A Run-hierarchical Description of Upper Mechanical Words with Irrational Slopes Using Continued Fractions

Hanna Uscka-Wehlou

Uppsala University, Department of Mathematics
Box 480, SE-751 06 Uppsala, Sweden
hania@wehlou.com
http://hania.wehlou.com

Abstract. The main result is a run-hierarchical description (by continued fractions) of upper mechanical words with slope $a \in]0,1[\setminus \mathbf{Q}$ and intercept 0. We compare this description with two classical methods of forming of such words. In order to be able to perform the comparison, we present a quantitative analysis of our method. We use the denominators of the convergents of the continued fraction expansion of the slope to compute the length of the prefixes obtained by our method. Due to the special treatment which is given to the elements equal to 1, our method gives in some cases longer prefixes than the two other methods. Our method reflects the hierarchy of runs, by analogy to digital lines, which can give a new understanding of the construction of upper mechanical words.

Keywords: upper mechanical word, characteristic word, digital line, irrational slope, continued fraction, run, hierarchy.

1 Introduction

In the presented paper we have basically two goals. The first one is to create a description of the construction of upper mechanical words (Def. 2) with irrational positive slope $a < 1$ and intercept 0, according to the hierarchy of runs, runs of runs, etc. Such a description can be a useful tool for examining of properties of upper and lower mechanical and characteristic words with irrational slopes, as has been shown in another paper by the author [10]. Our second goal is to show that our method works, in certain cases, faster than two well-known methods of forming of prefixes of characteristic words.

The theoretical base for this article are two earlier papers [8, 9] of the author. The run-hierarchical method is derived from the author's continued fraction (CF) based description of digitization of positive half lines $y = ax$. It is based on simple integer computations, thus can be used with advantage in computer programming. This qualitative description constitutes the first main result of the present paper (Theorem 3).

The second main result is a quantitative description of our method of forming prefixes of upper mechanical words (Theorem 6 and Corollary 1). We show there how to calculate the length of the prefixes of upper mechanical word formed according to our method. The length is expressed in terms of the denominators of the convergents of the CF expansion of the slope. These formulae allow us to compare our method with the classical and most frequently used descriptions by Venkov (1970) [11] and by Shallit (1991) [7]. In both of them one can express the length of the prefixes by these denominators (see Proposition 2 and Theorem 5).

The special treatment the CF elements equal to 1 get in our description (Theorem 3) makes that our process of forming words generates for some slopes longer prefixes than the similar classical recursive formulae presented by Shallit and Venkov.

We show that, for all a, the prefix P_k of the upper mechanical word $s'(a) = 1c(a)$ generated by our method is longer or of the same length compared to the prefix X_k of the corresponding word $c(a)$ generated by Shallit's method for each $k \in \mathbf{N}^+$ (Proposition 3). For some a our method generates *much* longer prefixes (Proposition 4 and Theorem 7).

The comparison with Venkov's method begins with Theorem 8. It depends on the set of 1's in the CF expansion of the slope a. For some a our method gives much longer prefixes than the method of Venkov after the same number of steps and our advantage can be as large as we want. For other slopes the method of Venkov generates longer prefixes (Proposition 5). However, Venkov's advantage in the k^{th} step for each $k \geq 3$ is always bounded by k (Proposition 6), while our advantage in case of slopes containing 1's in their CF expansion can be arbitrarily large. The *advantage* in this paper is expressed by the quotient of the length of the prefixes obtained when using the methods we compare.

The fact that we highlight some CF elements equal to 1 in the expansion of the slopes is not because they give us sometimes an advantage of forming longer prefixes than when using the well-known methods but because they determine the construction of lines (words) in terms of runs. This will be explained in what follows under Theorem 2 and in Section 6.

A list of references to papers concerning CF descriptions of characteristic words with irrational slopes can be found in Lothaire (2002) [4]. The most relevant for the present paper are Bernoulli (1772), Markoff (1882), Stolarsky (1976), Fraenkel et al. (1978) and Brown (1993). The first three papers correspond to the method of Venkov (described already much earlier by Markov), the last two correspond to the method of Shallit. The CF description method presented in Theorem 3 seems to be the only one which gives prefixes constructed according to the run hierarchy. This enables us to analyze the construction of upper mechanical words, which has been presented by the author in [10].

2 Continued Fractions — a Brief Introduction

The following algorithm gives the regular (or simple) CF for $a \in \mathbf{R} \setminus \mathbf{Q}$, which we denote by $[a_0; a_1, a_2, a_3, \ldots]$. We define a sequence of integers (a_n) and a sequence of real numbers (α_n) by: $\alpha_0 = a$; $a_n = \lfloor \alpha_n \rfloor$ and $\alpha_n = a_n + \frac{1}{\alpha_{n+1}}$ for $n \geq 0$. Then $a_n \geq 1$ and $\alpha_n > 1$ for $n \geq 1$. The integers $a_0, a_1, a_2, \ldots$ are called the *elements* of the

CF (or *terms*, or *partial quotients*). We use the word *elements*, following Khinchin (1997:1) [2]. Because a is irrational, so is each α_n, and the sequences (a_n) and (α_n) are infinite. A CF expansion exists and is unique for all $a \in \mathbf{R} \setminus \mathbf{Q}$; see [2], p. 16.

For $k \in \mathbf{N}$, the k^{th} *order convergent* of the CF $a = [a_0; a_1, a_2, \ldots]$ is the canonical representation of the number $s_k = [a_0; a_1, a_2, \ldots, a_k]$. We will denote it by p_k/q_k. The following theorem comes from the definition of CFs and can be found for example in Khinchin (1997:4) [2].

Theorem 1. *For the denominators of the convergents of each* $a = [a_0; a_1, a_2, \ldots]$ *we have* $q_0 = 1,\ q_1 = a_1,$ *and, for* $k \geq 2,\ \ q_k = a_k q_{k-1} + q_{k-2}$.

It follows immediately from the recursive formula in Theorem 1 that the sequence $(q_n)_{n \in \mathbf{N}^+}$ for each $a \in \mathbf{R} \setminus \mathbf{Q}$ is a **strictly increasing** sequence of natural numbers. We will exploit this fact heavily in what follows.

3 Earlier Results

In this section we recapitulate some results obtained by the author in [9]. Arithmetical description of the modified Rosenfeld digitization (R′-digitization) of the positive half line $y = ax$ for $a \in \,]0, 1[\, \setminus \mathbf{Q}$ as a subset of $\mathbf{Z}^2$ is the following:

$$D_{R'}(y = ax,\ x > 0) = \{(k, \lceil ak \rceil);\ \ k \in \mathbf{N}^+\}. \tag{1}$$

Our CF description from [9] was based on the description by digitization parameters from Uscka-Wehlou (2007) [8] and the following *index jump function*.

Definition 1. *For each* $a \in \,]0, 1[\, \setminus \mathbf{Q}$*, the* index jump function $i_a : \mathbf{N}^+ \to \mathbf{N}^+$ *is defined by* $i_a(1) = 1,\ i_a(2) = 2$ *and* $i_a(k+1) = i_a(k) + 1 + \delta_1(a_{i_a(k)})$ *for* $k \geq 2$*, where* $\delta_1(x) = \begin{cases} 1, x = 1 \\ 0, x \neq 1 \end{cases}$ *and* $a_1, a_2, \ldots \in \mathbf{N}^+$ *are the CF elements of* a.

The index jump function is a renumbering, which avoids elements following directly after some 1's in the CF expansion (in particular, it avoids every second element in the sequences of consecutive 1's with index greater than 1).

In both papers [8] and [9], digital lines were described according to the hierarchy of runs on all the digitization levels. The term *run* was already introduced by Azriel Rosenfeld (1974:1265) [6]. For the formal definition of runs and the modification of Rosenfeld digitization see [8]. We called $\mathrm{run}_k(j)$ for $k, j \in \mathbf{N}^+$ a *run of digitization level* k. Each $\mathrm{run}_1(j)$ can be identified with a subset of $\mathbf{Z}^2$:
$\{(i_0+1, j), (i_0+2, j), \ldots, (i_0+m, j)\}$, where m is the length $|\mathrm{run}_1(j)|$ of this run. For each $a \in \,]0, 1[\, \setminus \mathbf{Q}$ we have only two possible run_1 lengths: $\left\lfloor \frac{1}{a} \right\rfloor$ and $\left\lfloor \frac{1}{a} \right\rfloor + 1$. All the runs with one of those lengths always occur alone, i.e., do not have any neighbors of the same length in the sequence $(\mathrm{run}_1(j))_{j \in \mathbf{N}^+}$, while the runs of the other length can appear in sequences. The same holds for the sequences $(\mathrm{run}_k(j))_{j \in \mathbf{N}^+}$ on each level $k \geq 2$, i.e., runs on each level k can have one of two possible lengths (being

consecutive natural numbers) and runs with one of these lengths always appear alone in the sequence of runs_k. Runs of level $k+1$ for $k \in \mathbf{N}^+$ are defined recursively, as sets of runs_k and, in this context (but it will no longer be so in Section 5), by the length of run_{k+1} we mean its cardinality. Each run_{k+1} consists of one singly appearing run_k (called *short* run of level k and denoted S_k if its length is expressed by the least of the mentioned consecutive numbers for level k and called *long* run of level k and denoted L_k otherwise) and all the runs_k (L_k or S_k, respectively) which can appear in sequences comming between this single run_k and the next or the previous single run_k, depending on $\mathrm{run}_k(1)$, in the sequence $(\mathrm{run}_k(j))_{j \in \mathbf{N}^+}$. This means that runs_{k+1} for each $k \in \mathbf{N}^+$ can have one of following four shapes: $S_k^m L_k$, $L_k S_k^m$, $L_k^m S_k$ or $S_k L_k^m$, where m can be one of two consecutive positive integers which depend on the slope a and the level number k. For example, $S_k^m L_k$ means that the run_{k+1} consists of m short runs_k (S_k) and one long run_k (L_k) in this order. For the purpose of this paper this description suffices; for the formal definition see [8]. Moreover, Theorem 2, proven in [9], can serve as a definition of runs in the digitizations of straight lines $y = ax$ for $a \in]0,1[\setminus \mathbf{Q}$, since it presents a complete recurrent description of these. The theorem is completely CF based.

Theorem 2 (Main Result in [9]; description by CFs). *Let $a \in]0,1[\setminus \mathbf{Q}$ and $a = [0; a_1, a_2, \ldots]$. For the digital line with equation $y = ax$, we have $|S_1| = a_1$, $|L_1| = a_1 + 1$, and the forms of* runs_k ($\mathrm{form\_run}_k$) *for $k \geq 2$ are as follows:*

$$\mathrm{form\_run}_k = \begin{cases} S_{k-1}^m L_{k-1} & \text{if} \quad a_{i_a(k)} \neq 1 \quad \text{and} \quad i_a(k) \quad \text{is even} \\ S_{k-1} L_{k-1}^m & \text{if} \quad a_{i_a(k)} = 1 \quad \text{and} \quad i_a(k) \quad \text{is even} \\ L_{k-1} S_{k-1}^m & \text{if} \quad a_{i_a(k)} \neq 1 \quad \text{and} \quad i_a(k) \quad \text{is odd} \\ L_{k-1}^m S_{k-1} & \text{if} \quad a_{i_a(k)} = 1 \quad \text{and} \quad i_a(k) \quad \text{is odd,} \end{cases} \tag{2}$$

where $m = b_k - 1$ if the run_k *is short (S_k) and $m = b_k$ if the* run_k *is long (L_k). The function i_a is defined in Def. 1 and $b_k = a_{i_a(k)} + \delta_1(a_{i_a(k)}) a_{i_a(k)+1}$.*

We remark that the value of the index jump function for each natural $k \geq 2$ describes the index of the CF element which determines the construction of runs on level k in terms of runs of level $k-1$. If this CF element ($a_{i_a(k)}$) is equal to 1, the most frequent run on level $k-1$ is the long one (L_{k-1}). In all the other cases, i.e., if $a_{i_a(k)} > 1$, the most frequently appearing run on level $k-1$ is the short one (S_{k-1}). This means that the CF elements equal to 1 which are indexed by the values of the index jump function (greater than 1) play a very special role in the run hierarchical construction of digitized $y = ax$. In the author's paper [10] elements like this are called *essential* 1's. They have been used in [10] for a partition of all digital lines with slopes $a \in]0,1[\setminus \mathbf{Q}$ into equivalence classes. The equivalence relation is defined by the essential 1's of the CF expansions of the slopes and all the lines belonging to the same class have the same construction in terms of the forms of digitization runs. This partition was possible because of the description contained in Theorem 2 and can be of interest for combinatorics on words, due to the equivalence between digital lines and mechanical words.

4 Characteristic and (Upper, Lower) Mechanical Words and the Modified Rosenfeld Digitization

First we provide a brief introduction to characteristic and upper and lower mechanical words. The following definition comes from Lothaire (2002:53) [4].

Definition 2. *Given two real numbers* α *and* ρ *with* $0 \leq \alpha \leq 1$, *we define two infinite words* $s_{\alpha,\rho} : \mathbf{N} \to \{0,1\}$, $\quad s'_{\alpha,\rho} : \mathbf{N} \to \{0,1\}$ *by*

$$s_{\alpha,\rho}(n) = \lfloor \alpha(n+1) + \rho \rfloor - \lfloor \alpha n + \rho \rfloor, \quad s'_{\alpha,\rho}(n) = \lceil \alpha(n+1) + \rho \rceil - \lceil \alpha n + \rho \rceil.$$

The word $s_{\alpha,\rho}$ *is the* lower mechanical word *and* $s'_{\alpha,\rho}$ *is the* upper mechanical word *with* slope α *and* intercept ρ. *A lower or upper mechanical word is* irrational *or* rational *according as its slope is irrational or rational.*

In the present paper we deal with the special case when $\alpha \in \,]0,1[$ is irrational and $\rho = 0$. In this case we will denote the lower and upper mechanical words by $s = s(\alpha)$ and $s' = s'(\alpha)$ respectively. We have $s_0 = s_0(\alpha) = \lfloor \alpha \rfloor = 0$ and $s'_0 = s'_0(\alpha) = \lceil \alpha \rceil = 1$ and, because $\lceil x \rceil - \lfloor x \rfloor = 1$ for irrational x, we have

$$s = s(\alpha) = 0c(\alpha), \quad s' = s'(\alpha) = 1c(\alpha) \tag{3}$$

(meaning 0, resp. 1 concatenated to $c(\alpha)$). The word $c(\alpha)$ is called the *characteristic word* of α. For each $\alpha \in \,]0,1[\, \setminus \mathbf{Q}$, the characteristic word associated with α is thus the following infinite word $c = c(\alpha) : \mathbf{N}^+ \to \{0,1\}$:

$$c_n = \lfloor \alpha(n+1) \rfloor - \lfloor \alpha n \rfloor = \lceil \alpha(n+1) \rceil - \lceil \alpha n \rceil, \quad n \in \mathbf{N}^+. \tag{4}$$

The connection between characteristic words and digital lines is a well-known fact. See for example Lothaire (2002:53, 2.1.2 *Mechanical words, rotations*) [4], Pytheas Fogg (2002:143, 6. *Sturmian Sequences*) [5] or Klette and Rosenfeld (2004) [3]. In [8] the author remarks that the modified Rosenfeld digitization of the line $y = ax$, where $a \in \,]0,1[\, \setminus \mathbf{Q}$, is the subset of $\mathbf{Z}^2$ described by (1). This means, from (3) and (4), that the sequence $s'_0 = 1$, $s'_n = \lfloor (n+1)a \rfloor - \lfloor na \rfloor$ for $n \in \mathbf{N}^+$ describes the R′-digitization of $y = ax, x > 0$. So, for any $a \in \,]0,1[\, \setminus \mathbf{Q}$, the upper mechanical word $s'(a)$, as defined in Def. 2, describes completely the digitization of the positive half line $y = ax$. We can thus write $s'(a) = 10^{m_1}10^{m_2}10^{m_3}\ldots$, where $m_i \in \mathbf{N}$ for $i \in \mathbf{N}^+$. We have $|\mathrm{run}_1(i)| = 1 + m_i$, each run begins with a 1. Moreover, there exists $d_1 \in \mathbf{N}$ such that for all $i \in \mathbf{N}^+$ we have $m_i = d_1$ or $m_i = d_1 + 1$ and we know from the theory for digital lines that $d_1 = \lfloor \frac{1}{a} \rfloor - 1$. If $\lfloor \frac{1}{a} \rfloor = 1$, then $d_1 = 0$ and $|S_1| = 1$. Because of the correspondence between digital lines $y = ax$ and upper mechanical words $s'(a)$ for $a \in \,]0,1[\, \setminus \mathbf{Q}$, we also have the run hierarchical structure of upper mechanical words. Runs of level 1 are $S_1 = 10^{d_1}$ and $L_1 = 10^{d_1+1}$, where $d_1 = \lfloor \frac{1}{a} \rfloor - 1$ and we can defined recursively for each $k \in \mathbf{N}^+$ the runs of level $k+1$ as sets of runs of level k symbolically denoted as $S_k^{d_{k+1}} L_k$, $L_k S_k^{d_{k+1}}$, $L_k^{d_{k+1}} S_k$ or $S_k L_k^{d_{k+1}}$, where

d_{k+1} can be one of two consecutive positive integers which depend on the slope a and the level number $k+1$. We again use the notation of S for *short* and L for *long*, because words also have two possible run lengths (cardinalities) per level, due to the equivalence between digital lines and upper mechanical words.

The upper mechanical words s' for the slopes of all the lines with digitization around the origin as shown in the picture in Fig. 1 begin with 10001000.

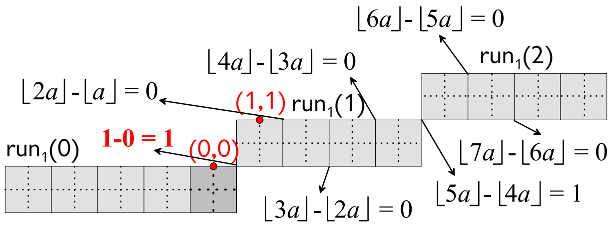

Fig. 1. Upper mechanical words $s'(a)$ and digital lines $y = ax$ for $a \in \left]\frac{1}{5}, \frac{1}{4}\right[\setminus \mathbf{Q}$.

This correspondence between the words and digital lines allows us to derive the following CF description of upper mechanical words from our result for digital lines. Because we have (3), our result will also give a description of lower mechanical words and characteristic words.

Theorem 3 (Main Result 1; a run-hierarchical CF description of upper mechanical words). *Let $a \in]0,1[\setminus \mathbf{Q}$ and $a = [0; a_1, a_2, \ldots]$. For $s'(a)$ as in Def. 2 we have $s'(a) = \lim_{k\to\infty} P_k$, where $P_1 = S_1 = 10^{a_1-1}$, $L_1 = 10^{a_1}$, and, for $k \geq 2$,*

$$P_k = \begin{cases} L_k = S_{k-1}^{a_{i_a(k)}} L_{k-1} & \text{if} \quad a_{i_a(k)} \neq 1 \quad \text{and} \quad i_a(k) \quad \text{is even} \\ S_k = S_{k-1} L_{k-1}^{a_{i_a(k)+1}} & \text{if} \quad a_{i_a(k)} = 1 \quad \text{and} \quad i_a(k) \quad \text{is even} \\ S_k = L_{k-1} S_{k-1}^{-1+a_{i_a(k)}} & \text{if} \quad a_{i_a(k)} \neq 1 \quad \text{and} \quad i_a(k) \quad \text{is odd} \\ L_k = L_{k-1}^{1+a_{i_a(k)+1}} S_{k-1} & \text{if} \quad a_{i_a(k)} = 1 \quad \text{and} \quad i_a(k) \quad \text{is odd,} \end{cases} \tag{5}$$

*where the function i_a is defined in Def. 1. The meaning of the symbols is the following: for $k \geq 1$, P_k - **P**refix number k, S_k - **S**hort run_k and L_k - **L**ong run_k. To make the recursive formula (5) complete, we add that for each $k \geq 2$, if $P_k = S_k$, then L_k is defined in the same way as S_k, with the only difference that the exponent defined by $a_{i_a(k)}$ (or by $a_{i_a(k)+1}$) is increased by 1. If $P_k = L_k$, then S_k is defined in the same way as L_k, with the only difference that the exponent defined by $a_{i_a(k)}$ (or by $a_{i_a(k)+1}$) is decreased by 1.*

Proof. We use Theorem 2 and the equivalence between the digital half lines $y = ax$ (where $a \in]0,1[\setminus \mathbf{Q}$ and $x > 0$) and the words $s'(a)$. We introduced P_k which corresponds to $\text{run}_k(1)$ for each $k \in \mathbf{N}^+$. According to Theorem 2, $\text{run}_{k-1}(1)$ for $k \geq 2$ is short if $i_a(k)$ is even (this result is represented by the first two rows of (2)) and long if $i_a(k)$ is odd (rows 3 and 4 in (2)). This means that $P_k = \text{run}_k(1)$ is short (S_k) if $i_a(k+1)$ is even and long (L_k) if $i_a(k+1)$ is odd. Because we have

$i_a(k+1) = i_a(k)+1+\delta_1(a_{i_a(k)})$, the parity of $i_a(k+1)$ is determined by the parity of $i_a(k)$ and $\delta_1(a_{i_a(k)})$, thus, in the cases described by the first and the fourth rows of (2), $P_k = L_k$, and, in the cases described by the second and the third rows, $P_k = S_k$. The exponents in (5) are computed according to the formula for b_k presented in Theorem 2. □

We have described $s'(a)$ by an increasing sequence of prefixes $(P_k)_{k\in\mathbf{N}^+}$. Prefix P_k for each $k \in \mathbf{N}^+$ corresponds to the first run of level k ($\mathrm{run}_k(1)$) in the digitization of $y = ax$, so this description reflects the hierarchy of runs.

Figure 2 shows a digital straight line segment (a prefix of upper mechanical word) and its hierarchy of runs. The picture shows the first digitization run on level 5, $\mathrm{run}_5(1) = S_5$ (the 5$^{\text{th}}$ prefix P_5) for the lines $y = ax$ (the words $s'(a)$) with slopes $a = [0; 1, 2, 1, 1, 3, 1, 1, a_8, a_9, \ldots]$, where $a_8, a_9, \ldots \in \mathbf{N}^+$. The dark squares on Fig. 2 represent the short runs_1. They can occur in sequences, while the long runs_1 (white) can only appear alone. We will revisit this example in Sect. 5 (Example 1). More about the hierarchy of runs can be found in Sect. 3.

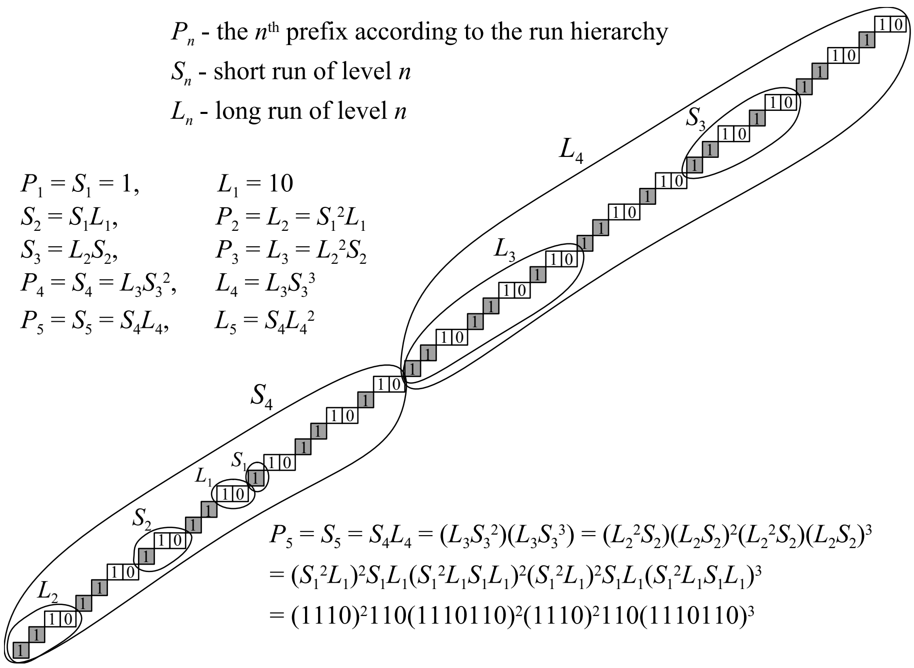

Fig. 2. Hierarchy of runs.

5 Comparison Between our Description by CFs and the Methods Described by Venkov and Shallit

In this section we consider only binary words over the two letter alphabet $\{0, 1\}$. For each such a word A, if it is finite, we denote by $|A|$ the length of A, being the total number of 0's and 1's forming A. In this section we no longer use the cardinality-wise run length as introduced in Section 3; we only use the (binary word)-length as defined above.

Let us first recall the well-known result formulated by the astronomer J. Bernoulli in 1772, proven by A. Markov in 1882 and described by Venkov (1970:67) [11].

Theorem 4 (Markov, Venkov). *For each irrational $a = [0; a_1, a_2, a_3, \ldots]$, the characteristic word is $c(a) = C_1C_2C_3\ldots$, where*

$$\begin{cases} C_1 = 0^{a_1-1}1 \\ D_1 = 0^{a_1}1 \end{cases}, \begin{cases} C_2 = C_1^{a_2-1}D_1 \\ D_2 = C_1^{a_2}D_1 \end{cases}, \ldots, \begin{cases} C_n = C_{n-1}^{a_n-1}D_{n-1} \\ D_n = C_{n-1}^{a_n}D_{n-1}. \end{cases}$$

Proposition 1 describes the length of C_n and D_n (meaning the number of 0's and 1's occurring in them), which leads immediately to Proposition 2.

Proposition 1. *With all the assumptions and the notation as in Theorem 4, we have $|C_k| = q_k$ and $|D_k| = q_k + q_{k-1}$ for all $k \in \mathbf{N}^+$, where q_k is the denominator of the k^{th} convergent of the CF expansion of a.*

Proof. By induction. For $k = 1$ we have $C_1 = 0^{a_1-1}1$, so $|C_1| = a_1 = q_1$ and $D_1 = 0^{a_1}1$, so $|D_1| = a_1 + 1 = q_1 + q_0$. Let's assume that $|C_k| = q_k$ and $|D_k| = q_k + q_{k-1}$ for some $k \geq 1$. By this assumption, combined with the definition of C_{k+1} and D_{k+1} and Theorem 1, we get $|C_{k+1}| = (a_{k+1} - 1)q_k + q_k + q_{k-1} = q_{k+1}$ and $|D_{k+1}| = a_{k+1}q_k + q_k + q_{k-1} = q_{k+1} + q_k$. $\square$

Proposition 2. *Let $a \in \,]0,1[\, \setminus \mathbf{Q}$. For each $n \in \mathbf{N}^+$, the length of the n^{th} prefix $C_1 \cdots C_n$ of $c(a)$ as defined in Theorem 4 is $|C_1 \cdots C_n| = q_1 + \cdots + q_n$, where q_k for $k \in \mathbf{N}^+$ is the denominator of the k^{th} convergent of the CF expansion of a.*

The second CF description of $c(a)$ we consider is that by Shallit (1991) [7], where $c(a)$ is formed as a limit of an increasing sequence of prefixes $(X_n)_{n\in\mathbf{N}^+}$; cf. the method by the *standard sequences* from Lothaire (2002:75, 76, 104, 105) [4].

Theorem 5 (Shallit 1991). *Let $a = [0; a_1, a_2, a_3, \ldots]$ be irrational and $c(a) = (\lfloor (n+1)a \rfloor - \lfloor na \rfloor)_{n\in\mathbf{N}^+}$ be its characteristic word. Let $X_0 = 0$. For the sequence of finite words $(X_n)_{n\in\mathbf{N}^+}$ being prefixes $X_n = c_1c_2\cdots c_{q_n}$ of $c(a)$ of length q_n, where q_n are the denominators of the convergents of the CF expansion of a, we have $X_1 = 0^{a_1-1}1$ and, for $n \geq 2$, $X_n = X_{n-1}^{a_n}X_{n-2}$.*

As we have seen (Proposition 2 and Theorem 5), the length of the prefixes of $c(a)$ obtained in both methods (Venkov's, Shallit's) can be expressed by the denominators of the convergents of the CF expansion of a. To be able to compare our result with their methods, we will now express the length of the prefixes P_k (from Theorem 3) of the upper mechanical word $s'(a) = 1c(a)$ in the same terms. The result is contained in Corollary 1. To get the corollary, we need the following theorem, which forms one of the main results in the present paper.

Theorem 6 (Main Result 2; a quantitative description of runs). *Let $a \in]0,1[\setminus \mathbf{Q}$ and $a = [0; a_1, a_2, \ldots]$. For the word $s'(a)$ we have for all $k \in \mathbf{N}^+$:*

$$|S_k| = q_{i_a(k+1)-1} \quad \text{and} \quad |L_k| = q_{i_a(k+1)-1} + q_{i_a(k+1)-2},$$

where i_a is the index jump function (Def. 1), $|S_k|$ and $|L_k|$ for $k \in \mathbf{N}^+$ denote the (binary word)-length of short, respectively long runs of level k as in Theorem 3, and q_k are the denominators of the convergents of the CF expansion of a.

Proof. By induction. We also use Def. 1 and Theorem 1. For $k = 1$ the statement is true, because $i_a(2) = 2$ and, due to Theorem 3, $|S_1| = a_1 = q_1$ and $|L_1| = a_1 + 1 = q_1 + q_0$. Let us now assume that the statement is true for some $n - 1 \geq 1$. We will show that it is also true for n. We consider four cases, as in Theorem 3:

- $a_{i_a(n)} \neq 1$ and $i_a(n)$ is even.
  We have $i_a(n+1) = i_a(n) + 1$ and $q_{i_a(n)} = a_{i_a(n)} q_{i_a(n)-1} + q_{i_a(n)-2}$, so:
  $|S_n| = (a_{i_a(n)} - 1) q_{i_a(n)-1} + q_{i_a(n)-1} + q_{i_a(n)-2} = q_{i_a(n)} - q_{i_a(n)-1} + q_{i_a(n)-1} = q_{i_a(n)} = q_{i_a(n+1)-1}$, $|P_n| = |L_n| = a_{i_a(n)} q_{i_a(n)-1} + q_{i_a(n)-1} + q_{i_a(n)-2} = q_{i_a(n)} + q_{i_a(n)-1} = q_{i_a(n+1)-1} + q_{i_a(n+1)-2}$.

- $a_{i_a(n)} = 1$ and $i_a(n)$ is even.
  We have $i_a(n+1) = i_a(n) + 2$ and $q_{i_a(n)} = q_{i_a(n)-1} + q_{i_a(n)-2}$, so:
  $|P_n| = |S_n| = q_{i_a(n)-1} + a_{i_a(n)+1} \cdot q_{i_a(n)} = q_{i_a(n)+1} = q_{i_a(n+1)-1}$,
  $|L_n| = q_{i_a(n)-1} + (1 + a_{i_a(n)+1}) \cdot q_{i_a(n)} = q_{i_a(n)+1} + q_{i_a(n)} = q_{i_a(n+1)-1} + q_{i_a(n+1)-2}$.

- $a_{i_a(n)} \neq 1$ and $i_a(n)$ is odd.
  We have $i_a(n+1) = i_a(n) + 1$ and $q_{i_a(n)} = a_{i_a(n)} q_{i_a(n)-1} + q_{i_a(n)-2}$, so:
  $|P_n| = |S_n| = q_{i_a(n)-1} + q_{i_a(n)-2} + (a_{i_a(n)} - 1) q_{i_a(n)-1} = q_{i_a(n)} + q_{i_a(n)-1} - q_{i_a(n)-1} = q_{i_a(n)} = q_{i_a(n+1)-1}$, $|L_n| = q_{i_a(n)-1} + q_{i_a(n)-2} + a_{i_a(n)} q_{i_a(n)-1} = q_{i_a(n)-1} + q_{i_a(n)} = q_{i_a(n+1)-2} + q_{i_a(n+1)-1}$.

- $a_{i_a(n)} = 1$ and $i_a(n)$ is odd.
  We have $i_a(n+1) = i_a(n) + 2$ and $q_{i_a(n)} = q_{i_a(n)-1} + q_{i_a(n)-2}$, so:
  $|S_n| = a_{i_a(n)+1} q_{i_a(n)} + q_{i_a(n)-1} = q_{i_a(n)+1} = q_{i_a(n+1)-1}$,
  $|P_n| = |L_n| = (1 + a_{i_a(n)+1}) q_{i_a(n)} + q_{i_a(n)-1} = q_{i_a(n)} + q_{i_a(n)+1} = q_{i_a(n+1)-2} + q_{i_a(n+1)-1}$.

The proof is complete. □

Corollary 1 (a quantitative description of prefixes). *Let $a \in]0,1[\setminus \mathbf{Q}$ and $a = [0; a_1, a_2, \ldots]$. The length of the prefixes P_k of the the upper mechanical word $s'(a)$ as defined in Theorem 3 is: $|P_1| = |S_1| = a_1$ and for all $k \geq 2$:*

$$|P_k| = \begin{cases} |L_k| = q_{i_a(k)} + q_{i_a(k)-1} & \text{if} \quad a_{i_a(k)} \neq 1 \quad \text{and} \quad i_a(k) \quad \text{is even} \\ |S_k| = q_{i_a(k)+1} & \text{if} \quad a_{i_a(k)} = 1 \quad \text{and} \quad i_a(k) \quad \text{is even} \\ |S_k| = q_{i_a(k)} & \text{if} \quad a_{i_a(k)} \neq 1 \quad \text{and} \quad i_a(k) \quad \text{is odd} \\ |L_k| = q_{i_a(k)+1} + q_{i_a(k)} & \text{if} \quad a_{i_a(k)} = 1 \quad \text{and} \quad i_a(k) \quad \text{is odd,} \end{cases} \tag{6}$$

where i_a is the index jump function and q_n for $n \in \mathbf{N}^+$ is the denominator of the n^{th} convergent of the CF expansion of a.

Proof. Follows from Theorems 3 and 6, and the fact that, for $k \geq 2$, $i_a(k+1) = i_a(k) + 1$ if $a_{i_a(k)} \neq 1$ and $i_a(k+1) = i_a(k) + 2$ if $a_{i_a(k)} = 1$. □

Let us remark that Corollary 1 shows that the sequences $(P_k)_{k\in\mathbf{N}^+}$ of prefixes of upper mechanical words $s'(a)$ generated by our method are usually (i.e., for most slopes a) **not** subsequences of $(X_k)_{k\in\mathbf{N}^+}$ generated by Shallit, even if we put the letter 1 in the front of each X_k and remove the last letter of each X_k, getting in this way prefixes of $s'(a) = 1c(a)$ with length equal to the denominator of a convergent of a. We have to impose two conditions on the slope a to make the corresponding $(P_k)_{k\in\mathbf{N}^+}$ be a subsequence of the corresponding $(X_k)_{k\in\mathbf{N}^+}$ (after this extra operation of putting the letter 1 in the front of each X_k and taking away the last letter of each X_k). These conditions imposed on the CF elements of a are:

- for each k for which $|P_k| = q_{i_a(k)} + q_{i_a(k)-1}$ it must be $a_{i_a(k)+1} = 1$, in order to get $q_{i_a(k)} + q_{i_a(k)-1} = q_{i_a(k)+1}$ (Theorem 1) so that P_k has the length equal to the denominator of a convergent of a, like $X_{i_a(k)+1}$ (Theorem 5).
- for each k for which $|P_k| = q_{i_a(k)+1} + q_{i_a(k)}$ it must be $a_{i_a(k)+2} = 1$, in order to get $q_{i_a(k)+1} + q_{i_a(k)} = q_{i_a(k)+2}$ so that P_k has the length equal to the denominator of a convergent of a, like $X_{i_a(k)+2}$.

All the lines as described in Example 2 below have this property (that the sequence of prefixes described by our method is a subsequence of the prefixes generated by Shallit's method - we use every second element of the sequence used by Shallit), but for the most slopes this is not the case. This also shows that the method by Shallit does not reflect the run hierarchical structure of words and that our method is different from his. We can say the same about the method by Venkov, but this is obvious, so we leave out the proof in this paper.

Example 1. The line segments $\mathrm{run}_k(1)$ for $k = 1, 2, 3, 4, 5$ on Fig. 2 correspond to the prefixes P_k of $s'(a)$ for all a such that $a = [0; 1, 2, 1, 1, 3, 1, 1, a_8, a_9, \ldots]$, where $a_8, a_9, \ldots \in \mathbf{N}^+$. For these a, $i_a(1) = 1$, $i_a(2) = 2$, $i_a(3) = 3$, $i_a(4) = 5$, $i_a(5) = 6$, $i_a(6) = 8$, and the denominators of the convergents are $q_1 = 1$, $q_2 = 3$, $q_3 = 4$, $q_4 = q_{i_a(4)-1} = 7$, $q_5 = q_{i_a(4)} = 25$, $q_6 = q_{i_a(5)} = 32$, $q_7 = q_{i_a(6)-1} = 57$. It is easy to check that the length $|P_k|$ of prefixes (runs) on Fig. 2 agrees with Corollary 1, so $|P_1| = 1$, $|P_2| = 4$, $|P_3| = 11$, $|P_4| = 25$, $|P_5| = 57$.

Example 2. Let $a = [0; a_1, 1, a_3, 1, a_5, 1, a_7, 1, a_9, \ldots]$, where $a_{2n+1} \in \mathbf{N}^+$ for all $n \in \mathbf{N}$. For $s'(a)$ we have $|P_k| = |S_k| = q_{2k-1}$ and $|L_k| = q_{2k}$ for all $k \in \mathbf{N}^+$ (the notation as in Theorem 3).

Indeed, the index jump function is $i_a(1) = 1$, $i_a(k) = 2k - 2$ for $k \geq 2$, so it is even for all $k \geq 2$. Moreover, $a_{i_a(k)} = 1$ for $k \geq 2$. From Theorem 6, $|L_k| = q_{i_a(k+1)-1} + q_{i_a(k+1)-2} = a_{i_a(k+1)} q_{i_a(k+1)-1} + q_{i_a(k+1)-2} = q_{i_a(k+1)} = q_{2k}$ and $|S_k| = q_{i_a(k+1)-1} = q_{2k-1}$, and from Corollary 1, $|P_k| = |S_k|$ for $k \in \mathbf{N}^+$.

Example 3. Let $a = [0; a_1, a_2, a_3, a_4, \ldots]$, where $a_1 \in \mathbf{N}^+$ and $a_n \geq 2$ for all $n \geq 2$ (thus $i_a(k) = k$ for all $k \in \mathbf{N}^+$). Due to Corollary 1, the lengths of the prefixes P_k (for $k \in \mathbf{N}^+$) of $s'(a)$ as defined in Theorem 3 are:

$$|P_k| = \begin{cases} |S_k| = q_k & \text{if } k \text{ is odd} \\ |L_k| = q_k + q_{k-1} & \text{if } k \text{ is even.} \end{cases} \tag{7}$$

Formulae (7) and the one from Proposition 1 look similar (we get the length q_k and $q_k + q_{k-1}$ in both cases), but they describe different parts of prefixes of $s'(a)$.

Now we will compare our method to that of Shallit.

Proposition 3. *Let $a \in]0,1[\setminus \mathbf{Q}$ and $a = [0; a_1, a_2, \ldots]$. We have $|P_k| \geq |X_k|$ for all $k \in \mathbf{N}^+$, where P_k is our k^{th} prefix of $s'(a) = 1c(a)$ and X_k is Shallit's k^{th} prefix of $c(a)$. There exists $k \geq 2$ for which the inequality is strict.*

Proof. For any a we have $|P_1| = q_1 = |X_1|$. The sequence $(q_n)_{n \in \mathbf{N}^+}$ is strictly increasing and for each $k \in \mathbf{N}^+$ we have $i_a(k) \geq k$, thus, from Theorems 6 and 5, we get $|P_k| \geq |S_k| = q_{i_a(k+1)-1} \geq q_k = |X_k|$ for $k \in \mathbf{N}^+$. The last statement follows from Corollary 1 and Example 3. The situation when only S_k with length q_k are prefixes is not possible and, if there is an element $a_s = 1$ $(s \geq 2)$ in the CF expansion of a, we have $i_a(s+1) = i_a(s) + 2$, so $q_{i_a(s+1)-1} = q_{i_a(s)+1} > q_s$. □

We have just shown that, for each $a \in]0,1[\setminus \mathbf{Q}$ and each $k \in \mathbf{N}^+$, our k^{th} prefix of $s'(a) = 1c(a)$ has the same length or is longer than Shallit's k^{th} prefix X_k of $c(a)$. The words are formed more quickly according to our method. Now we will show that our advantage (expressed by quotient) can be arbitrarily large.

Proposition 4. *For the methods from Theorems 3 and 5 we have the following:*

$$\forall \ (E_n)_{n \geq 2} \quad \exists \ a \in]0,1[\setminus \mathbf{Q} \quad \forall \ k \geq 2 \qquad |P_k| \geq E_k \cdot |X_k|, \tag{8}$$

where $(E_n)_{n \geq 2}$ is any infinite sequence of positive (large) numbers, P_k is our k^{th} prefix of $s'(a) = 1c(a)$ and X_k is Shallit's k^{th} prefix of $c(a)$.

Proof. Let $(E_n)_{n \geq 2}$ be any sequence of (large) positive numbers. We will show how to construct by induction a slope $a = [0; a_1, a_2, \ldots]$ fulfilling (8). We take any $a_1 \in \mathbf{N}^+$ and $a_2 = 1$. Because $a_2 = 1$, then, for every $k \geq 2$, we have $i_a(k+1) \geq k+2$. In the induction step, when we already have defined $a_1, \ldots, a_k$ for some $k \geq 2$, thus also have $q_1, \ldots, q_k$, we define a_{k+1} in order to get $|P_k|/|X_k| \geq E_k$.
According to Corollary 1 and Theorem 1 (the sequence $(q_n)_{n \in \mathbf{N}^+}$ is increasing), we have $|P_k| \geq |S_k| = q_{i_a(k+1)-1} \geq q_{k+1} = a_{k+1}q_k + q_{k-1}$, and, from Theorem 5, $|X_k| = q_k$, so we have $|P_k|/|X_k| \geq (a_{k+1}q_k + q_{k-1})/q_k \geq a_{k+1}$. This means that $|P_k|/|X_k| \geq E_k$ if $a_{k+1} \geq E_k$, so we can take for example $a_{k+1} = \lceil E_k \rceil$. □

Slopes with only one element equal to 1 in the CF expansion can already give us as large an advantage as we define *a priori*. It should be possible to get much better results for the slopes as in Example 2, where the quotient $|P_k|/|X_k|$ is equal to q_{2k-1}/q_k for all $k \in \mathbf{N}^+$. The following lemma (cf. [2] , p. 13) helps us perform further comparisons between our method and the method of Shallit.

Lemma 1. *Let $a = [0; a_1, a_2, \ldots]$. For all $k \geq 2$ we have $q_{2k-1} \geq 2^{\frac{k-2}{2}} q_k$, where q_n for $n \geq 2$ is the denominator of the n^{th} convergent of the CF expansion of a.*

Proof. For $k = 2$ we get $q_3 \geq q_2$, which is true. From Theorem 1 and because the sequence $(q_n)_{n\in\mathbf{N}^+}$ is increasing, we have $q_{4j+1} = a_{4j+1}q_{4j} + q_{4j-1} \geq q_{4j} + q_{4j-1} \geq 2q_{4j-1}$ for $j \geq 1$. Successive application of this inequality yields

$$q_{4j+1} \geq 2^s q_{4j-(2s-1)} \quad \text{for} \quad s = 1, 2, \ldots, 2j. \tag{9}$$

We put $s = j$ in (9) and we get $q_{2k-1} \geq 2^{\frac{k-1}{2}} q_k$, thus $q_{2k-1} \geq 2^{\frac{k-2}{2}} q_k$, for odd k. From Theorem 1 and (9), $q_{4j+3} = a_{4j+3}q_{4j+2} + q_{4j+1} \geq q_{4j+2} + q_{4j+1} \geq 2q_{4j+1} \geq 2 \cdot 2^{j-1} q_{2j+3} \geq 2^j q_{2j+2}$, which gives the statement for even k. □

Theorem 7. *For the slopes a as in Example 2 we have the following:*

- $\forall \ k \geq 2 \quad |P_k| = |X_{2k-1}|$,
- $\forall \ k \geq 2 \quad |P_k| \geq 2^{\frac{k-2}{2}} \cdot |X_k|$,

where P_k is our k^{th} prefix of $s'(a) = 1c(a)$ and X_k is Shallit's k^{th} prefix of $c(a)$. Moreover, for the methods from Theorems 3 and 5, we have the following:

$$\forall \ (E_n)_{n\geq 2} \quad \exists \ a \in]0, 1[\setminus \mathbf{Q} \quad \forall \ k \geq 2 \quad |P_k| \geq E_k \cdot |X_{2k-2}|, \tag{10}$$

where $(E_n)_{n\geq 2}$ is any infinite sequence of positive (large) numbers.

Proof. From Theorem 5, $|X_k| = q_k$ for $k \in \mathbf{N}^+$. From Example 2, $|P_k| = q_{2k-1}$ for $k \geq 2$, which proves the first two statements (for the second one we also use Lemma 1). To prove (10), we take any sequence $(E_n)_{n\geq 2}$ of positive (large) numbers and construct a slope $a = [0; a_1, 1, a_3, 1, a_5, 1, \ldots]$ as in Example 2. We will show how to choose a_{2k+1} for $k \in \mathbf{N}$ in order to get (10) for this $(E_n)_{n\geq 2}$. We proceed as follows. We take any $a_1 \in \mathbf{N}^+$. We choose a_{2k+1} for $k = 1, 2, 3, \ldots$ by induction. When we already have $a_1, \ldots, a_{2k-1}$ for some $k \geq 1$, then we also have $a_2 = \cdots = a_{2k} = 1$ and the denominators of the convergents $q_1, \ldots, q_{2k}$, and we define a_{2k+1} in order to get $|P_{k+1}|/|X_{2k}| \geq E_{k+1}$. Because, according to Example 2 and Theorem 1, $|P_{k+1}| = q_{2k+1} = a_{2k+1}q_{2k} + q_{2k-1}$ and, from Theorem 5, $|X_{2k}| = q_{2k}$, we have $|P_{k+1}|/|X_{2k}| = (a_{2k+1}q_{2k} + q_{2k-1})/q_{2k} \geq a_{2k+1}$, thus $|P_{k+1}|/|X_{2k}| \geq E_{k+1}$ if $a_{2k+1} \geq E_{k+1}$, and we can take for example $a_{2k+1} = \lceil E_{k+1} \rceil$. □

Theorem 7 shows that the advantage of using our method rather than Shallit's (when forming prefixes of $c(a)$ or $s'(a) = 1c(a)$) can be huge for the words with many

1's in the CF expansion of the slope. It is not only possible to get the advantage we choose *a priori*, but we also get arbitrarily longer prefixes in step k compared to Shallit's prefixes in step $2k-2$ for each $k \geq 2$.

A comparison between our method and the method of Venkov is contained in Theorem 8 and Propositions 5 and 6. Theorem 8 is a Venkov counterpart of Proposition 4 and Theorem 7. The statement (11) there is weaker than (12), but it is still worth to be formulated. The reason is that we can reach the advantage formulated in (11) already for slopes with only one CF element equal to 1. It is easy to find slopes fulfilling (11).

Theorem 8. *For each $a \in\,]0,1[\, \setminus \mathbf{Q}$ we have $|P_1| = |C_1|$ and $|P_2| \geq |C_1C_2|$. Moreover, for the methods from Theorems 3 and 4 we have the following:*

$$\forall\ (E_n)_{n\geq 2}\quad \exists\ a \in\,]0,1[\, \setminus \mathbf{Q}\quad \forall\ k \geq 2 \qquad |P_k| \geq E_k \cdot |C_1 \cdots C_k|, \tag{11}$$

$$\forall\ (E_n)_{n\geq 2}\quad \exists\ a \in\,]0,1[\, \setminus \mathbf{Q}\quad \forall\ k \geq 2 \qquad |P_k| \geq E_k \cdot |C_1 \cdots C_{2k-2}|, \tag{12}$$

where $(E_n)_{n\geq 2}$ is any infinite sequence of positive (large) numbers, P_k is our k^{th} prefix of $s'(a) = 1c(a)$ and $C_1 \cdots C_k$ is Venkov's k^{th} prefix of $c(a)$.

Proof. For all $a \in\,]0,1[\, \setminus \mathbf{Q}$ we have $|P_1| = a_1 = |C_1|$. For $k = 2$ we always have $C_1C_2 = q_1 + q_2$ and $|P_2|$ is equal, due to Corollary 1, to $q_2 + q_1$ if $a_2 \neq 1$ and to q_3 if $a_2 = 1$. In the case when $a_2 = 1$ we get $|C_1C_2| = q_1 + q_2 \leq a_3q_2 + q_1 = q_3 = |P_2|$.

To prove (11), we take any sequence $(E_n)_{n\geq 2}$ of (large) positive numbers. We will show how to construct a slope $a = [0; a_1, a_2, \ldots]$ fulfilling (11). The construction will be by induction. We take any $a_1 \in \mathbf{N}^+$ and $a_2 = 1$. Because $a_2 = 1$, then for every $k \geq 2$ we have $i_a(k+1) \geq k+2$. In the induction step, when we already have defined $a_1, \ldots, a_k$ for some $k \geq 2$, thus also have $q_1, \ldots, q_k$, we define a_{k+1} in order to get $|P_k|/|C_1 \cdots C_k| \geq E_k$. From Corollary 1 and Theorem 1, we have $|P_k| \geq |S_k| = q_{i_a(k+1)-1} \geq q_{k+1} = a_{k+1}q_k + q_{k-1}$ and, from Proposition 2, $|C_1 \cdots C_k| = \sum_{i=1}^{k} q_i$. This means that $|P_k|/|C_1 \cdots C_k| \geq (a_{k+1}q_k + q_{k-1})/\sum_{i=1}^{k} q_i \geq (a_{k+1}q_k)/(kq_k) = a_{k+1}/k$, and we get $|P_k|/|C_1 \cdots C_k| \geq E_k$ for $a_{k+1} \geq kE_k$, so we take for example $a_{k+1} = \lceil kE_k \rceil$.

To prove (12), we take any sequence $(E_n)_{n\geq 2}$ of (large) positive numbers. We construct a slope $a = [0; a_1, 1, a_3, 1, a_5, 1, \ldots]$ as in Example 2. We will show how to choose a_{2k+1} for $k \in \mathbf{N}$ in order to get (12) for this $(E_n)_{n\geq 2}$. We proceed as follows. We take any $a_1 \in \mathbf{N}^+$. We choose a_{2k+1} for $k = 1, 2, 3, \ldots$ by induction. Let us say that we already have $a_1, \ldots, a_{2k-1}$ for some $k \geq 1$. Then we also have $a_2 = \cdots = a_{2k} = 1$ and the denominators $q_1, \ldots, q_{2k}$, and we define a_{2k+1} in order to get $|P_{k+1}|/|C_1 \cdots C_{2k}| \geq E_{k+1}$. Because, according to Example 2 and Theorem 1, $|P_{k+1}| = q_{2k+1} = a_{2k+1}q_{2k} + q_{2k-1}$, and, from Proposition 2, $|C_1 \cdots C_{2k}| = \sum_{i=1}^{2k} q_i$, we get $|P_{k+1}|/|C_1 \cdots C_{2k}| = (a_{2k+1}q_{2k} + q_{2k-1})/\sum_{i=1}^{2k} q_i \geq (a_{2k+1}q_{2k})/(2kq_{2k}) = a_{2k+1}/(2k)$, so $|P_{k+1}|/|C_1 \cdots C_{2k}| \geq E_{k+1}$ if $a_{2k+1} \geq 2kE_{k+1}$. We can take for example $a_{2k+1} = \lceil 2kE_{k+1} \rceil$. □

The quotients $|P_k|/|C_1 \cdots C_k|$ and $|P_k|/|C_1 \cdots C_{2k-2}|$ can thus be arbitrarily large. The strongest result is (12), but (11) is the easiest one to reach.

Proposition 4, Theorem 7 and 8 show that, if there are some 1's in the CF expansion of the slope, our method can generate the longest prefixes of all three methods. The greater the number of 1's in the expansion, the greater advantage we get using our method. Because, from Def. 1, $k \leq i_a(k) \leq 2k-2$ for $k \geq 2$ for each $a \in \,]0,1[\, \setminus \mathbf{Q}$, slopes as in Example 2 can probably give us the largest possible advantage, depending on the choice of a_{2n+1} for $n \in \mathbf{N}^+$.

Also slopes $a = [0; a_1, a_2, 1, a_4, 1, a_6, \ldots]$ with $a_2 \geq 2$ give us a similar result. For the lines with such slopes we have $i_a(k) = 2k-3$ and $P_k = q_{2k-1}$ for $k \geq 3$.

Proposition 5. *Let $a = [0; a_1, a_2, a_3, a_4, \ldots]$, where $a_n \geq 2$ for all $n \geq 2$. Then $|P_1| = |C_1|$, $|P_2| = |C_1C_2|$ and $|P_k| < |C_1 \cdots C_k|$ for each $k \geq 3$, where P_k is our k^{th} prefix of $s'(a) = 1c(a)$ and $C_1 \cdots C_k$ is Venkov's k^{th} prefix of $c(a)$.*

Proof. From Proposition 2 and Example 3. For $k = 1$ and $k = 2$ we have clearly the above equality. If $k \geq 3$, then $|C_1 \cdots C_k| = q_1 + \cdots + q_k > q_{k-1} + q_k \geq |P_k|$. □

As we have seen in Proposition 5, it can easily happen that $|C_1 \cdots C_k| > |P_k|$ for some $a \in \,]0,1[\, \setminus \mathbf{Q}$ and $k \geq 3$. For the slopes as in Example 3, Venkov's prefixes for $k \geq 3$ are longer than ours. It is not possible, though, to make the quotient $|C_1 \cdots C_k|/|P_k|$ arbitrarily large, as it was in the opposite case (Theorem 8).

Proposition 6. *Let $a \in \,]0,1[\, \setminus \mathbf{Q}$. Then $|C_1 \cdots C_k| < k \cdot |P_k|$ for $k \geq 3$, where P_k is our k^{th} prefix of $s'(a) = 1c(a)$ and $C_1 \cdots C_k$ is Venkov's k^{th} prefix of $c(a)$.*

Proof. Let $k \geq 3$. It follows from Proposition 2, Corollary 1, and the fact that $(q_n)_{n \in \mathbf{N}^+}$ is strictly increasing, that $|C_1 \cdots C_k| = \sum_{i=1}^{k} q_i < kq_k \leq k \cdot |P_k|$. □

The quotient $|C_1 \cdots C_k|/|P_k|$ is thus bounded by k for each $k \geq 3$.

6 Conclusions and Some Topics for Future Research

We have presented a run-hierarchical CF based description of upper mechanical words $s'(a)$ with slope $a \in \,]0,1[\, \setminus \mathbf{Q}$ and intercept 0. We expressed the length of the prefixes obtained according to our method by the denominators of the convergents of the CF expansion of the slope. This allowed us to compare our result with other CF based methods (Venkov's, Shallit's) of forming such words. Due to the special treatment of the CF elements equal to 1, our method gives often longer prefixes after the same number of steps compared to the two other methods.

Our description uses an auxiliary function, the index jump function defined in Def. 1, while the two other methods do not use any extra functions. However, the index jump function is extremely simply constructed and computationally trivial. Another possible drawback of the method could be that it uses more elements of the CF expansion of the slope than the other methods, so the comparison might conceivably be thought of as being unfair. The run-hierarchical method presented in this paper is not meant to replace the existing methods, it should rather be seen as an additional possible method, which gives better results in some cases, as shown for example in Theorems 7 and 8. Moreover, as the author showed in [9] with numerous examples, in case of quadratic irrationals or even some transcendental numbers (like for example $\sqrt[n]{e}-1$ for $n \geq 2$, $\frac{e^2-1}{e^2+1}$), our method gives a compact description of all the runs with the knowledge of the CF elements which form the period (or, in case of the mentioned transcendental numbers, the knowledge of the periodic form of the CF expansion) and then it does not matter any longer that we use CF elements with a large index.

Corollary 1, together with Proposition 2 and Theorem 5, also shows that our method is the only one of the three presented methods which reflects the hierarchy of runs on all the levels. The run-hierarchical description enables us to analyze abstract properties of lines (words), which has been discussed in another paper of the author [10]. We have shown there how we can partition digital lines (upper mechanical words) with slopes $a \in \,]0,1[\, \setminus \mathbf{Q}$ into equivalence classes under two equivalence relations defined by means of CFs, based on the description from [9]. Hopefully these partitions can help us gain a better understanding of digital lines and maybe become a useful tool for combinatorics on words. Further work in this field could involve a fixed point theorem for Sturmian words and how to find a Sturmian word such that its letters are coding its own run hierarchical structure as defined in the presented paper. Words like this could be called *words with self-balanced construction*. It would be interesting to express the fixed points described above in terms of generalized balances introduced by I. Fagnot and L. Vuillon in [1].

Acknowledgments. I am grateful to Christer Kiselman for comments on earlier versions of the manuscript.

References

1. Fagnot, I.; Vuillon, L.: Generalized balances in Sturmian words. Technical Report 2000-02, Liafa (2000)
2. Khinchin, A. Ya.: Continued Fractions. Dover Publications, third edition (1997)
3. Klette, R.; Rosenfeld, A.: Digital straightness – a review. Discr. Appl. Math. **139** (1–3) (2004) 197–230
4. Lothaire, M.: Algebraic Combinatorics on Words, Cambridge Univ. Press (2002)
5. Pytheas Fogg, N.: Substitutions in Dynamics, Arithmetics and Combinatorics. Lecture Notes in Mathematics **1794**, Springer Verlag (2002)
6. Rosenfeld, A.: Digital straight line segments. IEEE Transactions on Computers **c-32**, No. 12, 1264–1269 (1974)
7. Shallit, J.: Characteristic Words as Fixed Points of Homomorphisms. Univ. of Waterloo, Dept. of Computer Science, Tech. Report CS-91-72 (1991)
8. Uscka-Wehlou, H.: Digital lines with irrational slopes. Theoret. Comp. Science **377** (2007) 157–169
9. Uscka-Wehlou, H.: Continued Fractions and Digital Lines with Irrational Slopes. In D. Coeurjolly et al. (Eds.): DGCI 2008, LNCS **4992**, pp. 93–104, 2008.
10. Uscka-Wehlou, H.: Two Equivalence Relations on Digital Lines with Irrational Slopes. A Continued Fraction Approach to Upper Mechanical Words. Submitted manuscript (2008)
11. Venkov, B. A.: Elementary Number Theory. Translated and edited by Helen Alderson, Wolters-Noordhoff, Groningen (1970)