The number of binary rotation words

A. Frid[*] and D. Jamet [†]

June 15, 2012

Abstract

We consider binary rotation words generated by partitions of the unit circle to two intervals and give a precise formula for the number of such words of length n. We also give the precise asymptotics for it, which happens to be $O(n^4)$. The result continues the line initiated by the formula for the number of all Sturmian words obtained by Lipatov in 1982, then independently by Berenstein, Kanal, Lavine and Olson in 1987, Mignosi in 1991, and then with another technique by Berstel and Pocchiola in 1993.

1 Introduction

Infinite words arising from rotations of the circle belong to the same family of infinite words defined by the means of dynamical systems as Sturmian words and interval exchange words in general. They were considered by G. Rote in 1992 [11] and can be defined using three parameters $\alpha, \beta, \gamma \in [0, 1)$ as $r = r_0 r_1 \cdots$, where for all i we have

$$r_i = \begin{cases} 1, & \text{if } \{i\alpha\} \in [\beta, \gamma), \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

(Here the interval $[\beta, \gamma)$ is denoted as usual if $\beta < \gamma$ and as $[\beta; 1) \cup [0, \gamma)$ otherwise.)

In the particular case when $\gamma - \beta = \alpha \pmod 1$, w is a *Sturmian* word. The family of Sturmian words is very well studied (see Chapter 2 of [9]); in particular, the total number of factors of all Sturmian words taken together is known to be

$$1 + \sum_{p=1}^{n} (n - p + 1)\varphi(p),$$

where φ is the Euler's totient function. This formula was rediscovered several times [8, 2, 10, 3]; the order of growth of this function is $\Theta(n^3/\pi^2)$.

In [6, 7] Cassaigne and the first author estimated and for some cases found the number of factors of length n of all rotation words with a given length $\gamma - \beta$ of the interval; it happens that it also grows as $O(n^3)$. In [1], Ambrož, Masáková, Pelantová and the first author estimated the number of all words arising from three-interval exchange, which continues the same line since Sturmian words are exactly two-interval exchange words; it happens that the number of three-interval exchange words grows as $O(n^4)$. In [5], Berstel and Vuillon coded rotation words by Sturmian words.

[*]Corresponding author, Sobolev Institute of Mathematics SB RAS and Université de Lorraine; supported in part by Presidential grant MK-4075.2012.1 and by RFBR grant 12-01-00089.

[†]Loria and Université de Lorraine.

At the beginning of the work on the project we present here, we hoped to estimate the number of all rotation words. However, we managed to do more, namely, to find a precise formula for it, predictably involving sums of the Euler's function. To write down the formula, we had to understand very clearly the structure of the set of rotation words, so in this abstract we not only give the formula, but also state some of the internal statements used to derive it.

2 Main statement

Theorem 1 *Starting from $n = 3$, the number of binary rotation words of length $n + 1$ is*

$$f(n+1) = n^2 + 3n + 4 + \frac{1}{2}\sum_{p=3}^{n} \varphi(p)(n^2 - p^2 + n + p) - f_1(n) - 2\sum_{l=2}^{n-1} f_2(n,l), \qquad (2)$$

where

$$f_1(n) = \begin{cases} 2\sum_{i=k}^{2k}\sum_{p=1}^{i+1}\varphi(p), & \text{if } n = 2k+1, \\ 2\sum_{i=k}^{2k-1}\sum_{p=1}^{i+1}\varphi(p) + \sum_{p=1}^{k}\varphi(p), & \text{if } n = 2k, \end{cases} \qquad (3)$$

$$g(n,l) = n - l + 1 + (n \bmod (l+1)), \qquad (4)$$

$$h(n,l) = \min(l+1, n-l),$$

and

$$f_2(n,l) = \left(\frac{1}{2}\left\lfloor\frac{n}{l+1}\right\rfloor g(n,l) - h(n,l)\right)(\varphi(l+1) - 1) + h(n,l)\left(\frac{\varphi(l+1)}{2} - 1\right). \qquad (5)$$

Note that the only addend of this formula growing faster than than $O(n^3)$ is the sum

$$\sum_{p=3}^{n} \varphi(p)(n^2 - p^2).$$

So, the asymptotics of the number of binary rotation words is equal to the asymptotics of this addend, which means that

$$f(n) = \frac{3n^4}{4\pi^2} + O(n^3 \log n).$$

The values of $f(n)$ for some values of n are shown in the table below.

n	6	7	10	15	20	30	40	50	75	100
$f(n)$	64	112	504	2804	9442	51306	168964	423814	2222984	7155096
$\frac{4\pi^2 f(n)}{3n^4} \approx$	0.65	0.61	0.66	0.73	0.78	0.83	0.87	0.89	0.92	0.94

3 Sketch of the proof

Clearly the bulky formula above reflects some structure of the set of rotation words. Here are some statements explaining this structure.

Denote the prefix of length $n + 1$ of the word r defined in (1) by $r(\alpha, \beta, \gamma, n)$. The parameter α is called the *slope* of the rotation word r.

Lemma 1 *It is sufficient to consider rotation words of slopes not greater than 1/2:*

$$f(n) = \#\{r(\alpha, \beta, \gamma, n-1) | \alpha \in (0, 1/2), \beta, \gamma \in \mathbb{R}/\mathbb{Z}.\}$$

Proof. Due to the symmetry, we have $r(\alpha, \beta, \gamma, n) = r(1-\alpha, 1-\beta, 1-\gamma, n)$ if $\{k\alpha\} \neq \beta$ or γ, that is, if the point $k\alpha$ is never equal to the end of an interval. But if it is, we just could slightly shift the interval to avoid it. So, slopes less than 1/2 and greater than 1/2 give exactly the same set of all rotation words. □

The following lemma is a particular case of the result of Berstel and Vuillon [5].

Lemma 2 *For any two-interval rotation word $r = r(\alpha, \beta, \gamma, n)$, where $\alpha \leq 1/2$, we have*

$$r_k = r_{k-1} + u_k - v_k \tag{6}$$

for some Sturmian words u and v of the slope α.

Hint for the proof. We take $u = r(\alpha, \beta, \beta + \alpha, n)$ and $v = r(\alpha, \gamma, \gamma + \alpha, n)$. □

Note that the symbols u_0 and v_0 are not used in the previous lemma, so, we see that a rotation word of length $n+1$ is uniquely defined by its first symbol and two Sturmian words of the same slope of length n. If these two Sturmian words are distinct, we can uniquely reconstruct from them the symbol r_0; if they are equal, both rotation words 0^{n+1} and 1^{n+1} can appear. It is clear also that each pair of Sturmian words of the same slope gives some rotation word (of that slope). This gives us the next lemma:

Lemma 3 *The number of binary rotation words is bounded as*

$$f(n+1) \leq \#\{(u, v) | u, v \in S(n, \alpha), \alpha \in (0, 1/2), u \neq v\} + 2. \quad \square \tag{7}$$

Here $S(n, \alpha)$ is the set of all Sturmian words of length n and of slope α.

The number of such pairs of Sturmian words can be calculated due to a result by Berstel and Pocchiola [4] and is equal to

$$n(n+1) + \frac{1}{2} \sum_{p=3}^{n} \varphi(p)(n^2 - p^2 + n + p).$$

This is already the main term of the formula (2). However, to pass to a precise formula, we should classify the cases when different pairs of Sturmian words give the same rotation word.

We start from the following

Lemma 4 *If a rotation word r appears from two different pairs of Sturmian words, then either 0s or 1s in r are isolated, that is, r does not contain the factor 11 or the factor 00.* □

The words 0^{n+1} and 1^{n+1} have been already excluded from consideration; so, in what follows we consider two cases: either $r = 0^i 1 0^{n-i}$ (or, symmetrically, $r = 1^i 0 1^{n-i}$) for some i; or r contains at least two 0s and two 1s.

Let us start from rotation words of the form $0^i 1 0^{n-i}$. Clearly, if $0 < i < n$, this word is generated by *all* pairs of Sturmian words of the form $(s10t, s01t)$ with $s10t, s01t \in S(n, \alpha)$ for some slope α. Here $|s| = i-1, |t| = n-i-1$. In particular, s is a right special word in $S(i-1, \alpha)$, and t is a left special word in $S(n-i-1, \alpha)$: here, as usual, a special word is a word which can be extended in the specified direction by two different letters. As it is well-known, the shorter of the words s and t is the suffix (prefix) of the mirror image of the longer one, so, the longer word determines it.

The number of special Sturmian words of length m is known to be $s(m) = \sum_{p=1}^{m+1} \varphi(p)$. Summing up these values (and adding those for the extreme cases of $i = 0$ and $i = n$), we get exactly $f_1(n)$ (see (3)) pairs of Sturmian words generating the $n + 1$ word of the given form.

From one hand, only a half of these pairs correspond to slopes less than 1/2; from the other hand, we should symmetrically consider the pairs generating the words $r = 1^i 0 1^{n-i}$. As a result, we see that $f_1(n)$ pairs generate $2(n + 1)$ rotation words, and to take them into account, we should subtract from the upper bound for $f(n + 1)$ the value $f_1(n) - 2(n + 1)$.

Most of technical details of our result are hidden in the following

Theorem 2 *Suppose a rotation word w is generated by two different pairs of Sturmian words of slope less than 1/2, and w contains at least two 1s and at least two 0s. Then $w = 0^i(10^l)^k 10^j$ or $w = 1^i(01^l)^k 01^j$ for some $i, j \geq 0$, $l \geq 2$, $k \geq 1$, and the number of pairs generating w is equal to $\varphi(l + 1)/2$ if $i, j \leq l$ and $\varphi(l + 1)$ otherwise.*

Its proof is based on the theory of *standard* Sturmian words and their construction with directive sequences (see Subsection 2.2.2 of [9]). The function $f_2(n, l)$ corresponds to the contribution of all words of the form $w = 0^i(10^l)^k 10^j$ for a given l to the difference between the upper bound and the precise value of the function f. After we multiply it by 2 to take into account the symmetric words $w = 1^i(01^l)^k 01^j$, and subtract from the upper bound their values for all l, we get exactly the formula (2). $\square$.

4 Acknowledgement

The authors are grateful to the participants of the working group in Saint-Dié des Vosges in April 2012 for stimulating discussions.

References

[1] P. Ambrož, A. Frid, Z. Masáková, E. Pelantová, *On the number of factors in codings of three interval exchange*, Discr. Math. Theoret. Comput. Sci. **13** (2011), 51–66.

[2] C. A. Berenstein and L. N. Kanal, D. Lavine and E. C. Olson, *A geometric approach to subpixel registration accuracy.* Comput. Vision Graph. **40** (1987), 334–360.

[3] J. Berstel, M. Pocchiola, *A geometric proof of the enumeration formula for Sturmian words.* Internat. J. Algebra Comput. **3** (1993), 349–355.

[4] J. Berstel, M. Pocchiola, *Random generation of finite Sturmian words.* Discr. Math. **153** (1996), 29–39.

[5] J. Berstel, L. Vuillon, *Coding rotations on intervals.* Theoret. Comput. Sci. **281** (2002), 99–107.

[6] J. Cassaigne, A.E. Frid, *On the arithmetical complexity of Sturmian words.* Theoret. Comput. Sci. **380** (2007) 304–316.

[7] A. Frid, *A lower bound for the arithmetical complexity of Sturmian words*, Siberian Electron. Math. Rep. **2**, 14–22 (in Russian, English abstract).

[8] E. P. Lipatov, *A classification of binary collections and properties of homogeneity classes.* Problemy Kibernet. **39** (1982), 67–84 (in Russian).

[9] M. Lothaire, *Algebraic Combinatorics on Words*, Cambridge University Press, Cambridge, (2002).

[10] F. Mignosi, *On the number of factors of Sturmian words.* Theoret. Comput. Sci. **82** (1991), 71–84.

[11] G. Rote, *Sequences with subword complexity* $2n$, J. Number Theory **46** (1994) 196–213.