ and } \mathbf{w_1} = \mathbf{u_1} \oplus_{\mathbf{d_1}} \mathbf{v_1})$$

We say that an ω-word $\mathbf{w}$ *allows a non-trivial self-shuffle* if we can write $\mathbf{w} = \mathbf{w} \oplus_{\mathbf{d}} \mathbf{w}$ for some non-constant ω-word $\mathbf{d}$. Evidently, for any ω-word $\mathbf{w}$, $\mathbf{w} = \mathbf{w} \oplus_{0^\omega} \mathbf{w} = \mathbf{w} \oplus_{1^\omega} \mathbf{w}$; we call these the *trivial self-shuffles* of $\mathbf{w}$. Write $x \preceq y$ (resp., $x \prec y$) to say that word x is no greater than (resp., less than) y in the natural lexicographic order where 0 precedes 1. Because we have the trivial self-shuffles, the lexicographically least ω-word $\mathbf{d}$ such that $\mathbf{w} = \mathbf{w} \oplus_{\mathbf{d}} \mathbf{w}$ is just $\mathbf{d} = 0^\omega$. Seeking the lexicographically least directive sequence *starting with 1* is a reasonable attempt to force non-trivial shuffling. Thus, if ω-word $\mathbf{w}$ allows a non-trivial self-shuffle, a natural question is

> What is the lexicographically least ω-word $\mathbf{d}$ *with prefix 1* such that $\mathbf{w} = \mathbf{w} \oplus_{\mathbf{d}} \mathbf{w}$?

3 Lexicographically least shuffles

In this section, $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ will be arbitrary but fixed effectively given ω-words.

Lemma 2. *Let a word $d_0 \in \{0,1\}^*$ be specified. Let*

$$D = \{\mathbf{d} \in \{0,1\}^\omega : \mathbf{w} = \mathbf{u} \oplus_{\mathbf{d}} \mathbf{v}\}.$$

If $D \cap d_0\{0,1\}^\omega$ is non-empty, then it has a lexicographically least element.

Proof. For a positive integer n, suppose that d_{n-1} has been defined and $D \cap d_{n-1}\{0,1\}^\omega$ is non-empty. It follows that at least one of $D \cap d_{n-1}0\{0,1\}^\omega$ and $D \cap d_{n-1}1\{0,1\}^\omega$ is non-empty. We can thus define an infinite sequence of words $\{d_n\}_{n=0}^\infty$, each d_n an extension of d_{n-1}, by

$$d_n = \begin{cases} d_{n-1}0, & \text{if } D \cap d_{n-1}0\{0,1\}^\omega \text{ is non-empty;} \\ d_{n-1}1, & \text{otherwise.} \end{cases}$$

Let $\bar{\mathbf{d}} = \lim_{n\to\infty} d_n$. We claim that $\bar{\mathbf{d}}$ is the lexicographically least element of $D \cap d_0\{0,1\}^\omega$. Each finite prefix d_n of $\bar{\mathbf{d}}$ has been chosen to be the prefix of a word of D, so that $\mathbf{w} = \mathbf{u} \oplus_{\bar{\mathbf{d}}} \mathbf{v}$. On the other hand, if for some $\hat{\mathbf{d}} \in D \cap d_0\{0,1\}^\omega$, $\hat{\mathbf{d}} \prec \bar{\mathbf{d}}$, consider the shortest prefix p of $\hat{\mathbf{d}}$ which is not a prefix of $\bar{d}$. For some positive n, $p = d_{n-1}0$, while $d_n = d_{n-1}1$. However, this implies that $D \cap d_{n-1}0\{0,1\}^\omega$ is empty, and $\hat{\mathbf{d}} \notin D \cap d_0\{0,1\}^\omega$. This is a contradiction. ☐

Remark 3. Suppose that

$$\mathbf{w} = \mathbf{u} \oplus_{\mathbf{d}} \mathbf{v}$$

has solutions $\mathbf{d} \in 1\{0,1\}^*$. For a fixed prefix w_0 of $\mathbf{w}$, we can effectively determine the lexicographically least element d_0 of $1\{0,1\}^*$ such that there exist prefixes u_0 and v_0 of $\mathbf{u}$ and $\mathbf{v}$, respectively, such that

$$w_0 = u_0 \oplus_{d_0} v_0. \tag{1}$$

There are only $2^{|w_0|-1}$ candidates for d_0. We can check for each candidate d_0, and the corresponding prefixes u_0, v_0 of $\mathbf{u}$, $\mathbf{v}$, with lengths $|d_0|_0$, $|d_0|_1$, whether (1) is satisfied. Note that the lengths of prefixes u_0, v_0 are always at most $|w_0|$.

Lemma 4. *Suppose that* $\mathbf{d}$ *is the lexicographically least element of* $1\{0,1\}^\omega$ *such that*

$$\mathbf{w} = \mathbf{u} \oplus_{\mathbf{d}} \mathbf{v}.$$

Let w_0 *be a fixed non-empty prefix of* $\mathbf{w}$. *Let* d_0 *be the lexicographically least element of* $1\{0,1\}^*$ *such that there exist prefixes* u_0 *and* v_0 *of* $\mathbf{u}$ *and* $\mathbf{v}$, *respectively, such that*

$$w_0 = u_0 \oplus_{d_0} v_0.$$

Suppose $d_0 \in \{0,1\}^*1$; *write* $\mathbf{w} = w_0'\mathbf{W}$, $\mathbf{u} = u_0\mathbf{U}$, $\mathbf{v} = v_0'\mathbf{V}$ *(so that* $w_0' = u_0 \oplus_{d_0'} v_0'$*). Suppose that there exists an element* $\delta \in 1\{0,1\}^\omega$ *such that*

$$\mathbf{W} = \mathbf{U} \oplus_\delta \mathbf{V}.$$

Then

$$\mathbf{d} = d_0'\mathbf{\Delta},$$

where $\mathbf{\Delta}$ *is the lexicographically least such* δ. *In particular,* d_0 *is a prefix of* $\mathbf{d}$.

Proof. Since $w_0' = u_0 \oplus_{d_0'} v_0'$ and $\mathbf{W} = \mathbf{U} \oplus_\delta \mathbf{V}$, by Remark 1, we have

$$\mathbf{w} = \mathbf{u} \oplus_{d_0'\Delta} \mathbf{v}.$$

Let d be the length $|w_0|$ prefix of $\mathbf{d}$. By the minimality of $\mathbf{d}$, $d \preceq d_0'1$.

Since

$$\mathbf{w} = \mathbf{u} \oplus_{\mathbf{d}} \mathbf{v},$$

it follows from Remark 1 that

$$w_0 = \hat{u}_0 \oplus_d \hat{v}_0,$$

where $\hat{u}_0$ is the length $|d|_0$ prefix of $\mathbf{u}$, and $\hat{v}_0$ is the length $|d|_1$ prefix of $\mathbf{v}$. By the lexicographic minimality of d_0, $d_0'1 = d_0 \preceq d$, so that $d_0 = d$.

Therefore, write $\mathbf{d} = d_0'\hat{\Delta}$, where $\hat{\Delta} \in 1\{0,1\}^\omega$. By Remark 1,

$$\mathbf{W} = \mathbf{U} \oplus_{\hat{\Delta}} \mathbf{V}.$$

By the minimality of Δ, $\Delta \preceq \hat{\Delta}$. However, by the minimality of $\mathbf{d}$, $d_0'\hat{\Delta} = \mathbf{d} \preceq d_0'\Delta$. Thus $\Delta = \hat{\Delta}$, so that $\mathbf{d} = d_0'\mathbf{\Delta}$. □

Corollary 5. *Let* $\mathbf{d}$ *be the lexicographically least element of* $1\{0,1\}^\omega$ *such that*

$$\mathbf{w} = \mathbf{u} \oplus_{\mathbf{d}} \mathbf{v}.$$

Suppose that for each positive integer i *there are finite words* W_i, U_i *and* V_i, *and* ω*-words* $\mathbf{w}_i$, $\mathbf{u}_i$ *and* $\mathbf{v}_i$, *where*

- $\mathbf{w}_1 = \mathbf{w}, \mathbf{u}_1 = \mathbf{u}$ *and* $\mathbf{v}_1 = \mathbf{v}$,
- W_i, U_i, V_i *are prefixes of length 2 or more of* $\mathbf{w}_i, \mathbf{u}_i, \mathbf{v}_i$, *respectively,*
- $\mathbf{w}_{i+1} = (W_i')^{-1}\mathbf{w}_i, \mathbf{u}_{i+1} = (U_i')^{-1}\mathbf{u}_i, \mathbf{v}_{i+1} = (V_i')^{-1}\mathbf{v}_i$.

so that, for each i,

$$\begin{aligned} \mathbf{w}_i &= \prod_{j=i}^{\infty} W_j' \\ \mathbf{u}_i &= \prod_{j=i}^{\infty} U_j' \\ \mathbf{v}_i &= \prod_{j=i}^{\infty} V_j'. \end{aligned}$$

For each i, *let* D_i *be the lexicographically least word starting with 1 such that*

$$W_i = \hat{u}_i \oplus_{D_i} \hat{v}_i$$

for some prefixes $\hat{u}_i$ *of* $\mathbf{u}_i$ *and* $\hat{v}_i$ *of* $\mathbf{v}_i$. *Suppose that, for each* i, D_i *ends in a 1,* $\hat{u}_i = U_i$ *and* $\hat{v}_i = V_i$. *Then*

$$\mathbf{d} = \prod_{i=1}^{\infty} D_i'.$$

Proof. This follows from the previous lemma by induction. □

4 The Thue-Morse word

Consider the binary version of the Thue-Morse word (A001285), namely, $\mathbf{t} = \mu^{\omega}(0)$ where $\mu(0) = 01$, $\mu(1) = 10$. Thus

$$\mathbf{t} = 0110100110010110 \cdots$$

The length 2 factors of the Thue-Morse word are 00, 01, 10, 11. If $\mathbf{t}[j..j+1] = ab$, a, $b \in \{0, 1\}$, then

$$\mathbf{t}[8j..8j+15] = \mu^3(ab)$$

and

$$\mathbf{t}[16j..16j+31] = \mu^4(ab).$$

It follows that

$$\langle \mathbf{t}[8j+1..8j+8], \mathbf{t}[8j+5..8j+13], \mathbf{t}[16j+6..16j+22] \rangle$$

takes on one of 4 possible values:

If $\mathbf{t}[j..j+1] = 00$, then

$$\begin{aligned} \mathbf{t}[8j..8j+15] &= 0\underline{1101\overline{0010}}\overline{11010}01 \\ \mathbf{t}[16j16j+31] &= 011010\underline{01100101100110100}110010110 \end{aligned}$$

so that

$$\begin{aligned} &\langle \mathbf{t}[8j+1..8j+8], \mathbf{t}[8j+5..8j+13], \mathbf{t}[16j+6..16j+22]\rangle \\ &\quad = \langle 11010010, 001011010, 01100101100110100\rangle. \end{aligned}$$

Arguing similarly in the other three cases, we find that

$$\langle \mathbf{t}[8j+1..8j+8], \mathbf{t}[8j+5..8j+13], \mathbf{t}[16j+6..16j+22]\rangle \in \langle U_i, V_i, W_i\rangle$$

where the values of the U_i, V_i, W_i are as follows:

i	U_i	V_i	W_i
1	11010010	001011010	01100101100110100
2	11010011	001100101	01100101101001011
3	00101100	110011010	10011010010110100
4	00101101	110100101	10011010011001011

For each non-negative integer j, let $i_j \in \{1, 2, 3, 4\}$ be the unique value such that

$$\mathbf{t}[8j+1..8j+8] = U_{i_j}.$$

Let $D_1 = 10001110100011101$, $D_2 = 10001001100111101$.
One checks that

$$\begin{aligned} W_1 &= U_1 \oplus_{D_1} V_1 \\ W_2 &= U_2 \oplus_{D_2} V_2 \\ W_3 &= U_3 \oplus_{D_2} V_3 \\ W_4 &= U_4 \oplus_{D_1} V_4. \end{aligned}$$

For a given value of j, consider the ω-words $\mathbf{U} = \mathbf{t}[8j+1..\infty]$, $\mathbf{V} = \mathbf{t}[8j+5..\infty]$, $\mathbf{W} = \mathbf{t}[16j+6..\infty]$. Let the length 17 prefix of $\mathbf{W}$ be W_0. Thus $W_0 \in \{W_1, W_2, W_3, W_4\}$. As per Remark 3, one can determine the lexicographically least D_0 with prefix 1 such that $W_0 = U_0 \oplus_{D_0} V_0$ for some prefixes U_0 of U and V_0 of V; we need only consider prefixes of $\mathbf{U}$ and $\mathbf{V}$ of lengths at most 17. It is therefore a finite computation to show that whenever $W_0 \in \{W_1, W_4\}$, then $D_0 = D_1$ and when $W_0 \in \{W_2, W_3\}$, then $D_0 = D_2$. For convenience, define $\delta : \{1, 2, 3, 4\} \to \{1, 2\}$ by $\delta(1) = \delta(4) = 1$, $\delta(2) = \delta(3) = 2$.

Let $T_0 = 0110100$, the length 7 prefix of the Thue-Morse word $\mathbf{t}$. A short computation (feasible by hand) shows that the lexicographically least word Δ_0 with prefix 1 such that $T_0 = T_1 \oplus_{\Delta_0} T_2$ for prefixes T_1, T_2 of $\mathbf{t}$ is $\Delta_0 = 1111101$.

We remark that each of D_1, D_2 and Δ_0 ends in a 1.

Theorem 6. *The lexicographically least word d with prefix 1 such that $\mathbf{t} = \mathbf{t} \oplus_d \mathbf{t}$ is*

$$d = 111110 \prod_{j=0}^{\infty} (D_{\delta(i_j)})'.$$

Proof. Note that

$$\mathbf{t} = 011010 \prod_{j=0}^{\infty} W'_{i_j} = 0 \prod_{j=0}^{\infty} U'_{i_j} = 01101 \prod_{j=0}^{\infty} V'_{i_j}.$$

The result thus follows from Corollary 5. □

Remark 7. One verifies that this is the shuffle given by Charlier et al. in [1].

5 Acknowledgments

Thanks to the anonymous and thorough referee, and to Narad Rampersad, for their comments.

References

[1] E. Charlier, T. Kamae, S. Puzynina, and L. Q. Zamboni, Self-shuffling words, *J. Combin. Theory Ser. A* **128** (2014), 1–40.

[2] J. Endrullis and D. Hendriks, The optimal self-shuffle of the Thue-Morse word, preprint available at http://www.cs.vu.nl/~diem/publication/pdf/self-shuffle.pdf, 2014.

[3] D. Henshall, N. Rampersad, and J. Shallit, Shuffling and unshuffling, *Bull. Europ. Assoc. Theoret. Comput. Sci.* No. 107 (June 2012), 131–142.

[4] M. Lothaire, *Algebraic Combinatorics on Words*, Cambridge University Press, 2002.

[5] Axel Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiana* **1** (1912), 1–67. Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, 1977, pp. 413–478.

2010 *Mathematics Subject Classification*: Primary 68R15.
Keywords: Word shuffling, Thue-Morse word, lexicographical order.

(Concerned with sequence A001285).

Received September 12 2014; revised versions received October 11 2014; October 12 2014. Published in *Journal of Integer Sequences*, November 2 2014.

Return to Journal of Integer Sequences home page.