AUSTRALASIAN JOURNAL OF COMBINATORICS
Volume **85(2)** (2023), Pages 150–158

The periodic complexity function of the Thue–Morse word, the Rudin–Shapiro word, and the period-doubling word

Narad Rampersad

Department of Mathematics and Statistics
University of Winnipeg
Winnipeg MB, R3B 2E9
Canada
narad.rampersad@gmail.com

Abstract

We revisit the periodic complexity function $h_{\mathbf{w}}(n)$ introduced by Mignosi and Restivo. This function gives the average of the first n local periods of a recurrent infinite word $\mathbf{w}$. Our method for computing the asymptotics of the periodic complexity function is different than that of Mignosi and Restivo. We apply it to the Thue–Morse word, the Rudin–Shapiro word, and the period-doubling word.

1 Introduction

Mignosi and Restivo [3] introduced a new complexity measure for infinite words called the *periodic complexity*. This function is defined based on the *local period* at each position of the infinite word. Let $\mathbf{w} = w_0w_1w_2\cdots$ be an infinite word. The *periodicity function* $p_{\mathbf{w}}(i)$ is defined as follows. The value of $p_{\mathbf{w}}(i)$ is the length of the shortest prefix u of $w_iw_{i+1}w_{i+2}\cdots$ such that either u *is a suffix of* $w_0\cdots w_{i-1}$ or $w_0\cdots w_{i-1}$ *is a suffix of* u, if such a word u exists. If no such u exists, then $p_{\mathbf{w}}(i) = \infty$. However, if $\mathbf{w}$ is recurrent (i.e., every factor of $\mathbf{w}$ occurs infinitely often in $\mathbf{w}$), which will always be the case in this paper, then $p_{\mathbf{w}}(i) < \infty$ for all i.

For example, if $\mathbf{w} = 1011001011\cdots$, then $p_{\mathbf{w}}(4) = 6$, since 001011 has suffix 1011, and $p_{\mathbf{w}}(5) = 1$, since 10110 has suffix 0.

Since the values of $p_{\mathbf{w}}(i)$ can fluctuate wildly, it is not that suitable as a complexity function. Mignosi and Restivo therefore defined the *periodic complexity function* $h_{\mathbf{w}}(i)$ as the *average of the periodicity function*; that is, if

$$P_{\mathbf{w}}(i) = \sum_{j=0}^{i-1} p_{\mathbf{w}}(j)$$

ISSN: 2202-3518

is the *summatory function* of $p_{\mathbf{w}}(i)$, then $h_{\mathbf{w}}(i) = (1/i)P_{\mathbf{w}}(i)$ for $i \geq 1$.

Mignosi and Restivo studied the periodicity function and the periodicity complexity function for both the *Thue–Morse word*

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

and the *Fibonacci word*

$$\mathbf{f} = 0100101001001010\cdots.$$

They proved that $h_{\mathbf{t}}(n) = \Theta(n)$ and $h_{\mathbf{f}}(n) = \Theta(\log n)$. (Recall that $f(n) = \Theta(g(n))$ means that there are positive constants C and D such that $Cg(n) \leq f(n) \leq Dg(n)$ for n sufficiently large.) Schaeffer [5] studied the periodicity function of Sturmian words using the Ostrowski representation of natural numbers.

In this paper we study $p_{\mathbf{t}}(i)$ and $h_{\mathbf{t}}(i)$ with the aid of the computer program Walnut [4]. We get a more precise description of these functions than the ones given in [3] and we show how to apply these techniques to other automatic sequences, such as the Rudin–Shapiro sequence.

2 Periodic complexity of the Thue–Morse word

The Thue–Morse word $\mathbf{t} = t_0t_1t_2\cdots$ is defined by

$$t_i = \begin{cases} 0 & \text{if the number of 1's in the binary representation of } i \text{ is even,} \\ 1 & \text{otherwise.} \end{cases}$$

Table 1 shows some initial values of $p_{\mathbf{t}}(i)$.

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$p_{\mathbf{t}}(i)$	1	3	1	6	2	12	1	12	1	24	1	24	2	24	1	24

Table 1: Initial values of $p_{\mathbf{t}}(i)$

Consider the following logical formulas:

$$\begin{aligned}\phi_1(i,j,n) &:= \forall k,\ k < n \Rightarrow \mathbf{t}[i+k] = \mathbf{t}[j+k] \\ \phi_2(i,n) &:= (i \geq n \wedge \phi_1(i, i-n, n)) \mid (i < n \wedge \phi_1(0, n, i)) \\ \phi_3(i,n) &:= n > 0 \wedge \phi_2(i,n) \wedge (\forall m,\ (m > 0 \wedge m < n) \Rightarrow \neg\phi_2(i,m))\end{aligned}$$

Then the pairs $(i, p_{\mathbf{t}}(i))$ are exactly the pairs that satisfy ϕ_3.

We can get an automaton that computes the binary representation of $p_{\mathbf{t}}(i)$ with the following Walnut commands (see [6, Section 10.8.12]):

```
def tmEq "?msd_2 Ak (k<n) => T[i+k]=T[j+k]":

def tmRepWd "?msd_2 (i>=n & $tmEq(i,i-n,n)) | (n>i & $tmEq(0,n,i))":

def tmLocPer "?msd_2 (n>0) & $tmRepWd(i,n) & Am (m>0 & m<n) =>
    ~$tmRepWd(i,m)":
```

This produces the automaton in Figure 1. By examining this automaton, one obtains the following result, which is a more precise version of [3, Proposition 3.18].

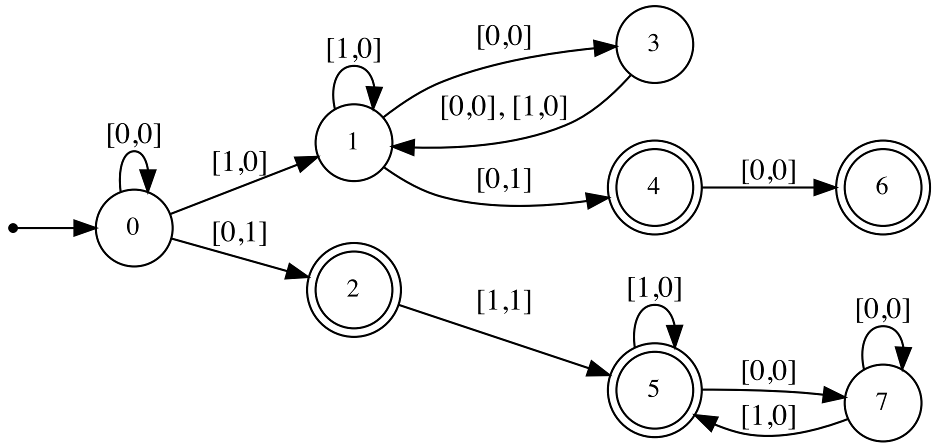

Figure 1: Automaton for the pair $(i, p_{\mathbf{t}}(i))$

Proposition 2.1. *We have*

- $p_{\mathbf{t}}(i) \in \{1, 2\}$ *if i is even; and,*
- $p_{\mathbf{t}}(i) = 3 \cdot 2^t$ *if i is odd and $2^t \le i < 2^{t+1}$.*

We can then bound the summatory function of $p_{\mathbf{t}}(i)$.

Proposition 2.2. *For $n \ge 1$, we have*

$$\frac{3}{8}(n-1)^2 + \frac{n}{2} \;\le\; P_{\mathbf{t}}(n) \;\le\; \frac{3}{4}n^2 + n + 1.$$

Proof. We split the sum $P_{\mathbf{t}}(n) = \sum_{i=0}^{n-1} p(i)$ into even and odd indexed terms. By Proposition 2.1, we have

$$\frac{n}{2} \;\le\; \sum_{\substack{i=0 \\ i \text{ even}}}^{n-1} p(i) \;\le\; n+1.$$

Again, by Proposition 2.1, we have

$$\sum_{\substack{i=0 \\ i \text{ odd}}}^{n-1} p(i) \;\le\; \sum_{\substack{i=0 \\ i \text{ odd}}}^{n-1} 3i \;\le\; 3(n/2)^2 \;=\; \frac{3}{4}n^2.$$

and

$$\sum_{\substack{i=0\\ i \text{ odd}}}^{n-1} p(i) \;\geq\; \sum_{\substack{i=0\\ i \text{ odd}}}^{n-1} 3i/2 \;\geq\; (3/2)[(n-1)/2]^2 \;=\; \frac{3}{8}(n-1)^2.$$

Hence,

$$\frac{3}{8}(n-1)^2 + \frac{n}{2} \;\leq\; P_{\mathbf{t}}(n) \;\leq\; \frac{3}{4}n^2 + n + 1.$$

□

This gives the following bounds on the periodic complexity of $\mathbf{t}$, which are an improvement on the inequalities from the proof of [3, Proposition 3.19].

Theorem 2.3. *For $n \geq 1$, we have*

$$3n/8 - 1/4 \;\leq\; h_{\mathbf{t}}(n) \;\leq\; 3n/4 + 2.$$

In particular, we have $h_{\mathbf{t}}(n) = \Theta(n)$.

In this case, we were fortunate that the automaton in Figure 1 was rather simple. For more complicated sequences, this may not be the case, so next we explore other methods for analyzing the asymptotics of $P_{\mathbf{t}}(i)$. To apply these methods, we first need a *linear representation* for $p_{\mathbf{t}}(i)$. That is, we need an integer row vector v, an integer column vector w, and a pair of integer matrices M_0 and M_1, such that

$$p_{\mathbf{t}}(i) = vM_{i_{\ell-1}}M_{i_{\ell-2}} \cdots M_{i_0}w,$$

where $i_{\ell-1}i_{\ell-2}\cdots i_0$ is the binary representation of i. Walnut can produce a linear representation for $p_{\mathbf{t}}(i)$ with the command

```
eval tmLocPer_enum i "?msd_2 En $tmLocPer(i,n) & m<n & ~$tmLocPer(i,m)":
```

The output of this command is a Maple worksheet containing the following values for v, w, M_0 and M_1.

$$v = [1,0,1,0,0,0],$$

$$M_0 = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 1\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 2\\ 0 & 1 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}, \qquad M_1 = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0\\ 0 & 1 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 2 & 0 & 0\\ 0 & 1 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 2 & 0 & 0 \end{bmatrix},$$

$$w = [1,1,0,1,1,0]^T.$$

Sequences defined by such linear representations are called 2*-regular sequences* (in general, q*-regular sequences*). Dumas [1] obtained a description of the asymptotics of the summatory function of q-regular sequences (also see Heuberger and Krenn [2]).

To make use of these results, we need a number of definitions (see [2, Section 3.2]). Let $X(N) = \sum_{n=0}^{N-1} x(n)$ be the summatory function of a sequence $x(n)$ for which we have a linear representation consisting of a row vector $v \in \mathbb{C}^d$, a column vector $w \in \mathbb{C}^d$, and q matrices $M_0, \ldots, M_{q-1} \in \mathbb{C}^{d\times d}$. That is,

$$x(n) = vM_{n_{\ell-1}}M_{n_{\ell-2}} \cdots M_{n_0}w, \tag{1}$$

where $n_{\ell-1}n_{\ell-2}\cdots n_0$ is the base-q representation of n. Let $\|\cdot\|$ denote any norm on $\mathbb{C}^d$, as well as its induced matrix norm. Define $M := M_0 + M_1 + \ldots + M_{q-1}$. Choose $R > 0$ such that $\|M_{r_1}M_{r_2}\cdots M_{r_\ell}\| = O(R^\ell)$ holds for all $\ell \geq 0$ and all $r_1, \ldots, r_\ell \in \{0, \ldots, q-1\}$. That is, the number R is an upper bound for the *joint spectral radius* of $M_0, \ldots, M_{q-1}$. Let $\sigma(M)$ denote the set of eigenvalues of M. For $\lambda \in \mathbb{C}$, if $\lambda \in \sigma(M)$, let $m(\lambda)$ denote the size of the largest Jordan block of M associated with λ, and let $m(\lambda) = 0$ otherwise. The following result is essentially [1, Theorem 1] as presented in the first part of [2, Theorem A].

Theorem 2.4. *With the above definitions, we have*

$$X(N) = \sum_{\substack{\lambda\in\sigma(M)\\|\lambda|>R}} N^{\log_q \lambda} \sum_{0\leq k<m(\lambda)} \frac{(\log N)^k}{k!}\Phi_{\lambda k}(\log_q N) + O\left(N^{\log_q R}(\log N)^{\max\{m(\lambda):|\lambda|=R\}}\right),$$

where the $\Phi_{\lambda k}$ are certain 1-periodic continuous functions. The big O "error term" can be omitted if there are no eigenvalues $\lambda \in \sigma(M)$ with $|\lambda| \leq R$.

Note that we have defined the linear representation of $x(n)$ in terms of the most-significant-digit first representation of n. It can also be defined using the least-significant-digit first representation of n (as it is in [2]). One can easily convert from one representation to the other by taking the transpose of v, $M_0, \ldots, M_{q-1}$, w, and the transpose of Equation 1. Since the eigenvalues of a matrix and its transpose are the same, we can still apply Theorem 2.4, regardless of the choice of representation.

If we return to the linear representation of $p_{\mathbf{t}}(i)$ that we computed earlier, we have

$$M = M_0 + M_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0\\ 0 & 1 & 0 & 1 & 1 & 1\\ 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 2 & 0 & 2\\ 0 & 2 & 0 & 2 & 0 & 0\\ 0 & 0 & 0 & 2 & 0 & 2 \end{bmatrix}.$$

The set of eigenvalues of M is $\sigma(M) = \{4, 2, 1, 0, -1\}$, where each eigenvalue has multiplicity 1, except the eigenvalue 0, which has multiplicity 2. To compute R it is convenient for us to choose the $\|\cdot\|_\infty$ norm on $\mathbb{C}^6$ (i.e., the maximum norm), which induces the maximum row sum norm on $\mathbb{C}^{6\times 6}$. Since the maximum row sum of M_0 and M_1 is 2, we can take $R = 2$. This is enough information to apply Theorem 2.4 to $P_{\mathbf{t}}(n)$, which gives the following result.

Theorem 2.5. *We have*

$$P_{\mathbf{t}}(n) = n^2\Phi(\log_2 n) + O(n\log n),$$

and

$$h_{\mathbf{t}}(n) = n\Phi(\log_2 n) + O(\log n),$$

for some 1*-periodic continuous function* Φ.

Without a precise description of the function Φ (which is beyond the scope of this paper), the statement of Theorem 2.5 is not so satisfying, and indeed the estimates of Proposition 2.2 and Theorem 2.3 are more informative. However, as we see in the next section, Theorem 2.4 can often quickly give the asympototics in cases where we cannot easily obtain precise estimates.

3 Periodic complexity of the Rudin–Shapiro sequence

We can determine the asymptotic growth of $h_{\mathbf{x}}(n)$ for other automatic sequences $\mathbf{x}$ by first using Walnut to compute a linear representation for $p_{\mathbf{x}}(i)$, and then applying Theorem 2.4. Let

$$\mathbf{rs} = r_0r_1r_2\cdots = 0001001000011101\cdots$$

be the *Rudin–Shapiro sequence*, defined by

$$r_i = \begin{cases} 0 & \text{if the number of 11's in the binary representation of } i \text{ is even,} \\ 1 & \text{otherwise.} \end{cases}$$

If we use Walnut to compute a linear representation for $p_{\mathbf{rs}}(i)$, the matrices M_0 and M_1 that we get are 31×31, so we do not show them here. They each have maximum row sum 2, so again we can take $R = 2$. The set of eigenvalues of the matrix $M := M_0 + M_1$ is $\sigma(M) = \{4, 2, 1, 0, -1, -2\}$. From the Jordan form of M, we find $m(4) = 1$ and $m(2) = m(-2) = 2$. Applying Theorem 2.4 thus gives the following result.

Theorem 3.1. *We have*

$$P_{\mathbf{rs}}(n) = n^2\Psi(\log_2 n) + O(n\log^2 n),$$

and

$$h_{\mathbf{rs}}(n) = n\Psi(\log_2 n) + O(\log^2 n),$$

for some 1*-periodic continuous function* Ψ.

4 Periodic complexity of the period-doubling sequence

Next we determine the asymptotic behaviour of $P_{\mathbf{pd}}(n)$ and $h_{\mathbf{pd}}(n)$, where $\mathbf{pd}$ is the *period-doubling word*, i.e., the fixed point of the morphism $0 \to 01$, $1 \to 00$. Table 2 shows some initial values of $p_{\mathbf{pd}}(i)$.

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$p_{\mathbf{pd}}(i)$	1	2	4	1	1	8	2	2	2	2	16	1	1	4	4	1

Table 2: Initial values of $p_{\mathbf{pd}}(i)$

Our goal is to show that $h_{\mathbf{pd}}(n) = \Theta(\log n)$ (i.e., its periodic complexity is rather more like that of the Fibonacci word (see [3]) than the Thue–Morse word).

We begin by using Walnut to compute the automaton for the pair $(i, p_{\mathbf{pd}}(i))$, which is given in Figure 2. This time, it is not so easy to see an analogue to Propo-

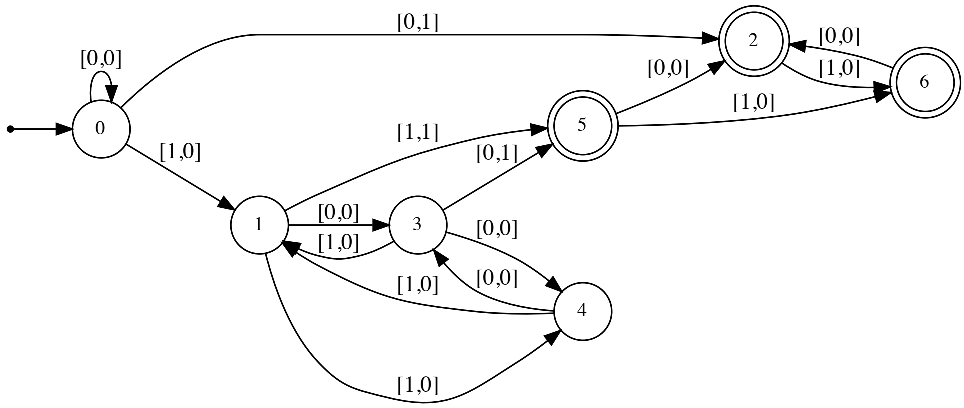

Figure 2: Automaton for the pair $(i, p_{\mathbf{pd}}(i))$

sition 2.1 by direct inspection of the automaton, so let us proceed by analyzing the corresponding linear representation:

$$v = [1, 0, 0, 0, 0, 0],$$

$$M_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad M_1 = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \end{bmatrix},$$

$$w = [1, 1, 1, 1, 1, 1]^T.$$

Now, if we try to apply Theorem 2.4 to $P_{\mathbf{pd}}(n)$, we run into the following problem. The maximum row sum of M_0 and M_1 is 2, so we could take $R = 2$, but the largest eigenvalue of $M := M_0 + M_1$ is also 2 (with $m(2) = 2$), which means that in this case the "error term" in Theorem 2.4 dominates, and we do not obtain the desired asymptotics. However, using other methods, we can obtain the following bounds.

Theorem 4.1. *For $n \geq 1$, we have*

$$(1/3\log_2 n - 1/18)n + 4/9 \ \leq \ P_{\mathbf{pd}}(n) \ \leq \ (4/3\log_2 n + 22/9)n + 5/9$$

and

$$1/3\log_2 n - 1/18 \ \leq \ h_{\mathbf{pd}}(n) \ \leq \ 4/3\log_2 n + 3.i$$

Proof. For $\ell \geq 0$, we have

$$\begin{aligned} P_{\mathbf{pd}}(2^\ell) &= \sum_{i<2^\ell} p(i) \\ &= \sum_{i_0,\ldots,i_{\ell-1}\in\{0,1\}} vM_{i_{\ell-1}}\cdots M_{i_0}w \\ &= v(M_0+M_1)^\ell w \\ &= vM^\ell w. \end{aligned}$$

From the Jordan form of M we find that

$$vM^\ell w = (A + B\ell)2^\ell + C(-2)^\ell + D + E(-1)^\ell, \tag{2}$$

for some constants $A, \ldots, E$. To compute these constants, we compute $vM^\ell w$ (i.e., $P_{\mathbf{pd}}(2^\ell)$) for $\ell = 0, \ldots, 4$, which gives the values $1, 3, 8, 21, 52$. We then substitute these values into (2) to obtain a system of linear equations in the variables $A, \ldots, E$. When we solve this system of linear equations we get

$$A = 5/9,\ B = 2/3,\ C = 0,\ D = 1/2,\ E = -1/18.$$

Thus, we have

$$P_{\mathbf{pd}}(2^\ell) = (5/9 + (2/3)\ell)2^\ell + 1/2 - 1/18(-1)^\ell,$$

and so

$$(5/9 + (2/3)\ell)2^\ell + 4/9 \ \leq \ P_{\mathbf{pd}}(2^\ell) \ \leq \ (5/9 + (2/3)\ell)2^\ell + 5/9.$$

Now write $2^\ell \leq n < 2^{\ell+1}$, so that $\ell \leq \log_2 n < \ell + 1$. Then

$$\begin{gathered} (5/9 + (2/3)\ell)2^\ell + 4/9 \leq P_{\mathbf{pd}}(n) \leq (5/9 + 2/3(\ell+1))2^{\ell+1} + 5/9 \\ (5/9 + 2/3(\log_2 n - 1))(n/2) + 4/9 \leq P_{\mathbf{pd}}(n) \leq (5/9 + 2/3(\log_2 n + 1))(2n) + 5/9 \\ (1/3\log_2 n - 1/18)n + 4/9 \leq P_{\mathbf{pd}}(n) \leq (4/3\log_2 n + 22/9)n + 5/9, \end{gathered}$$

and

$$1/3\log_2 n - 1/18 \leq h_{\mathbf{pd}}(n) \leq 4/3\log_2 n + 3.$$

□

We note that we could have also applied this method to the Thue–Morse word to obtain very similar bounds to those of Proposition 2.2 and Theorem 2.3, except with different constants. In fact, with this method we get a slightly better upper bound and a slightly worse lower bound.

Acknowledgments

We thank Jeffrey Shallit for suggesting the approach used in the proof of Theorem 4.1.

References

[1] P. Dumas, Joint spectral radius, dilation equations, and asymptotic behavior of radix-rational sequences, *Linear Algebra Appl.* 438 (2013), 2107–2126.

[2] C. Heuberger and D. Krenn, Asymptotic analysis of regular sequences, *Algorithmica* 82 (2020), 429–508.

[3] F. Mignosi and A. Restivo, A new complexity function for words based on periodicity, *Int. J. Algebra Comput.* 23 (2013), 963–987.

[4] H. Mousavi, Automatic theorem proving in Walnut; documentation (2016–2021) available at `https://arxiv.org/abs/1603.06017`.

[5] L. Schaeffer, Ostrowski numeration and the local period of Sturmian words, in: Proc. LATA 2013, LNCS 7810, pp. 493–503, Springer, 2013.

[6] J. Shallit, The Logical Approach to Automatic Sequences: Exploring Combinatorics on Words with `Walnut` (in press).

(Received 30 Dec 2021; revised 7 Oct 2022, 21 Nov 2022)