Some Remarks on Automatic Sequences, Toeplitz Words and Perfect Shuffling

Gabriele Fici

Università di Palermo, Italy

One World Combinatorics on Words Seminar

6th December 2021

Words

Definition

A (right) infinite word over an alphabet Σ is a map from $\mathbb{N}$ to Σ.

Example of constructions of words:

- *Purely morphic words*:
  Iterate a substitution
- *Morphic words*:
  Apply a coding to a purely morphic word
- *Toeplitz words*:
  Iteratively fill holes of a pattern with another pattern

Toeplitz Words

Definition

A word $x = x(0)x(1)x(2)\cdots$ is a Toeplitz word if for every n there exists p such that $x(n) = x(n+p) = x(n+2p) = \ldots$

Toeplitz construction: Let $s_1, s_2, \ldots$ be a sequence of finite words over $\Sigma \cup \{?\}$ such that each s_i contains $?$ and infinitely many s_i do not start with $?$.

Starting from the infinite word $s_1^\omega = s_1 s_1 s_1 \cdots$, replace the subsequence of $?$ by s_2^ω; repeat for every $i \geq 2$.

Toeplitz Words

Example

$s_i = 01?$ if i is odd or $10?$ if i is even.

$$01?01?01?01?01?01?01?01?01?01?01?01?01?01?01?01? \cdots$$

$$01101001?01101001?01101001?01101001?01101001?011 \cdots$$

$$0110100100110100110110100 1?0110100100110100 11011 \cdots$$

$\ldots$

We obtain the *Sierpiński gasket word* (A156595 in OEIS)

$$sg = 011010010011010011011010011011010010011010011011010 01 \cdots$$

which is also the fixed point of the substitution $0 \mapsto 011, 1 \mapsto 010$.

Toeplitz Words

Theorem (Boccuto, Carpi, 2020)

Let x be a Toeplitz word generated by the sequence of (primitive) words $(s_n)_{n\geq 0}$ of maximal length k. Then x is k-free.

Definition

A Toeplitz word is a *paperfolding word* if $s_i \in \{0?1?, 1?0?\}$ for every i.

We will only consider *simple* Toeplitz words, i.e., whose associated sequence is (ultimately) periodic.

Purely Morphic and Toeplitz

Some words are fixed point of morphisms (substitutive rules), some are Toeplitz (substitutive patterns), some are both, some are none.

Examples:

word	symbol	substitution	pattern
Thue-Morse	t	$0 \mapsto 01, 1 \mapsto 10$	NO
Fibonacci	f	$0 \mapsto 01, 1 \mapsto 0$	NO
period-doubling	d	$0 \mapsto 01, 1 \mapsto 00$	$010?$
regular paperfolding	p	NO	$0?1?$
alternate paperfolding	a	NO	$0?1?, 1?0?$
Sierpiński gasket	sg	$0 \mapsto 011, 1 \mapsto 010$	$01?, 10?$
Rudin-Shapiro	rs	NO	NO

Automatic Words

We are interested in words that are related to the string that represents n in some base b, e.g. $b = 2$.

Given a regular language L over $\Sigma_b = \{0, 1, \ldots, b-1\}$, and a partition $a_1, a_2 \cdots, a_\sigma$ of the set of states of a DFA accepting L, we can construct an infinite word $x = x(0)x(1)x(2)\cdots$ over $\Sigma = \{a_1, \cdots, a_\sigma\}$ by defining $x(n)$ to be a_i if the b-ary representation of n ends in a state belonging to a_i.

Such words are called b-automatic, and can be obtained by applying a coding to a fixed point of a b-uniform morphism (Cobham, 1972).

Proposition

A paperfolding word is automatic if and only if it is a simple Toeplitz word.

Automatic Words

In particular, one can define any binary automatic word from a rational property of the string representing n in base b by considering the partition induced by the accepting states.

Example

$L = \{$binary representations of n that contain an odd number of $1\}$ defines the Thue-Morse word $0110100110010110\cdots$

$L' = \{$ternary representations of n that end with 1, 02 or $122\}$ defines the Sierpiński gasket word $sg = 0110100100110100110110100110110\cdots$

b-adic Valuation

Definition

The 2-adic valuation of $n > 0$ is $\sup\{j : 2^j \text{ divides } n\}$.

$=$ the length of the longest 0-suffix of the binary representation of n.

For example, the 2-adic valuation of $n = 18 = 1001\underline{0}$ is 1.

The b-adic valuation, for any $b > 2$, is defined similarly.

Sometimes, it's convenient to describe a b-automatic word in terms of the b-adic valuation of n instead that in terms of the b-ary representation of n.

The Ruler Word

The sequence of 2-adic valuations of positive integers is the sequence

$$r_2 = 0102010301020104010201030102010501020103010201040\cdots$$

It's called *ruler sequence*, *Zimin word*, *infini-bonacci word*, etc.

- It's the lexicographically smallest square-free word;
- After deleting all 0s, one obtains the (isomorphic) word $r_2 + 1$;
- It's the fixed point of the (infinite) morphism

$$0 \mapsto 01,\ 1 \mapsto 02,\ \ldots, (n-1) \mapsto 0n,\ \ldots$$

- It's a Toeplitz word generated by the sequence $0?$, $1?$, $\cdots$, $n?$, $\cdots$

The 3-Ruler Word

The sequence of 3-adic valuations of positive integers is the sequence

$r_3 = 001001002001001002001001003001001002001001002001001003\cdots$

- It's the lexicographically smallest cube-free word;
- After deleting all 0s, one obtains the (isomorphic) word $r_3 + 1$;
- It's the fixed point of the (infinite) morphism

$$0 \mapsto 001,\ 1 \mapsto 002,\ \ldots,(n-1) \mapsto 00n,\ \ldots$$

- It's a Toeplitz word generated by the sequence $00?$, $11?$, $22?$, $\cdots$, $nn?$, $\cdots$

Generalization: The sequence r_b of b-adic valuations of positive integers is the lexicographically smallest word avoiding b-powers, etc.

Some Well-Known 2-Automatic Words

n	ruler 2-adic	per.doub. r mod 2	reg.paperfold. (r+2)th to last
00001	0	0	0
00010	1	1	0
00011	0	0	1
00100	2	0	0
00101	0	0	0
00110	1	1	1
00111	0	0	1
01000	3	1	0
01001	0	0	0
01010	1	1	0
01011	0	0	1
01100	2	0	1
01101	0	0	0
01110	1	1	1
01111	0	0	1
10000	4	0	0

m-bonacci words

n	ruler 2-adic	Fibonacci skip 00	Tribonacci skip 000
1	0	0	0
10	1	1	1
11	0	0	0
100	2		2
101	0	0	0
110	1	1	1
111	0	0	0
1000	3		
1001	0		0
1010	1	1	1
1011	0	0	0
1100	2		2
1101	0	0	0
1110	1	1	1
1111	0	0	0
10000	4		

Some Well-Known 2-Automatic Words

n	Thue-Morse odd $\#1$'s in n	Rudin-Shapiro odd $\#11$'s in n
0	0	0
1	1	0
10	1	0
11	0	1
100	1	0
101	0	0
110	0	1
111	1	0
1000	1	0
1001	0	0
1010	0	0
1011	1	1
1100	0	1
1101	1	1
1110	1	0
1111	0	1

Some Less-Known 2-Automatic Words

Let's do another example.

Skipping the first term, if one takes the sequence of sums of blocks of length 2 in the Thue-Morse word $t = 0110100110010\cdots$, one obtains the *Hall word* (a.k.a. *Thue squarefree word*, *variant of Thue-Morse*, etc.):

$$h = 210201210120210201202101210201210120210121020120\cdots$$

often defined as the fixed point of $2 \mapsto 210$, $1 \mapsto 20$, $0 \mapsto 1$. It is also the sequence that counts the number of 1's between two consecutive 0's in t.

As observed by Berstel in 1979, h can be obtained from the fixed point of $0 \mapsto 12$, $1 \mapsto 13$, $2 \mapsto 20$, $3 \mapsto 21$ by applying the coding $0, 3 \mapsto 1$; $1 \mapsto 0$, $2 \mapsto 2$.

Taking h modulo 2 one obtains the period-doubling word $d = 0100010101000100010\cdots$.

Some Less-Known 2-Automatic Words

Let's now do the same with the Rudin-Shapiro word.

Skipping the first term, if one takes the sequence of sums of two consecutive elements in the Rudin-Shapiro word $rs = 0001001000011 \cdots$ one obtains the word:

$$011002110111201001100212211 \cdots$$

which is not (yet!) in OEIS...

However, taking this word modulo 2, one obtains the alternate paperfolding word

$$a = 0110001101110010011000100111001110 \cdots$$

that is, the word obtained by Toeplitz construction alternating the patterns $0?1?$ and $1?0?$

Some Less-Known 2-Automatic Words

		sum	mod 2		sum	mod 2
	Thue-M.	Hall	per.doub.	Rud.-Shap.	(not in OEIS)	altern.pap.
1	1			0		
10	1	2	0	0	0	0
11	0			1		
100	1	1	1	0	1	1
101	0			0		
110	0	0	0	1	1	1
111	1			0		
1000	1	2	0	0	0	0
1001	0			0		
1010	0	0	0	0	0	0
1011	1			1		
1100	0	1	1	1	2	0
1101	1			1		
1110	1	2	0	0	1	1
1111	0			1		
10000	1	1	1	0	1	1

Perfect Shuffle of Two Words

The Thue-Morse word

$$t = 0110100110010110 \cdots$$

is the binary word starting with 0 that is equal to the perfect shuffling of itself with its complement:

$$t = t ⧢ (1 - t)$$

so it's a solution of the equation

$$x = x ⧢ (1 - x)$$

Definition

We call x_{even} and x_{odd} the two words whose perfect shuffling yields x. That is, $x = x_{even} ⧢ x_{odd}$

We are interested in describing x_{even} and x_{odd} for other automatic sequences.

Remark 1: x is automatic $\Leftrightarrow$ x_{even} and x_{odd} are both automatic.
Remark 2: x is Toeplitz $\Leftrightarrow$ x_{even} and x_{odd} are both Toeplitz.

Perfect Shuffle of Two 2-Automatic Words

For example, we have:

- ruler: $r_2 = 0^\omega \sqcup\!\sqcup (r_2 + 1)$
  0 **1** 0 **2** 0 **1** 0 **3** 0 **1** 0 **2** 0 **1** 0 **4** 0 **1** 0 **2** 0 **1** 0 **3** 0 **1** 0 **2** 0 **1** 0 **5**
- period-doubling: $d = 0^\omega \sqcup\!\sqcup (1 - d)$
  0 **1** 0 **0** 0 **1** 0 **1** 0 **1** 0 **0** 0 **1** 0 **0** 0 **1** 0 **0** 0 **1** 0 **1** 0 **1** 0 **0** 0 **1** 0 **0**
- regular paperfolding: $p = (01)^\omega \sqcup\!\sqcup\, p$
  0 **0** 1 **0** 0 **1** 1 **0** 0 **0** 1 **1** 0 **1** 1 **0** 0 **0** 1 **0** 0 **1** 1 **1** 0 **0** 1 **1** 0 **1** 1 **0**
- alternate paperfolding: $a = (01)^\omega \sqcup\!\sqcup (1 - a)$
  0 **1** 1 **0** 0 **0** 1 **1** 0 **1** 1 **1** 0 **0** 1 **0** 0 **1** 1 **0** 0 **0** 1 **0** 0 **1** 1 **1** 0 **0** 1 **1**

Remark: in all these cases, the equation yields a self-constructing rule.

The Hall Word

Let's now look at the Hall word

$$h = 2\mathbf{1}0\mathbf{2}0\mathbf{1}2\mathbf{1}0\mathbf{1}2\mathbf{0}2\mathbf{1}0\mathbf{2}0\mathbf{1}2\mathbf{0}2\mathbf{1}0\mathbf{1}2\mathbf{1}0\mathbf{2}0\mathbf{1}2\mathbf{1}0\mathbf{1}2\mathbf{0}\cdots$$

We have

$$h = (2 \times (1 - t)) ⧢ x$$

where

$$x = 1^\omega ⧢ h$$

$x \mod 2 = 1 - d$ (actually, replacing every other 0 by 2 in $1 - d$ one obtains x), where d is the period-doubling word.

A Companion of the Rudin-Shapiro Word

Consider the morphism: $0 \mapsto 01$, $1 \mapsto 02$, $2 \mapsto 31$, $3 \mapsto 32$.

Its fixed point starting with 0 is the word:

$$0102013101023202010201313231013101020131010232023231\cdots$$

Applying the coding $0, 1 \mapsto 0$; $2, 3 \mapsto 1$ one gets the Rudin-Shapiro word

$$rs = 0001001000011101000100101110001000010010000111011110\cdots$$

while applying the coding $0, 2 \mapsto 0$; $1, 3 \mapsto 1$ one gets the word

$$rs' = 0100011101001000010001111011011101000111010010001011\cdots$$

The sequence rs' is not (yet!) in OEIS...

The Word rs' is not in OEIS

The OEIS Foundation is supported by donations from users of the OEIS and by a grant from the Simons Foundation.

THE ON-LINE ENCYCLOPEDIA OF INTEGER SEQUENCES ®

founded in 1964 by N. J. A. Sloane

0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1 Search Hints

(Greetings from The On-Line Encyclopedia of Integer Sequences!)

Search: **seq:0,1,0,0,0,1,1,1,0,1,0,0,1,0,0,0,0,1,0,0,0,1,1,1,1,0,1,1,0,1,1**

Sorry, but the terms do not match anything in the table.

If your sequence is of general interest, please submit it using the form provided and it will (probably) be added to the OEIS! Include a brief description and if possible enough terms to fill 3 lines on the screen. We need at least 4 terms.

Search completed in 0.000 seconds

Lookup | Welcome | Wiki | Register | Music | Plot 2 | Demos | Index | Browse | More | WebCam
Contribute new seq. or comment | Format | Style Sheet | Transforms | Superseeker | Recent
The OEIS Community | Maintained by The OEIS Foundation Inc.

License Agreements, Terms of Use, Privacy Policy. .

Last modified October 23 10:13 EDT 2021. Contains 348211 sequences. (Running on oeis4.)

A Companion of the Rudin-Shapiro Sequence

$$rs = 0\mathbf{0}0\mathbf{1}0\mathbf{0}1\mathbf{0}0\mathbf{0}0\mathbf{1}1\mathbf{1}0\mathbf{1}0\mathbf{0}0\mathbf{1}0\mathbf{0}1\mathbf{0}1\mathbf{1}1\mathbf{0}0\mathbf{0}1\mathbf{0}0\mathbf{0}0\mathbf{1}0\mathbf{0}1\mathbf{0}0\mathbf{0}\cdots$$

We have

$$rs = rs \sqcup\!\sqcup\, rs'$$

and

$$rs' = rs \sqcup\!\sqcup\, (1 - rs)$$

Remark: rs' is the sequence of the parity of 11's in the binary expansions of odd integers.

3-automatic Words

Let's now look at some 3-automatic words.

Some 3-automatic Words

n	$ln2$ last non-2 digit	Mephisto-Waltz odd $\#2$'s in n	Stewart choral n ends in 21*
000	0	0	0
001	1	0	0
002	0	1	1
010	0	0	0
011	1	0	0
012	1	1	1
020	0	1	0
021	1	1	1
022	0	0	1
100	0	0	0
101	1	0	0
102	0	1	1
110	0	0	0
111	1	0	0
112	1	1	1
120	0	1	0
121	1	1	1

Some 3-automatic Words

word	symbol	substitution	pattern
Last Non-2	$ln2$	$0 \mapsto 010, 1 \mapsto 011$	$01?$
Stewart choral	st	$0 \mapsto 001, 1 \mapsto 011$	$0?1$
Mephisto-Waltz	mw	$0 \mapsto 001, 1 \mapsto 110$	NO
Sierpiński gasket	sg	$0 \mapsto 011, 1 \mapsto 010$	$01?, 10?$
Noname (A189706)	nn	$0 \mapsto 011, 1 \mapsto 001$	$0?1, 1?0$

sg can be obtained from mw by taking the sums modulo 2 of two consecutive elements (the same way $1 - d$ can be obtained from t).

mw avoids $(3 + \varepsilon)$-powers but contains arbitrarily long cubes.

$ln2$, st, sg and nn avoid cubes (and factors of the form $xxyyxx$) but contain $(3 - \varepsilon)$-powers; they all have factor complexity $2n$ and the same set of palindromic factors, which has cardinality 15.

Some 3-automatic Words

The word

$$ln2 = 010011010010011011010011 0\cdots$$

is the fixed point of $0 \mapsto 010, 1 \mapsto 011$.

The fixed point of its conjugate morphism $0 \mapsto 001, 1 \mapsto 101$ is the word $0 \cdot ln2$, which can be *almost* generated as a Toeplitz word taking the constant sequence $?01$, provided that one replaces the first $?$ by 0 in the limit word.

Ferenczi in 1995 observed that applying the morphism $0 \mapsto 0, 1 \mapsto 10$ to the word $0 \cdot ln2$, one gets the Chacon word

$$c = 0010001010010001000101001010010001 01\cdots$$

fixed point of the (nonprimitive) morphism $0 \mapsto 0010, 1 \mapsto 1$.

Why We Only Consider $ln2$

Definition

lnD (Last-Non-D), $D = 0, 1, 2$, is the sequence of the last digit different form D in the ternary representation of n.

$ln2_{(n \geq 0)} = 0100110100100110110100110100100110100100110110 \cdots$
fixed point of $0 \mapsto 010, 1 \mapsto 011$ and generated by the pattern $01?$

$ln0_{(n \geq 1)} = 1211221211211221221211221211211221211211221221 \cdots$
fixed point of $1 \mapsto 121, 2 \mapsto 122$ and generated by the pattern $12?$

Thus, $ln0 = 1 + ln2$.

$ln1_{(n \geq 0)} = 00200202200200202200202202200200202200200202200 \cdots$
fixed point of $0 \mapsto 002, 2 \mapsto 022$ and generated by the pattern $0?2$

Thus, $ln1 = 2 \times st$.

Perfect Shuffle of Two Words

Let us consider the Mephisto-Waltz word, fixed point of $0 \mapsto 001, 1 \mapsto 110$:

$$mw = 001001110001001110110110001001001110001001110 11\cdots$$

In 2005, Rampersad, Shallit and Wang described mw_{even} and mw_{odd}.

Let

$$g = 0123020313210120310123213023210313020123\cdots$$

be the (squarefree) fixed point $0 \mapsto 012, 1 \mapsto 302, 2 \mapsto 031, 3 \mapsto 321$.

Let $\mu_1 : 0 \mapsto 010, 1 \mapsto 100, 2 \mapsto 011, 3 \mapsto 101$.
Let $\mu_2 : 0 \mapsto 001, 1 \mapsto 101, 2 \mapsto 010, 3 \mapsto 110$.

$$mw_{even} = \mu_1(g) = 010100011101010011010101100101011100 0101\cdots$$

$$mw_{odd} = \mu_2(g) = 0011010101100010100011101011100101010011\cdots$$

Rampersad, Shallit and Wang proved that mw_{even} and mw_{odd} avoid large squares (while mw does not).

Perfect Shuffle of Two Words

For the word

$$ln2 = 0\mathbf{1}0\mathbf{0}1\mathbf{1}0\mathbf{1}0\mathbf{0}1\mathbf{0}0\mathbf{1}1\mathbf{0}1\mathbf{1}0\mathbf{1}0\mathbf{0}1\mathbf{1}0\mathbf{1}0\mathbf{0}1\mathbf{0}0\mathbf{1}1\mathbf{0}1\mathbf{0}0\mathbf{1}0\mathbf{0}1\mathbf{1}\cdots$$

we have

$$ln2 = st \sqcup\!\sqcup (1 - ln2)$$

For the Sierpiński gasket word

$$sg = 0\mathbf{1}1\mathbf{0}1\mathbf{0}0\mathbf{1}0\mathbf{0}1\mathbf{1}0\mathbf{1}0\mathbf{0}1\mathbf{1}0\mathbf{1}1\mathbf{0}1\mathbf{0}0\mathbf{1}1\mathbf{0}1\mathbf{1}0\mathbf{1}0\mathbf{0}1\mathbf{0}0\mathbf{1}1\mathbf{0}1\mathbf{0}\cdots$$

we have

$$sg = nn \sqcup\!\sqcup (1 - sg)$$

Perfect Shuffle of Two Words

For the Stewart choral word st, we have that:

- st_{even} is the Toeplitz word obtained alternating $01?$ and $?01$, and it is the fixed point of $0 \mapsto 010010011, 1 \mapsto 011010011$;
- st_{odd} is the Toeplitz word obtained alternating $?01$ and $01?$, and it is the fixed point of $0 \mapsto 001101001, 1 \mapsto 001101101$;

For the word nn, we have that:

- nn_{even} is the Toeplitz word obtained alternating $01?$ and $?10$, and it is the fixed point of $0 \mapsto 010011010, 1 \mapsto 011011010$;
- nn_{odd} is the Toeplitz word obtained alternating $?01$ and $10?$, and it is the complement of the fixed point of $0 \mapsto 010110010, 1 \mapsto 010110110$.

Open Problems

Some questions/further directions:

- Can we characterize the intersection of the sets of automatic words and Toeplitz words?
- What other classes of words are closed under perfect shuffle? (for example, the even part of the Fibonacci word, A339051, starts with 0011, hence it's not Sturmian)
- How much can differ the factor complexities of three words x, y, z satisfying the equation $x = y \sqcup\!\sqcup z$?
- The class of Toeplitz words generated by a sequence of patterns in the set $\{abc \mid \{a, b, c\} = \{0, 1, ?\}\}$ seems very interesting (critical exponent 3, factor complexity $2n$, 15 palindromic factors, etc.)

Thank you