Discrete Mathematics and Theoretical Computer Science DMTCS vol. **13:3**, 2011, 51–66

On the number of factors in codings of three interval exchange

Petr Ambrož[1] Anna Frid[2] Zuzana Masáková[1] Edita Pelantová[1]

[1]*Doppler Institute & Department of Mathematics, FNSPE, Czech Technical University in Prague, Czech Republic*
[2]*Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia*

received 24th April 2009, revised 27th December 2010, accepted 2nd June 2011.

We consider exchange of three intervals with permutation $(3, 2, 1)$. The aim of this paper is to count the cardinality of the set $3\mathrm{iet}(N)$ of all words of length N which appear as factors in infinite words coding such transformations. We use the strong relation of 3iet words and words coding exchange of two intervals, i.e., Sturmian words. The known asymptotic formula $\#2\mathrm{iet}(N)/N^3 \sim \frac{1}{\pi^2}$ for the number of Sturmian factors allows us to find bounds $\frac{1}{3\pi^2} + o(1) \leq \#3\mathrm{iet}(N)/N^4 \leq \frac{2}{\pi^2} + o(1)$.

Keywords: interval exchange, enumeration of factors, Sturmian words

1 Introduction

The study of infinite words arising from an exchange of several intervals was initiated by Rauzy, whereas dynamical systems connected with exchange of intervals has been already studied by Katok and Stepin [15]. The most explored subfamily of such words are Sturmian words corresponding to an exchange of two intervals [22]. Infinite words coding exchange of 3 intervals form the next interesting family.

The special case of Sturmian words is very well understood. For surveys on the properties of Sturmian words see [18, 22].

Much less is known about codings of r intervals for $r > 2$. Combinatorial characterization of the language of such words was given first for $r = 3$ in [11], then for general r in [6]. Boshernitzan and Caroll [8] have found a sufficient condition for substitutivity of an infinite word coding an exchange of r intervals. The necessity of this condition in the special case $r = 3$ and reverse-order permutation was demonstrated in [1]. A criterion for substitution invariance of words coding exchange of 3 intervals with permutation (3,2,1) is given in [3] and [4].

Actually, infinite words coding exchange of 3 intervals with permutation (3,2,1), here called 3iet words, constitute the central theme of this paper. They stand out among interval exchange words by the fact that, similarly to Sturmian words, they can be geometrically represented by cut-and-project sequences. This was first remarked in [16], then developed in [12]. Results of these papers are connected to the well known 3-distance theorem [23, 2], and can be also explained in terms of the theorem of Katok [14] which states that when a k-interval exchange transformation is induced on a subinterval, the resulting transformation is a k or $k + 1$ interval exchange.

An even more explicit connection between 3iet words and Sturmian words is displayed in [3]. This connection allowed us to estimate the number of factors of 3iet words using the following enumeration formula for Sturmian words, first given by Lipatov [17]; see also [19, 7].

Theorem 1 *[17, 19, 7] The number $\#2\mathrm{iet}(N)$ of Sturmian factors of length N is given by*

$$\#2\mathrm{iet}(N) = 1 + \sum_{k=1}^{N}(N+1-k)\varphi(k) = \frac{1}{\pi^2}N^3 + \mathcal{O}(N^2\log N)\,,$$

The aim of this paper is to show the following theorem.

Theorem 2 *The asymptotic growth of the number of 3iet factors of length N is given by*

$$\frac{1}{3\pi^2}N^4 + \mathcal{O}(N^{3+\delta}) \le \#3\mathrm{iet}(N) \le \frac{2}{\pi^2}N^4 + \mathcal{O}(N^{3+\delta})\,,$$

for arbitrary $\delta > 0$.

The proof of Theorem 2 is divided into two parts, the upper bound is given in Theorem 9 in Section 5, the lower bound in Theorem 12 in Section 6. The lemmas on Sturmian words used in these sections may be of independent interest. Sections 3 and 4 describing properties of the Euler function and of Sturmian factors are auxiliary.

Precise values of $\#3\mathrm{iet}(N)$ for $N \le 10$ are found in Section 7.

2 Definitions of interval exchange words

Let $\mathcal{A}$ be a finite alphabet. A concatenation $w = w_0w_1\cdots w_{n-1}$ of letters $w_i \in \mathcal{A}$ is called a word of length $|w| = n$. The number of occurrences of a letter a in the word w is denoted by $|w|_a$. The set of all finite words over $\mathcal{A}$ together with the empty word is denoted by $\mathcal{A}^*$. We can also consider one- or two-sided infinite words $u = u_0u_1\cdots \in \mathcal{A}^{\mathbb{N}}$, resp. $u = \cdots u_{-2}u_{-1}u_0u_1u_2\cdots \in \mathcal{A}^{\mathbb{Z}}$. A finite word w is a factor of an infinite word $u = (u_n)$ if $w = u_iu_{i+1}\cdots u_{i+n-1}$ for some i. The factors of an infinite word u form the language of u, which is denoted by $\mathcal{L}(u)$. We denote $\mathcal{L}_n(u) = \mathcal{L}(u)\cap\mathcal{A}^n$. The factor complexity of u is the function $\mathcal{C}_u : \mathbb{N} \mapsto \mathbb{N}$ given by $\mathcal{C}_u(n) = \#\mathcal{L}_n(u)$.

Sturmian words are aperiodic words with the smallest possible complexity, namely $\mathcal{C}_u(n) = n+1$ for all $n \in \mathbb{N}$. It is known [20] that Sturmian words on $\{0,1\}$ are exactly the lower and upper mechanical words, which means that each Sturmian word $u = (u_n)_{n\in\mathbb{Z}} \in \{0,1\}^{\mathbb{Z}}$ is defined by

$$\begin{aligned} u_n &= \lfloor (n+1)\alpha+\beta\rfloor - \lfloor n\alpha+\beta\rfloor \quad \text{or}\\ u_n &= \lceil (n+1)\alpha+\beta\rceil - \lceil n\alpha+\beta\rceil\,, \end{aligned}$$

where $\alpha \in (0,1)$ is irrational and $\beta \in [0,1)$. Another representation of Sturmian words is given by the two interval exchange. Define $\varepsilon = 1-\alpha \in (0,1)$, where α is the parameter used above. Consider the intervals $I_0 = [0,\varepsilon)$, $I_1 = [\varepsilon,1)$, $I_0\cup I_1 = [0,1)$, and the corresponding transformation $T_\varepsilon : [0,1) \mapsto [0,1)$ defined by

$$T_\varepsilon(x) = \begin{cases} x+1-\varepsilon & \text{for } x \in I_0\,,\\ x-\varepsilon & \text{for } x \in I_1\,. \end{cases} \tag{1}$$

We can code the orbit of a point $x_0 \in [0,1)$ by the infinite word $u_{\varepsilon,x_0} = (u_n)_{n\in\mathbb{Z}} \in \{0,1\}^{\mathbb{Z}}$:

$$u_n = \begin{cases} 0 & \text{if } T_\varepsilon^n(x_0) \in I_0\,, \\ 1 & \text{if } T_\varepsilon^n(x_0) \in I_1\,. \end{cases} \tag{2}$$

The word u_{ε,x_0} is the upper mechanical word corresponding to $\alpha = 1-\varepsilon$ and $\beta = x_0$. Similarly, lower mechanical words are obtained by coding exchange of intervals of the type $(\cdot,\cdot]$. It is the representation of Sturmian words by two interval exchange which will be the most used here.

Since the language of a Sturmian word does not depend on the initial point x_0, but rather only on the parameter ε, we can denote the language of u_{ε,x_0} by $\mathcal{L}(\varepsilon)$ and the set of factors of length M of u_{ε,x_0} by $\mathcal{L}_M(\varepsilon)$.

We study infinite words which generalize Sturmian words in the way that they code exchange of three intervals. Let ε, ℓ be real numbers satisfying

$$\varepsilon \in (0,1)\setminus\mathbb{Q}\,, \qquad \max\{\varepsilon, 1-\varepsilon\} < \ell < 1\,. \tag{3}$$

The mapping $T_{\varepsilon,\ell}(x) : [0,\ell) \mapsto [0,\ell)$ defined by

$$T_{\varepsilon,\ell}(x) = \begin{cases} x+1-\varepsilon & \text{for } x \in I_A := [0, \ell-1+\varepsilon)\,, \\ x+1-2\varepsilon & \text{for } x \in I_B := [\ell-1+\varepsilon, \varepsilon)\,, \\ x-\varepsilon & \text{for } x \in I_C := [\varepsilon, \ell)\,, \end{cases} \tag{4}$$

is called exchange of three intervals[(i)] with permutation $(3,2,1)$. The orbit of a point $x_0 \in [0,\ell)$ under the transformation $T_{\varepsilon,\ell}$ of (4) can be coded by the infinite word $u_{\varepsilon,\ell,x_0} = (u_n)_{n\in\mathbb{Z}}$ in the alphabet $\{A,B,C\}$, where

$$u_n = \begin{cases} A & \text{if } T_{\varepsilon,\ell}^n(x_0) \in I_A\,, \\ B & \text{if } T_{\varepsilon,\ell}^n(x_0) \in I_B\,, \\ C & \text{if } T_{\varepsilon,\ell}^n(x_0) \in I_C\,. \end{cases} \tag{5}$$

A visual representation of a 3 interval exchange transformation can be found, e. g., in [1].

Since ε is irrational, the infinite word u is aperiodic. We call these words 3iet words. Again, the language of a 3iet word does not depend on the initial point x_0, thus we denote it by $\mathcal{L}(\varepsilon,\ell)$. The factor complexity of 3iet words depends on the parameter ℓ in the following way: $\mathcal{C}(n) = 2n+1$, if $\ell \notin \mathbb{Z}+\varepsilon\mathbb{Z}$, otherwise it satisfies $\mathcal{C}(n) = n + C$ for some constant C and sufficiently large n, see [1]. Words with the latter complexity are called quasisturmian and described for example in [9, 10].

3 Asymptotic behaviour of the totient function

In determining the asymptotic behaviour of the number of 3iet factors, we shall strongly use asymptotic properties of the Euler totient function,

$$\varphi(n) = \#\{k \in \mathbb{N} \mid k \le n,\ k \perp n\} = n \prod_{\substack{p\mid n \\ p \text{ prime}}} \Big(1 - \frac{1}{p}\Big)\,,$$

(i) Usually one defines exchange of three intervals of lengths $\alpha, \beta, \gamma > 0$ satisfying $\alpha+\beta+\gamma = 1$. Our choice of parameters (3) implies relation $\alpha + 2\beta + \gamma = 1$, where $\varepsilon = \alpha+\beta$ and $\ell = \alpha+\beta+\gamma$.

where $k \perp n$ means $\gcd(k,n) = 1$. In our formulas, we will use the Landau big $\mathcal{O}$ notation; instead of a function $f(n)$, we write $\mathcal{O}(g(n))$ if there exists a constant K such that $|f(n)| \leq K|g(n)|$ for all $n \in \mathbb{N}$.

The proof of the first asymptotic formula for the totient function can be found for example in [21]:

$$\sum_{k=1}^{n} \varphi(k) = \frac{3}{\pi^2} n^2 + \mathcal{O}(n \log n)\,. \tag{6}$$

Using the first equality from Theorem 1 and this formula, one can easily derive the following estimate found already in [17]:

$$\#2\mathrm{iet}(N) = 1 + \sum_{k=1}^{N} (N+1-k)\varphi(k) = \frac{1}{\pi^2} N^3 + \mathcal{O}(N^2 \log N)\,. \tag{7}$$

Yet another formula involving the totient function will be useful. It can be easily derived that for any q, numbers coprime to q are in a certain sense uniformly distributed, namely

$$\#\{p \in \mathbb{N} \mid p \perp q,\ p \leq i\} = \frac{\varphi(q)}{q} i + \mathcal{O}(2^{\omega(q)})\,,$$

where $\omega(q)$ is the number of prime divisors of q. From the result of Hardy and Wright [13] on the asymptotic behavior of $\omega(q)$, it follows that

$$\#\{p \in \mathbb{N} \mid p \perp q,\ j \leq p \leq i\} = \frac{\varphi(q)}{q}(i-j) + \mathcal{O}(q^\delta)\,, \qquad \text{for any } \delta > 0. \tag{8}$$

4 Sturmian factors and Farey numbers

As we have already mentioned, for every fixed irrational $\varepsilon \in (0,1)$ and fixed $M \in \mathbb{N}$, the number of factors of length M in the language $\mathcal{L}(\varepsilon)$ of a Sturmian word is exactly $M+1$. Obviously, the set $\mathcal{L}_M(\varepsilon)$ does not determine ε uniquely: the same set of $M+1$ factors appears for uncountably many irrational ε which form an interval. As shown in [19], if for some irrational $\varepsilon_1 < \varepsilon_2$ it holds that $\mathcal{L}_M(\varepsilon_1) \neq \mathcal{L}_M(\varepsilon_2)$, then there exists a rational $\frac{p}{q}$ such that $\varepsilon_1 < \frac{p}{q} < \varepsilon_2$ and $q \leq M$. Therefore the interval $(0,1)$ of admissible parameters ε is partitioned into small intervals determining classes of Sturmian words with distinct $\mathcal{L}_M(\varepsilon)$. The partition is given by the Farey fractions of order M, i.e., all the reduced fractions in $[0,1]$ with denominator smaller than or equal to M. It is useful to consider the Farey fractions as an ordered list

$$\mathcal{F}_M: \quad f_0 = \frac{0}{1} < f_1 < \cdots < f_{r-1} < f_r = \frac{1}{1}\,.$$

The partition $[0,1] = \cup_{i=0}^{r}[f_i, f_{i+1}]$, where $\mathcal{F}_M = \{f_i \mid i = 0,\ldots,r\}$, is essential for describing Sturmian factors of length M. We shall thus speak about classes of Sturmian words of length M determined by intervals (f_i, f_{i+1}), $i = 0,\ldots,r-1$. Every class contains $M+1$ factors and of course, the same factor may belong to different classes. For $M = 4$, the division is illustrated in Figure 1 together with the corresponding sets of factors of length 4.

The possible numbers of letters 0 and 1 in factors corresponding to an interval (f_i, f_{i+1}) can be easily determined as follows.

$$0 \qquad \tfrac{1}{4} \qquad \tfrac{1}{3} \qquad \tfrac{1}{2} \qquad \tfrac{2}{3} \qquad \tfrac{3}{4} \qquad 1$$

0111	0110	0101	0010	0001	0000
1011	0111	0110	0100	0010	0001
1101	1011	1010	0101	0100	0010
1110	1101	1011	1001	1000	0100
1111	1110	1101	1010	1001	1000

Fig. 1: Classes of Sturmian words of length 4.

Lemma 3 *Let $M \in \mathbb{N}$ be fixed, and let f_i, f_{i+1} be two consecutive Farey fractions of order M. Let ε be an irrational in (f_i, f_{i+1}). Then for every factor w of length M of a Sturmian word with parameter ε we have*

$$|w|_0 \in \big\{\lfloor Mf_i\rfloor, \lfloor Mf_i\rfloor + 1\big\}\,, \qquad |w|_1 \in \big\{M - \lfloor Mf_i\rfloor, M - \lfloor Mf_i\rfloor - 1\big\}\,.$$

Proof: Denote $\alpha = 1 - \varepsilon$. If $w = w_0 w_1 \cdots w_{M-1}$ is a factor of a Sturmian word with parameter ε, there must exist $\beta \in [0, 1)$ such that

$$w_i = \big\lfloor (i+1)\alpha + \beta\big\rfloor - \big\lfloor i\alpha + \beta\big\rfloor \in \{0, 1\}\,, \qquad \text{for } i = 0, 1, \ldots, M-1\,.$$

Therefore

$$|w|_1 = \sum_{i=0}^{M-1} w_i = \lfloor M\alpha + \beta\rfloor \in \big\{\lfloor M\alpha\rfloor,\ \lfloor M\alpha\rfloor + 1\big\}\,.$$

The language $\mathcal{L}_M(\varepsilon)$ is the same for all $\varepsilon \in (f_i, f_{i+1})$ and we have $\lfloor M\alpha\rfloor = \big\lfloor M(1-\varepsilon)\big\rfloor = M - \lceil M\varepsilon\rceil = M - \lfloor M\varepsilon\rfloor - 1 = M - \lfloor Mf_i\rfloor - 1$. $\square$

Now consider words coding the transformations T_{f_i} defined as in (1). If $f_i = \frac{p}{q}$, then such a word is purely periodic and looks as $v^{(i)}v^{(i)}v^{(i)} \cdots = \big(v^{(i)}\big)^{\omega}$, where $v^{(i)}$ is a primitive word of length q. Recall that a finite word v is said to be primitive if it cannot be written in the form $v = ww\cdots w = w^k$ for any integer $k \geq 2$. Therefore the language of the periodic word coding the transformation T_{f_i} has exactly q factors of length M for every $M \geq q$. Let us denote the set of such factors by $L_M^{(i)}$. As it is derived in [19],

$$\mathcal{L}_M(\varepsilon) = L_M^{(i)} \cup L_M^{(i+1)}\,, \quad \text{for all } \varepsilon \in (f_i, f_{i+1})\,.$$

We have $\#\mathcal{L}_M(\varepsilon) = M + 1$, $\#L_M^{(i)} = q_i$, and $\#L_M^{(i+1)} = q_{i+1}$, where q_i, q_{i+1} are the denominators of Farey fractions f_i, f_{i+1}, respectively. Therefore we have

$$\#\Big(L_M^{(i)} \cap L_M^{(i+1)}\Big) = q_i + q_{i+1} - M - 1\,. \tag{9}$$

5 The upper bound

The main tool which we use in this section is a link between 3iet words and Sturmian words over the alphabet $\{0, 1\}$ by means of morphisms $\sigma_{01}, \sigma_{10} : \{A, B, C\}^* \mapsto \{0, 1\}^*$ defined by

$$\begin{aligned} \sigma_{01}: &\quad A \mapsto 0\,,\ B \mapsto 01\,,\ C \mapsto 1\,,\\ \sigma_{10}: &\quad A \mapsto 0\,,\ B \mapsto 10\,,\ C \mapsto 1\,. \end{aligned} \tag{10}$$

In [3], the following statement is proved.

Theorem 4 *A ternary word u is a 3iet word if and only if both $\sigma_{01}(u)$ and $\sigma_{10}(u)$ are Sturmian words.*

Based on this strong relation, we define the notion of b-amicability.

***Definition* 1** *Let $w^{(1)}, w^{(2)}$ be finite words over the alphabet $\{0,1\}$. We say that $w^{(1)}$, $w^{(2)}$ form a b-amicable pair, if there exists a 3iet factor w over $\{A,B,C\}$ with exactly b letters B such that $w^{(1)} = \sigma_{01}(w)$ and $w^{(2)} = \sigma_{10}(w)$.*

$$\begin{array}{r|c|c|c|c|c|c|}
\cline{2-7}
w^{(1)} = & 0 & 1 & 0 & 0\,1 & 0 & 1 \\
w^{(2)} = & 0 & 1 & 0 & 1\,0 & 0 & 1 \\
\cline{2-7}
\multicolumn{1}{r}{w =} & \multicolumn{1}{c}{A} & \multicolumn{1}{c}{C} & \multicolumn{1}{c}{A} & \multicolumn{1}{c}{B} & \multicolumn{1}{c}{A} & \multicolumn{1}{c}{C}
\end{array}$$

Fig. 2: Finite words $w^{(1)} = 0100101$ and $w^{(2)} = 0101001$ forming a 1-amicable pair. The corresponding ternary word is $w = ACABAC$.

The notion is illustrated in Figure 2. Note that due to Theorem 4, both words of a b-amicable pair are Sturmian factors from the same language $\mathcal{L}(\varepsilon)$ for some irrational $\varepsilon \in (0,1)$. A 3iet factor w of length N with precisely b letters B corresponds to a b-amicable pair of Sturmian factors of length $N+b$. Obviously, different b-amicable pairs give rise to different 3iet factors. Therefore we can express the number of 3iet words as the number of corresponding b-amicable pairs.

Proposition 5 *Let $N \in \mathbb{N}$. Then*

$$\#\mathrm{3iet}(N) = \sum_{b=0}^{N} \#\mathrm{3iet}(N,b) = \sum_{b=0}^{N} \#\ \textit{b-amicable pairs of length}\ (N+b)\,,$$

where $\mathrm{3iet}(N,b)$ denotes the set of 3iet factors of length N with precisely b letters B.

Let us focus on b-amicable pairs of Sturmian factors. First, it is obvious that words which form a b-amicable pair must both contain at least b letters 0 and b letters 1. As a consequence of Lemma 3, we have the following statement.

Corollary 6 *Let $b, M \in \mathbb{N}$ and let ε be an irrational in (f_i, f_{i+1}), where f_i, f_{i+1} are consecutive Farey fractions of order M. If $\mathcal{L}_M(\varepsilon)$ contains a b-amicable pair, then*

$$b \le \min\Big\{ \lfloor M f_i \rfloor + 1, M - \lfloor M f_i \rfloor \Big\}\,.$$

Lemma 7 *Let $Z_{M,b}$ be the family of classes of Sturmian words of length M containing at least one b-amicable pair. Then for $M < 2b$ we have $\#Z_{M,b} = 0$, and otherwise,*

$$\#Z_{M,b} \ \le \frac{3}{\pi^2}(M-2b)M + \mathcal{O}(M^{1+\delta})\,, \quad \textit{for arbitrary}\ \delta > 0\,.$$

Proof: Obviously, words forming a b-amicable pair must be of length at least $M \geq 2b$. Consider an interval $[f_i, f_{i+1})$ such that the language $\mathcal{L}_M(\varepsilon)$ for $\varepsilon \in (f_i, f_{i+1})$ contains an amicable pair and $f_i = \frac{p}{q}$, $p \perp q$. Corollary 6 then implies that b must satisfy

$$b \leq M\frac{p}{q} + 1 \qquad \text{and} \qquad b \leq M - M\frac{p}{q} + 1\,,$$

providing bounds on p, namely,

$$q\frac{b-1}{M} \leq p \leq q\left(1 - \frac{b-1}{M}\right)\,.$$

Thus, we are looking for the number

$$\sum_{q=1}^{M} \#\Big\{p \in \mathbb{N} \,\Big|\, p \perp q \text{ and } q\tfrac{b-1}{M} \leq p \leq q\big(1 - \tfrac{b-1}{M}\big)\Big\}$$

With the use of (8) we can write

$$\begin{aligned}\#Z_{M,b} &\leq \sum_{q=1}^{M} \frac{\varphi(q)}{q} q\Big(1 - \frac{2(b-1)}{M}\Big) + M\mathcal{O}\big(M^{\delta}\big) = \\ &= \frac{M-2b+2}{M} \sum_{q=1}^{M} \varphi(q) + \mathcal{O}(M^{1+\delta}) \ \text{ for all } \delta > 0.\end{aligned}$$

Together with the asymptotic behavior of $\sum_{q=1}^{M} \varphi(q)$ given in (6), this implies the statement of the lemma. □

Lemma 8 *Let $\varepsilon \in (0,1)$ be irrational. Let $b, M \in \mathbb{N}$, $2b \leq M$. Then the number of distinct b-amicable pairs in class $\mathcal{L}_M(\varepsilon)$ is at most $M - b$.*

Proof: The language $\mathcal{L}_M(\varepsilon)$ contains $M+1$ factors of length M. Choose such a factor w and consider the set of points $x \in [0,1)$ such that the orbit $\{T_\varepsilon^j(x)\}$, $j = 0, 1, \ldots, M-1$, is coded by w. It can be easily shown that this set is a subinterval of $[0,1)$ of the type $[\cdot,\cdot)$. Thus $[0,1)$ is divided into $M+1$ disjoint subintervals; the division is done by M division points $T_\varepsilon^{-j}(\varepsilon)$, $j = 0, 1, \ldots, M-1$. Let us denote these intervals J_i, $i = 1, 2, \ldots, M+1$, ordered so that $x < y$ for any $x \in J_i$ and any $y \in J_{i+1}$. The corresponding words of length M are denoted by $w^{(i)}$, $i = 1, 2, \ldots, M+1$. Note that $T^j(J_i)$ is an interval for all $j = 0, 1, \ldots, M-1$.

Let us first find an upper bound for the number of 1-amicable pairs of Sturmian factors of length M. The following procedure is illustrated in Figure 3. Take two consecutive intervals, J_i, J_{i+1} for some $i = 1, \ldots, M$. It means that $J_i \cup J_{i+1}$ is an interval which contains in its interior exactly one of the division points, say $T_\varepsilon^{-k}(\varepsilon)$. If $k = 0$, then the word $w^{(i)}$ starts with 0 and the word $w^{(i+1)}$ starts with 1. More generally, one can show that words $w^{(i)}$ and $w^{(i+1)}$ have a common prefix v of length k and the set $T_\varepsilon^k(J_i \cup J_{i+1})$ is still an interval. Since $T_\varepsilon^{-k}(\varepsilon) \in (J_i \cup J_{i+1})^\circ$, we have $\varepsilon \in T_\varepsilon^k\big((J_i \cup J_{i+1})^\circ\big)$, and therefore the word $w^{(i)}$ has prefix $v0$ and the word $w^{(i+1)}$ has prefix $v1$.

If $k = M-1$, then $w^{(i)} = v0$ and $w^{(i+1)} = v1$ and they are not 1-amicable. On the other hand, if $k < M-1$, then $T^{k+1}(J_i) = T^k(J_i)+1-\varepsilon$ is an interval of the form $[\cdot, 1)$, and $T^{k+1}(J_{i+1}) = T^k(J_{i+1})-\varepsilon$ is an interval of the form $[0,\cdot)$. We further have

$$\begin{aligned} T^{k+2}(J_i) &= T^{k+1}(J_i) - \varepsilon = T^k(J_i) + 1 - 2\varepsilon\,, \\ T^{k+2}(J_{i+1}) &= T^{k+1}(J_{i+1}) + 1 - \varepsilon = T^k(J_{i+1}) + 1 - 2\varepsilon\,. \end{aligned}$$

and therefore the set $T^j(J_i \cup J_{i+1})$ is again an interval for all j, $k+2 \le j \le M-1$, and thus the words $w^{(i)}$ and $w^{(i+1)}$ have a common suffix v' of length $M-k-2$. Therefore we have $w^{(i)} = v01v'$ and $w^{(i+1)} = v10v'$ and the words $w^{(i)}$ and $w^{(i+1)}$ form a 1-amicable pair. Together, we have a 1-amicable pair of words $w^{(i)}$ and $w^{(i+1)}$ for $i = 1,\ldots,M$, $i \neq k$. Their number is therefore $M-1$.

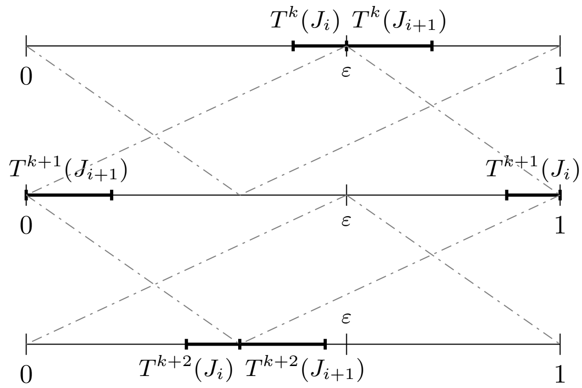


Fig. 3: Iterations of consecutive intervals J_i, J_{i+1} show that the corresponding words $w^{(i)}$, $w^{(i+1)}$ form a 1-amicable pair.

The previous considerations imply that the words $w^{(i)}$, $i = 1,\ldots,M+1$, are ordered lexicographically, where consecutive words differ only by interchange of $01 \leftrightarrow 10$ at one place. The only exception is the pair of words $w^{(i)}$ and $w^{(i+1)}$ whose intervals J_i, J_{i+1} are separated by the iteration $T_\varepsilon^{-M+1}(\varepsilon)$. These words are of the form $w^{(i)} = v0$ and $w^{(i+1)} = v1$, which corresponds to 'half' of the interchange $01 \leftrightarrow 10$.

From that, it is clear that b-amicable pairs can only be the pairs of words $w^{(i)}$, $w^{(i+b)}$, $i = 1,\ldots,M+1-b$, with exceptions, namely those indices i, such that the interval $J_i \cup \cdots \cup J_{i+b}$ contains in its interior the iteration $T_\varepsilon^{-M+1}(\varepsilon)$. However, since we do not know the position of this point, it is difficult to determine the number of these exceptions, but it is at least one. Therefore the number of b-amicable pairs is at most $M-b$. □

Now we are ready to prove the upper bound from Theorem 2.

Theorem 9 *The number of 3iet factors of length N satisfies*

$$\#3\mathrm{iet}(N) \le \frac{2}{\pi^2}N^4 + \mathcal{O}(N^{3+\delta})$$

for arbitrary $\delta > 0$.

Proof: Combining Proposition 5 and Lemma 8, we see that

$$\begin{aligned}
\#3\mathrm{iet}(N) &= \sum_{b=0}^{N} \#3\mathrm{iet}(N,b) = \sum_{b=0}^{N} \#\ b\text{-amicable pairs of length } (N+b) \leq \\
&\leq \sum_{b=0}^{N} \sum_{\mathcal{K}\in Z_{N+b,b}} \#\ b\text{-amicable pairs of length } (N+b) \text{ in the class } \mathcal{K} \leq \\
&\leq \sum_{b=0}^{N} N\#Z_{N+b,b}\,.
\end{aligned}$$

Putting in the estimate on $\#Z_{N+b,b}$ from Lemma 7, we obtain

$$\begin{aligned}
\#3\mathrm{iet}(N) &\leq N\sum_{b=0}^{N} \frac{3}{\pi^2}(N-b)(N+b) + N\sum_{b=0}^{N} \mathcal{O}\big((N+b)^{1+\delta}\big) = \\
&= \frac{3}{\pi^2}N\Big(N^2(N+1) - \frac{1}{6}N(N+1)(2N+1)\Big) + \mathcal{O}(N^{3+\delta}) = \\
&= \frac{2}{\pi^2}N^4 + \mathcal{O}(N^{3+\delta})\,.
\end{aligned}$$

□

6 The lower bound

Consider the following geometric representation of Sturmian and 3iet words by cut-and-project sequences. Given a strip in the Euclidean plane parallel to the (irrationally oriented) straight line $y = \varepsilon x$, take all points of the lattice $\mathbb{Z}^2$ and project them orthogonally to the straight line. It is known [16, 12] that in such a way, one obtains a sequence of points with two or three distances between them. If the distances are two, say Δ_1, Δ_2, we can code them by two letters, and the resulting word is a Sturmian word. If the distances are three, they are of the form Δ_1, Δ_2, $\Delta_1 + \Delta_2$. Coding these distances by three letters, we obtain a 3iet word. Moreover, Sturmian word is obtained only for a discrete set of values of the width of the considered strip. For details about this construction we refer to [12].

On the other hand, any Sturmian word u_{ε,x_0} or a 3iet word u_{ε,ℓ,x_0} can be represented in this way, the parameter ε being the slope of the straight line and ℓ corresponding to the width of the strip [12, 16]. If without loss of generality we take $\varepsilon \in (0,1)$, the widths of the strips giving rise to Sturmian words form a sequence

$$\cdots < \max\{\varepsilon, 1-\varepsilon\} < 1 < 1+\varepsilon < \cdots$$

All parameters ℓ in between these values yield 3iet words.

This geometric representation is illustrated in Figure 4, where $\ell_1 < \ell_2$ are two consecutive discrete values for which the cut-and-project scheme yields a Sturmian word. Projections of points in the strip S_1 with parameter ℓ_1 correspond to a Sturmian word $u^{(1)}$ over the alphabet $\{A, B\}$, where A stands for the projection of the horizontal side and B of the diagonal of a unit square. The strip S_2 corresponding to

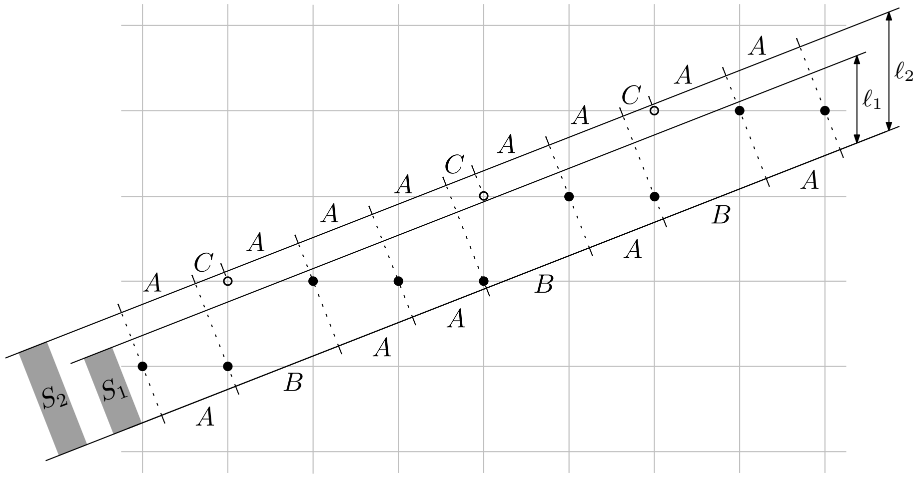


Fig. 4: Geometric representation of Sturmian and 3iet words by cut-and-project sequences.

the value ℓ_2 gives a Sturmian word $u^{(2)}$ over $\{A, C\}$, where again, A is the projection of the horizontal side and C of the vertical side of the unit square. Note that the projection of a diagonal is the sum of projections of the vertical and horizontal sides. This corresponds to the fact that the infinite word $u^{(2)}$ arises from $u^{(1)}$ by replacing every B by AC.

Any strip S such that $S_1 \subset S \subset S_2$ gives rise to a 3iet word over the alphabet $\{A, B, C\}$ where only some letters B have been replaced by AC. Enlarging the strip S from S_1 to S_2 causes more and more lattice points to fall into S and thus to split more and more distances B into CA. This procedure is illustrated in Figure 5 on a finite segment taken from Figure 4.

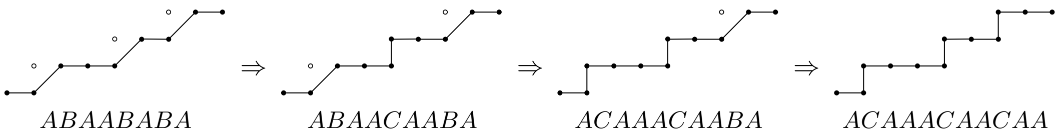


Fig. 5: Construction of 3iet factors with given number of letters C by substituting AC for suitable letters B in a Sturmian factor.

Let now $w^{(1)} \in \{A, B\}^*$ be a Sturmian factor of length M with at least b letters B, $b \leq M$. By a suitable choice of the enlarged strip S, we obtain from $w^{(1)}$ a 3iet factor w of length $M + b$ containing exactly b letters C. Such a word w satisfies $|w|_A \geq |w|_C$, where the equality holds only if $w = (CA)^M$. Since the role of letters A and C is symmetric over the set of all 3iet factors, the interchange $A \leftrightarrow C$ in the factor w gives again a 3iet factor. Denoting by $H_{M,b}$ the number of Sturmian factors of length M with at least b letters B, we can summarize the above considerations by saying that

$$\#\text{3iet}(N) \geq 2 \sum_{b=0}^{N} H_{N-b,b} \, .$$

Moreover, since a Sturmian factor with at least b letters B is of length at least b, it follows that the sum is taken only over b such that $N - b \geq b$, i.e., $b \leq \lfloor N/2 \rfloor$. This gives the following

Proposition 10 *Let $N \in \mathbb{N}$. Then*

$$\#3\mathrm{iet}(N) \geq 2 \sum_{b=0}^{\lfloor N/2 \rfloor} \# \textit{ Sturmian factors of length } (N-b) \textit{ with at least } b \textit{ letters B} .$$

To estimate the number in the right part of the equation, we use the relation (9). Let us mention that an asymptotic formula for the number of Sturmian factors w with given $|w|$ and $|w|_1$ was derived in [5].

Lemma 11 *The number $H_{M,b}$ of Sturmian factors of length M with at least b letters 1 satisfies*

$$H_{M,b} \geq \frac{1}{\pi^2}(M-b)M^2 + \mathcal{O}(M^{2+\delta}), \qquad \textit{for any } \delta > 0 .$$

Proof: Let f_i be the i-th Farey fraction $f_i = \frac{p_i}{q_i}$. Consider the set $L_M^{(i)}$. By Lemma 3, if $b \leq M - Mf_i - 1$, then every factor in $L_M^{(i)}$ has at least b letters 1. The inequality $M - Mf_i - 1 \geq b$ can be rewritten as

$$p_i \leq \frac{M-b-1}{M} q_i . \tag{11}$$

By (8), the number of Farey fractions with the same denominator $q = q_i \leq M$ satisfying (11) is equal to

$$\frac{M-b-1}{M}\varphi(q) + \mathcal{O}(q^\delta), \qquad \text{for any } \delta > 0 . \tag{12}$$

Since the values of Farey fractions increase with i, there exists a maximal index, say s, such that every factor in the set $\bigcup_{i=0}^{s} L_M^{(i)}$ has sufficient number of letters 1. Therefore

$$H_{M,b} \geq \# \bigcup_{i=0}^{s} L_M^{(i)} .$$

In order to determine the cardinality of this union, we write it as a disjoint union, namely

$$\bigcup_{i=0}^{s} L_M^{(i)} = L_M^{(0)} \cup \bigcup_{i=1}^{s} \left(L_M^{(i)} \setminus L_M^{(i-1)} \right) .$$

Since $\#L_M^{(i)} = q_i$ and $\#L_M^{(i-1)} = q_{i-1}$ and equality (9) holds, we obtain

$$\#\left(L_M^{(i)} \setminus L_M^{(i-1)} \right) = M + 1 - q_{i-1} ,$$

which implies

$$H_{M,b} \geq 1 + \sum_{i=1}^{s} (M + 1 - q_{i-1}) = \sum_{i=0}^{s} (M + 1 - q_i) - M + q_s .$$

We group summands with the same values of $q_i = q$. The number of such summands (for each q) is given by (12). Therefore

$$H_{M,b} \geq \sum_{q=1}^{M}(M+1-q)\frac{M-b-1}{M}\varphi(q) + \mathcal{O}(M^{2+\delta}) = \\ = \frac{M-b}{M}\frac{M^3}{\pi^2} + \mathcal{O}(M^{2+\delta}),$$

where we have used (7). The result easily follows. □

The following theorem provides a slightly better lower bound on $\#3\mathrm{iet}(N)$ than that announced in Theorem 2. Instead of the constant $\frac{1}{3}$ we obtain $\frac{17}{48} > \frac{1}{3}$.

Theorem 12 *The number of 3iet factors of length N satisfies*

$$\#3\mathrm{iet}(N) \geq \frac{17}{48\pi^2}N^4 + \mathcal{O}(N^{3+\delta}),$$

for arbitrary $\delta > 0$.

Proof: Combining the formula from Proposition 10 and Lemma 11, we have

$$\begin{aligned}\#3\mathrm{iet}(N) \geq 2\sum_{b=0}^{\lfloor N/2\rfloor} H_{N-b,b} \geq \frac{2}{\pi^2}\sum_{b=0}^{\lfloor N/2\rfloor}(N-2b)(N-b)^2 + \sum_{b=0}^{\lfloor N/2\rfloor}\mathcal{O}\big((N-b)^{2+\delta}\big) = \\ = \frac{2}{\pi^2}\sum_{b=0}^{\lfloor N/2\rfloor}(N^3 - 4N^2b + 5Nb^2 - 2b^3) + \mathcal{O}(N^{3+\delta}) = \\ = \frac{2}{\pi^2}\Big(N^3\frac{N}{2} - 4N^2\frac{(N/2)^2}{2} + 5N\frac{(N/2)^3}{3} - 2\frac{(N/2)^4}{4}\Big) + \mathcal{O}(N^{3+\delta}) = \\ = \frac{2N^4}{\pi^2}\Big(\frac{1}{2} - \frac{1}{2} + \frac{5}{24} - \frac{1}{32}\Big) + \mathcal{O}(N^3\log N) = \frac{17}{48\pi^2}N^4 + \mathcal{O}(N^{3+\delta}).\end{aligned}$$

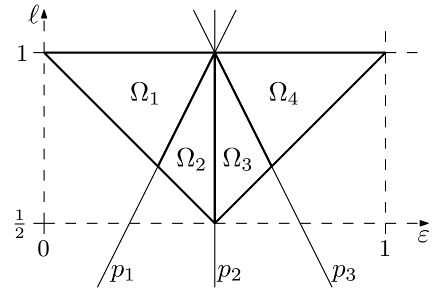
□

7 Precise values for small lengths

We have seen for Sturmian words, that given $N \in \mathbb{N}$, the interval $(0,1)$ of possible values of the parameter ε was divided into $\sum_{k=1}^{N}\varphi(k)$ areas such that all Sturmian words with parameter within one area had the same set of $N+1$ factors of length N. Similarly, one can expect that the family of possible pairs of parameters $\varepsilon \in (0,1)$ and $\ell \in \big(\max\{\varepsilon, 1-\varepsilon\}, 1\big)$ will be divided into regions with the same set of factors of length N. The number of factors of a 3iet word with parameters in the interior of the regions will be equal to $2N+1$.

Let us describe the way to obtain the list of factors of length N for fixed parameters ε, ℓ. Denote $d_1 = \varepsilon$, $d_2 = \ell - 1 + \varepsilon$, the discontinuity points of the transformation $T_{\varepsilon,\ell}$. The domain of $T_{\varepsilon,\ell}$, namely $[0,\ell)$, is divided by points $T_{\varepsilon,\ell}^{-j}(d_1)$, $T_{\varepsilon,\ell}^{-j}(d_2)$, $j = 0, 1, \dots, N-1$ generically into $2N+1$ subintervals. (This happens for $\ell \notin \mathbb{Z} + \mathbb{Z}\varepsilon$ and for small N also for other ℓ. Otherwise, some of these iterations coincide and the number of subintervals is smaller.) Each of these subintervals corresponds to one factor of length N occurring in the language of the infinite word u_{ε,ℓ,x_0} for any x_0. The ordering of iterations $T_{\varepsilon,\ell}^{-j}(d_1)$, $T_{\varepsilon,\ell}^{-j}(d_2)$ in $[0, \ell)$ depends on parameters ε, ℓ; different orderings give rise to different lists of 3iet factors. Values of ε, ℓ providing the same ordering are given by linear inequalities in ε, ℓ. Figure 6 shows the division of the region of parameters ε, ℓ by these inequalities for factors of length 2. The set of all 3iet factors of length 2 is the union of lists given in the figure, i.e.,

$$\mathrm{3iet}(2) = \{AA, AB, AC, BA, BB, BC, CA, CB, CC\}\,, \qquad \text{and} \qquad \#\mathrm{3iet}(2) = 9\,.$$



$p_1 : \ell = 2\varepsilon,\ p_2 : \varepsilon = \frac{1}{2},\ p_3 : \ell = 2(1-\varepsilon).$

$$\begin{aligned}
\Omega_1 &: \quad AC, BC, CA, CB, CC \\
\Omega_2 &: \quad AC, BB, BC, CA, CB \\
\Omega_3 &: \quad AB, AC, BA, BB, CA \\
\Omega_4 &: \quad AA, AB, AC, BA, AC
\end{aligned}$$

Fig. 6: Division of the region of parameters ε, ℓ by inequalities for factors of length 2 and the corresponding lists of factors.

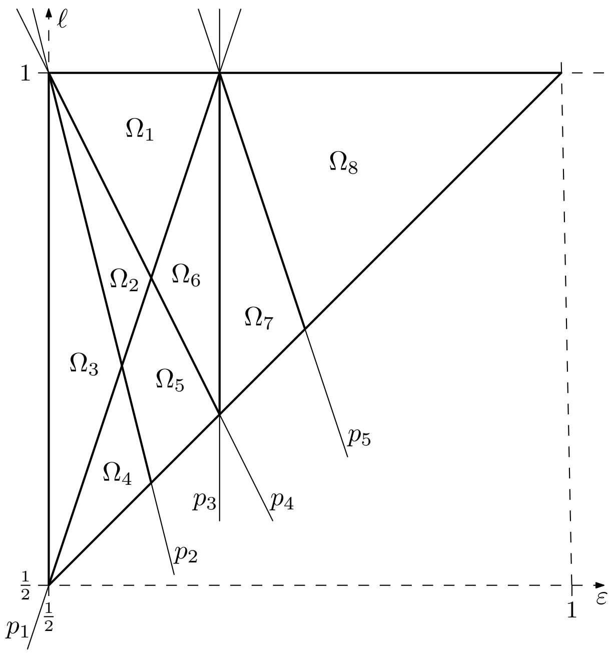
In Figure 7, we give the same analysis for factors of length $N = 3$. Note that the region and its division into areas according to factors of given length must be symmetric with respect to the axis $\varepsilon = 1/2$. This corresponds to interchange of letters $A \leftrightarrow C$ in all the factors. Figure 7 shows only one half of the picture.

The lists of factors of length 3 in individual areas of Figure 7 are given as follows.

$$\begin{aligned}
\Omega_1 &: \quad AAC, ABA, ACA, BAC, CAA, CAB, CAC \\
\Omega_2 &: \quad ABA, ABB, ACA, BAC, BBA, CAB, CAC \\
\Omega_3 &: \quad ABB, ACA, BAC, BBA, BBB, CAB, CAC \\
\Omega_4 &: \quad ABB, ACA, BAB, BAC, BBA, BBB, CAB \\
\Omega_5 &: \quad ABA, ABB, ACA, BAB, BAC, BBA, CAB \\
\Omega_6 &: \quad AAC, ABA, ACA, BAB, BAC, CAA, CAB \\
\Omega_7 &: \quad AAB, AAC, ABA, ACA, BAA, BAB, CAA \\
\Omega_8 &: \quad AAA, AAB, AAC, ABA, ACA, BAA, CAA
\end{aligned}$$

The lists of factors in areas $\Omega_9, \dots, \Omega_{16}$ not shown in Figure 7 are obtained from $\Omega_1, \dots, \Omega_8$ by the interchange $A \leftrightarrow C$. The set of all 3iet factors of length 3 is therefore equal to

$$\mathrm{3iet}(3) = \{A, B, C\}^3 \setminus \{ABC, CBA\}$$



$p_1: \quad \ell = 3\varepsilon - 1$

$p_2: \quad \ell = 3 - 4\varepsilon$

$p_3: \quad \varepsilon = \frac{2}{3}$

$p_4: \quad \ell = 2(1-\varepsilon)$

$p_5: \quad \ell = 3(1-\varepsilon)$

Fig. 7: Division of the region of parameters ε, ℓ by inequalities for factors of length 3.

and hence $\#3\mathrm{iet}(3) = 25$.

A computer evaluation provides the following table of values $\#3\mathrm{iet}(N)$ for $N = 1, \ldots, 10$.

N	1	2	3	4	5	6	7	8	9	10
$\#3\mathrm{iet}(N)$	3	9	25	55	113	199	339	531	809	1165
$\frac{\pi^2 \#3\mathrm{iet}(N)}{N^4} \approx$	29.6	5.55	3.05	2.12	1.78	1.52	1.39	1.28	1.22	1.15

As we have proved above, asymptotically, the values from the last string of the table should fall between $17/48$ and 2. It seems from the table that these bounds start to hold quite early, and perhaps it is possible to improve them.

Acknowledgements

We acknowledge financial support by the Czech Science Foundation grant 201/09/0584 and by the grants MSM6840770039 and LC06002 of the Ministry of Education, Youth, and Sports of the Czech Republic. The second author was supported also by RFBR grant 09-01-00244.

References

[1] B. Adamczewski. *Codages de rotations et phénomènes d'autosimilarité*. J. Théor. Nombres Bordeaux **14** (2002), 351–386.

[2] P. Alessandri, V. Berthé. *Three distance theorems and combinatorics on words*. Enseign. Math. **44** (1998), 103–132.

[3] P. Arnoux, V. Berthé, Z. Masáková, E. Pelantová. *Sturm numbers and substitution invariance of 3iet words*. Integers **8** (2008), A14, 17.

[4] P. Baláži, Z. Masáková, E. Pelantová. *Characterization of substitution invariant words coding exchange of three intervals*. Integers **8** (2008), A20, 21.

[5] N. Bédaride, E. Domenjoud, D. Jamet, J.-L. Rémy. *On the number of balanced words of given length and height over a two-letter alphabet*. Discrete mathematics and theoretical computer science **12** (2010), 41–62.

[6] A. Ya. Belov, A. L. Chernyat'ev. *Words with low complexity and interval exchange transformations*. Russ. Math. Surv. **63** (2008), 158–160.

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

[8] M. D. Boshernitzan, C. R. Carroll. *An extension of Lagrange's theorem to interval exchange transformations over quadratic fields*. J. Anal. Math. **72** (1997), 21–44.

[9] J. Cassaigne. *Sequences with grouped factors*. Developments in Language Theory III (DLT'97), Aristotle University of Thessaloniki, (1998), 211–222.

[10] A. L. Chernyat'ev. *Words with minimal growth function*. Moscow Univ. Math. Bull. **63** (2008), 262–264.

[11] S. Ferenczi, C. Holton, L. Q. Zamboni. *Structure of three-interval exchange transformations. II. A combinatorial description of the trajectories*. J. Anal. Math. **89** (2003), 239–276.

[12] L.-S. Guimond, Z. Masáková, E. Pelantová. *Combinatorial properties of infinite words associated with cut-and-project sequences*. J. Théor. Nombres Bordeaux **15** (2003), 697–725.

[13] G. H. Hardy, E. M. Wright. *An Introduction to the Theory of Numbers*. Oxford University Press, Oxford, sixth edition, (2008). Revised by D. R. Heath-Brown and J. H. Silverman.

[14] A. Katok. *Interval exchange transformations and some special flows are not mixing*. Israel J. Math. **35** (1980), 301–310.

[15] A. B. Katok, A. M. Stepin. *Approximations in ergodic theory*. Usp. Math. Nauk. **22** (1967), 81–106.

[16] M. Langevin. *Stimulateur cardiaque et suites de Farey*. Period. Math. Hungar. **23** (1991), 75–86.

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

[18] M. Lothaire. *Algebraic Combinatorics on Words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, (2002).

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

[20] M. Morse and G. A. Hedlund. *Symbolic dynamics II. Sturmian trajectories*. Amer. J. Math. **62** (1940), 1–42.

[21] M. B. Nathanson. *Elementary methods in number theory*, volume 195 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, (2000).

[22] N. Pytheas Fogg. *Substitutions in Dynamics, Arithmetics and Combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer Verlag, (2002).

[23] V. T. Sós. *On the distribution mod 1 of the sequence* $n\alpha$. Ann. Univ. Sci. Budapest. Eötvös Sect. Math. **1** (1958), 127–134.