Advances in Applied Mathematics 148 (2023) 102536

Contents lists available at ScienceDirect

Advances in Applied Mathematics

journal homepage: www.elsevier.com/locate/yaama

Two-block substitutions and morphic words

Michel Dekking [a,b,*], Michael Keane [c,d]

[a] *CWI, Amsterdam, the Netherlands*
[b] *3TU Applied Mathematics Institute and Delft University of Technology, Faculty EWI, P.O. Box 5031, 2600 GA Delft, the Netherlands*
[c] *3TU Applied Mathematics Institute and Delft University of Technology, Faculty EWI, the Netherlands*
[d] *Mathematical Institute, University of Leiden, Niels Bohrweg 1, 2333 CA Leiden, the Netherlands*

A R T I C L E   I N F O

Article history:
Received 12 March 2023
Received in revised form 30 March 2023
Accepted 4 April 2023
Available online xxxx

MSC:
68R15

Keywords:
Two-block substitutions
Kolakoski sequence
Morphic words
Base 3/2

A B S T R A C T

We consider in general two-block substitutions and their fixed points. We prove that some of them have a simple structure: their fixed points are morphic sequences. Others are intrinsically more complex, such as the Kolakoski sequence. We prove this for the Thue-Morse sequence in base 3/2.

* Corresponding author at: 3TU Applied Mathematics Institute and Delft University of Technology, Faculty EWI, P.O. Box 5031, 2600 GA Delft, the Netherlands.

E-mail addresses: f.m.dekking@tudelft.nl, Michel.Dekking@cwi.nl (M. Dekking), m.s.keane@tudelft.nl (M. Keane).

https://doi.org/10.1016/j.aam.2023.102536

1. Introduction

Let $A = \{0, 1\}$, A^* the monoid of all words over A, and let T^* be the submonoid of 0-1-words of even length. A *two-block substitution* κ is a map

$$\kappa : \{00, 01, 10, 11\} \to A^*.$$

A two-block substitution κ acts on T^* by defining for $w_1w_2 \dots w_{2m-1}w_{2m} \in T^*$

$$\kappa(w_1w_2 \dots w_{2m-1}w_{2m}) = \kappa(w_1w_2) \dots \kappa(w_{2m-1}w_{2m}).$$

In the case that $\kappa(T^*) \subseteq T^*$, we call κ 2-*block stable*. This property entails that the iterates κ^n are all well-defined for $n = 1, 2, \dots$.

The most interesting example of a two-block substitution that is *not* 2-block stable is the Oldenburger-Kolakoski two-block substitution κ_{K} given by

$$\kappa_{\mathrm{K}}(00) = 10, \quad \kappa_{\mathrm{K}}(01) = 100, \quad \kappa_{\mathrm{K}}(10) = 110, \quad \kappa_{\mathrm{K}}(11) = 1100.$$

The fact that κ_{K} is not 2-block stable, and so its iterates κ_{K}^n are not defined, makes it very hard to establish properties of the fixed point $x_{\mathrm{K}} = 110010\dots$ (usually written as 221121...) of κ_{K}, see, e.g., [4].

In Section 2 we show that even if a two-block substitution κ_{K} is *not* 2-block stable, then still it can be well-behaved in the sense that its fixed points are pure morphic words.

In Section 3 we prove that the Thue-Morse word in base $3/2$ is not well-behaved: it cannot be generated as a coding of a fixed point of a morphism.

This is a remarkable contrast with the behaviour of the sum of digits function for two seemingly more complicated bases: the Fibonacci base, and the golden mean base—see the paper [6].

2. Two-block substitutions with conjugated morphisms

Let κ be a two-block substitution on T^*, and let σ be a morphism on A^* with $\sigma(T^*) \subseteq T^*$. We say κ and σ commute if $\kappa\sigma(w) = \sigma\kappa(w)$ for all w from T^*.

In this case we say that σ is *conjugated* to κ.

Note that if $\kappa\sigma = \sigma\kappa$, then for all $n \geq 1$ one has $\kappa\sigma^n = \sigma^n\kappa$ on T^*.

Let $\sigma : A^* \to A^*$ be a morphism. Then σ induces a two-block substitution κ_σ by defining

$$\kappa_\sigma(ij) = \sigma(ij) \quad \text{for } i, j \in A.$$

We mention the following property of κ_σ, which is easily proved by induction.

Proposition 1. *Let $\sigma : A^* \to A^*$ be a morphism, let n be a positive integer, and suppose that κ_σ is two-block stable. Then $\kappa_\sigma^n = \kappa_{\sigma^n}$.*

We call σ the *trivial conjugated morphism* of the two block substitution κ_σ.

Not all morphisms σ can occur as trivial conjugated morphisms, but many will be according to the following simple property.

Proposition 2. *Any morphism σ on $\{0,1\}$ with the lengths of $\sigma(0)$ and $\sigma(1)$ both odd or both even is conjugated to the two-block substitution $\kappa = \kappa_\sigma$.*

Example: for the Fibonacci morphism φ defined by $\varphi(0) = 01, \varphi(1) = 0$, one can take the third power φ^3 to achieve this (cf. [13, A143667]).

In the remaining part of this section we discuss non-trivial conjugated morphisms.

Theorem 3. *Let κ be a two-block substitution on T^* conjugated with a morphism σ on A^*. Suppose that there exist i, j from A such that $\kappa(ij)$ has prefix ij, and such that ij is also prefix of a fixed point x of σ. Then also κ has fixed point x.*

Proof. Letting $n \to \infty$ in $\kappa\sigma^n(ij) = \sigma^n\kappa(ij) = \sigma^n(ij\dots)$ gives $\kappa(x) = x$. □

The Pell word $w_{\mathrm{P}} = 0010010001001\dots$ is the unique fixed point of the Pell morphism π given by

$$\pi : \begin{cases} 0 \to 001 \\ 1 \to 0. \end{cases}$$

The following result proves a conjecture from R.J. Mathar in [13, A289001]. The difficulty here is that since the 2-block substitution in Theorem 4 has the property that $\kappa(0010) = 0010010$ has odd length, the two-block substitution κ is not 2-block stable.

Theorem 4. *Let κ be the two-block substitution[1]:*

$$\kappa : \begin{cases} 00 \to 0010 \\ 01 \to 001 \\ 10 \to 010. \end{cases}$$

Then the unique fixed point of κ is the Pell word w_{P}.

Proof. We apply Theorem 3 with $ij = 00$.

Note first that $\pi(T^*) \subseteq T^*$. Next, we have to establish that κ and π commute on T^*.

It suffices to check this for the three generators 00, 01 and 10 from the four generators of T^*:

[1] Here it is not necessary to define $\kappa(11)$, since 11 does not occur in images of words without 11 under κ.

$$\kappa\pi(00) = \kappa(001001) = 0010010001 = \pi(0010) = \pi\kappa(00),$$
$$\kappa\pi(01) = \kappa(0010) = 0010010 = \pi(001) = \pi\kappa(01),$$
$$\kappa\pi(10) = \kappa(0001) = 0010001 = \pi(010) = \pi\kappa(10). \quad \square$$

3. Thue-Morse in base 3/2

A natural number N is written in base 3/2 if N has the form

$$N = \sum_{i\geq 0} d_i \left(\frac{3}{2}\right)^i, \tag{1}$$

with digits $d_i = 0, 1$ or 2.

We write these expansions as

$$\mathrm{SQ}(N) = d_R(N) \dots d_1(N) d_0(N) = d_R \dots d_1 d_0.$$

Let for $N \geq 0$, $s_{3/2}(N) := \sum_{i=0}^{i=R} d_i(N)$ be the sum of digits function of the base 3/2 expansions. The Thue-Morse word in base 3/2 is the word $(x_{3/2}(N)) := (s_{3/2}(N) \bmod 2) = 0100101011011010101\dots$

Theorem 5. *([5]) Let the two-block substitution κ_{TM} be defined by*

$$\kappa_{\mathrm{TM}} : \begin{cases} 00 & \to 010 \\ 01 & \to 010 \\ 10 & \to 101 \\ 11 & \to 101 \end{cases}$$

Then the word $x_{3/2}$ is the fixed point of κ_{TM} starting with 0.

The Thue-Morse word t is fixed point with prefix 0 of the Thue-Morse morphism τ : $0 \to 01$, $1 \to 10$. It satisfies the recurrence relations $t(2N) = t(N)$, $t(2N+1) = 1 - t(N)$.

The fixed point $x_{3/2}$ satisfies very similar recurrence relations:

$$x_{3/2}(3N) = x_{3/2}(2N),\ x_{3/2}(3N+1) = 1 - x_{3/2}(2N),\ x_{3/2}(3N+2) = x_{3/2}(2N).$$

We call κ_{TM} the *Thue-Morse two-block substitution.*

We now discuss the Kolakoski word x_{K}. This word was introduced by Kolakoski (years after Oldenburger [12]) as a problem in [8]. The problem was to prove that x_{K} is not eventually periodic. Its solution in [9] is however incorrect (The claim that words w with minimal period N in $www\dots$ map to words with period N_1 satisfying $N < N_1 < 2N$ by replacing run lengths by the runs themselves is false. For example, if the period word is $w = 21221$, then ww maps to the period word 2212211211211221, or its binary complement image.) A stronger result was proved by both Carpi [3] and Lepistö [10]: x_{K} does not contain any cubes. The fixed point $x_{3/2}$ of κ_{TM} has more repetitiveness. It contains for example the fourth power 01010101.

The Thue-Morse word is a purely morphic word, i.e., fixed point of a morphism. It is known that the Kolakoski word is not purely morphic ([4]). However it is still open whether the Kolakoski word is morphic, i.e., image under a coding (letter to letter map) of a fixed point of a morphism. The tool here is the subword complexity function $(p(N))$, which gives the number of words of length N occurring in an infinite word. A well known result tells us that when the subword complexity function increases too fast, faster than N^2, then a word can not be morphic. There is one example of a two-block substitution which yields a word that is not morphic given by Lepistö in the paper [11].

Theorem 6. *([11]) Let the two-block substitution κ_{L} be defined by*

$$\kappa_{\mathrm{L}}:\begin{cases}00 & \to 011\\ 01 & \to 010\\ 10 & \to 001\\ 11 & \to 000\end{cases}$$

Then the fixed point 010011000011 . . . *of κ_{L} has subword complexity function $p(N)$ satisfying $p(N) > C \cdot N^t$ for some $C > 0$ and $t > 2$.*

We do not know how to prove this ‘faster than quadratic’ property for the base 3/2 Thue-Morse word, but still we can use Lepistö’s result to obtain the following.

Theorem 7. *The base* 3/2 *Thue-Morse word $x_{3/2}$ is not a morphic word.*

The proof of Theorem 7 will be based on what we call the base 3/2 Toeplitz word.

Recall (see, e.g., [1, Lemma 3]) that the binary base Toeplitz word $z = 01000\ldots$ is directly derived from the binary Thue-Morse word $t = 01101001\ldots$ by putting $z(N) = t(N) + t(N+1) + 1 \mod 2$. It appears that for the generalization to base 3/2, there is a subtle move: $z(N) = t(N) + t(N+1) + 1 \mod 2$ is equivalent to $z(N) = t(2N) + t(2N+2) + 1 \mod 2$. We therefore define the *base 3/2 Toeplitz word* x_{T} for $N \geq 0$ by

$$x_{\mathrm{T}}(N) = x_{3/2}(3N) + x_{3/2}(3N+3) + 1 \mod 2. \tag{2}$$

So $x_{\mathrm{T}} = 101100111100\ldots$.

With some effort one can find in the paper [7, Theorem 3.2] a completely different proof of our next result.

Theorem 8. *The base 3/2 Toeplitz word x_{T} is the unique fixed point of the two-block substitution given by*

$$\kappa_{\rm T}: \begin{cases} 00 & \to 111 \\ 01 & \to 110 \\ 10 & \to 101 \\ 11 & \to 100 \end{cases}$$

Proof. In this proof $\equiv$ denotes equality modulo 2. The goal is to show that $x_{\rm T}$ satisfies for $m \geq 0$ the recurrence relations in Equations (3), (4), (5). This implies directly that $x_{\rm T}$ is fixed point of the 2-3-block substitution $a, b \to 1, a+1, b+1$. Taking $a, b = 0, 1$ one then obtains $\kappa_{\rm T}$.

$$x_{\rm T}(3m) \equiv 1, \tag{3}$$

$$x_{\rm T}(3m+1) \equiv x_{\rm T}(2m) + 1, \tag{4}$$

$$x_{\rm T}(3m+2) \equiv x_{\rm T}(2m+1) + 1. \tag{5}$$

The proof of these equations is based on the properties of the 6-9-block substitution generated by $\kappa_{\rm TM}$:

$$\lambda_{\rm TM}: \begin{cases} 010010 & \to 010010101 \\ 010101 & \to 010010010 \\ 101010 & \to 101101101 \\ 101101 & \to 101101010 \end{cases}$$

It is easy to see that $x_{3/2}$ is the fixed point of $\lambda_{\rm TM}$ starting with 010010. We first prove Equation (3). Consider $N = 3m$. Then $3N = 9m$, and $3N + 3 = 9m + 3$. So by Equation (2) we have

$$x_{\rm T}(3m) \equiv x_{3/2}(9m) + x_{3/2}(9m+3) + 1.$$

But $x_{3/2}(9m)$ and $x_{3/2}(9m+3)$ are the first and the fourth letter in an image block of length 9 of $\lambda_{\rm TM}$, which are generated by the first and the third letter of the corresponding source block of $\lambda_{\rm TM}$. For any source block these two letters are equal (simply because the source blocks occur at a position $0 \mod 3$ in $x_{3/2}$).

The conclusion is that $x_{\rm T}(3m) = x_{3/2}(9m) + x_{3/2}(9m+3) + 1 \equiv 1$ for all m.

To prove Equation (4), consider $N = 3m+1$. Then $3N = 9m+3$, and $3N+3 = 9m+6$. So by Equation (2) we have

$$x_{\rm T}(3m+1) \equiv x_{3/2}(9m+3) + x_{3/2}(9m+6) + 1.$$

But $x_{3/2}(9m+3)$ and $x_{3/2}(9m+6)$ are the fourth letter and the seventh letter in an image block of length 9 of $\lambda_{\rm TM}$, which are generated by the third and the fifth letter of the corresponding source block of $\lambda_{\rm TM}$. These are at positions $6m+2$, respectively $6m+4$. So

$$x_{3/2}(9m+3) = x_{3/2}(6m+2),\ x_{3/2}(9m+6) = x_{3/2}(6m+4).$$

On the other hand, by Equation (2) we have

$$x_{\mathrm{T}}(2m) \equiv x_{3/2}(6m) + x_{3/2}(6m+3)\ +1.$$

But $x_{3/2}(6m) = x_{3/2}(6m+2)$, because they are the first and the third letter in a block 010 or 101. Also, $x_{3/2}(6m+3)+1 \equiv x_{3/2}(6m+4)$, because $x_{3/2}(6m+3)$, respectively $x_{3/2}(6m+4)$ are the first and the second letter in a block 010 or 101.

The conclusion is that for all m

$$\begin{aligned} x_{\mathrm{T}}(3m+1) &\equiv x_{3/2}(9m+3) + x_{3/2}(9m+6) + 1 \equiv x_{3/2}(6m) + x_{3/2}(6m+3) + 1 + 1 \\ &\equiv x_{\mathrm{T}}(2m) + 1. \end{aligned}$$

To prove Equation (5), consider $N = 3m+2$. Then $3N = 9m+6$, and $3N+3 = 9m+9$. So by Equation (2) we have

$$x_{\mathrm{T}}(3m+2) \equiv x_{3/2}(9m+6) + x_{3/2}(9m+9) + 1.$$

But $x_{3/2}(9m+6)$ and $x_{3/2}(9m+9)$ are the seventh letter and the first letter in an image block of length 9 of λ_{TM}, which are generated by the third and the first letter of the corresponding source block of λ_{TM}. These are at positions $6m+4$, respectively $6m+6$. So

$$x_{3/2}(9m+6) = x_{3/2}(6m+4),\ x_{3/2}(9m+9) = x_{3/2}(6m+6).$$

On the other hand, by Equation (2) we have

$$x_{\mathrm{T}}(2m+1) \equiv x_{3/2}(6m+3) + x_{3/2}(6m+6)\ +1.$$

But $x_{3/2}(6m+3) \equiv x_{3/2}(6m+4)+1$, because they are the first and the second letter in a block 010 or 101. The conclusion is that for all m

$$\begin{aligned} x_{\mathrm{T}}(3m+2) &\equiv x_{3/2}(9m+6) + x_{3/2}(9m+9) + 1 \equiv x_{3/2}(6m+3) + 1 + x_{3/2}(6m+6) + 1 \\ &\equiv x_{\mathrm{T}}(2m+1) + 1. \quad \square \end{aligned}$$

Proof of Theorem 7. The crucial observation is that the base 3/2 Toeplitz two-block substitution κ_{T} is just the binary complement of the κ_{L} two-block substitution. In particular Theorem 6 also holds for the base 3/2 Toeplitz word, and so x_{T} cannot be a morphic word.

Suppose that the base 3/2 Thue-Morse word $(x_{3/2}(N))$ is a morphic word. Then an application of [2, Theorem 7.9.1] yields that the word $(x_{3/2}(3N))$ is morphic. Next, [2, Theorem 7.6.4] gives that the direct product word $([x_{3/2}(3N), x_{3/2}(3(N{+}1))])$ is morphic.

Finally, another application of [2, Theorem 7.9.1] yields that according to Equation (2) this direct product word maps to a morphic word $(x_{\mathrm{T}}(N))$ under the morphism $[0,0] \mapsto 1$, $[0,1] \mapsto 0$, $[1,0] \mapsto 0$, $[1,1] \mapsto 1$. But this contradicts the fact that $(x_{\mathrm{T}}(N))$ is not morphic. Hence the base 3/2 Thue-Morse word is not a morphic word. $\square$

Acknowledgment

We are grateful to Jean-Paul Allouche for several useful comments.

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, Discrete Math. 139 (1995) 455–461.

[2] J.-P. Allouche, J. Shallit, Automatic Sequences: Theory, Applications, Generalizations, Cambridge University Press, 2003.

[3] A. Carpi, Repetitions in the Kolakovski sequence, Bull. Eur. Assoc. Theor. Comput. Sci. 50 (1993) 194–196.

[4] F.M. Dekking, What is the long range order in the Kolakoski sequence?, in: R.V. Moody (Ed.), Proceedings of the NATO Advanced Study Institute, Waterloo, ON, August 21–September 1, 1995, Kluwer, Dordrecht, Netherlands, 2009, pp. 115–125.

[5] F.M. Dekking, The Thue-Morse sequence in base 3/2, J. Integer Seq. 26 (2023) 23.2.3.

[6] F.M. Dekking, The sum of digits functions of the Zeckendorf and the base phi expansions, Theor. Comput. Sci. 859 (2021) 70–79.

[7] T. Edgar, H. Olafson, J. Van Alstine, Some combinatorics of rational base representations, preprint, available at https://community.plu.edu/~edgartj/preprints/basepqarithmetic.pdf, 2014.

[8] W. Kolakoski, Self generating runs, Problem 304, Am. Math. Mon. 72 (1965) 674.

[9] W. Kolakoski, N. Ücoluk, Solution to self generating runs, Problem 304, Am. Math. Mon. 73 (1966) 681–682.

[10] A. Lepistö, Repetitions in Kolakoski sequence, in: Developments in Language Theory, 1994, pp. 130–143.

[11] A. Lepistö, On the power of periodic iteration of morphisms, in: ICALP 1993, in: Lect. Notes Comp. Sci, vol. 700, 1993, pp. 496–506.

[12] R. Oldenburger, Exponent trajectories in symbolic dynamics, Trans. Am. Math. Soc. 46 (3) (1939) 453–466.

[13] The On-Line Encyclopedia of Integer Sequences, founded by N.J.A Sloane, electronically available at https://oeis.org.