# Ternary Square-Free Partial Words
# with Many Wildcards

Daniil Gasnikov and Arseny M. Shur[(✉)]

Ural Federal University, Ekaterinburg, Russia
legionlon@gmail.com, arseny.shur@urfu.ru

**Abstract.** We contribute to the study of square-free words. The classical notion of a square-free word has a natural generalization to partial words, studied in several papers since 2008. We prove that the maximal density of wildcards in the ternary infinite square-free partial word is surprisingly big: 3/16. In addition, we introduce a related characteristic of infinite square-free words, called flexibility, and find its values for some interesting words and classes of words.

**Keywords:** Partial word · Square-free word · Letter density

## 1 Introduction

Partial words are a natural generalization of "ordinary" words. A partial word is a word with some positions undefined; more formally, a partial word over an alphabet $\Sigma$ is a word over the alphabet $\Sigma \cup \{\diamond\}$, where the *wildcard* symbol $\diamond$ has a special meaning. Namely, when two words are being compared, a wildcard matches any symbol. Thus, the partial word $a\diamond bc$ matches $\diamond c\diamond c$. In the study of partial words, the matching relation replaces equality in such notions as periods, powers, etc. The main feature of the matching relation is its nontransitivity. This makes many problems on partial words hard. For example, the pattern matching problem, studied for partial words since 1974 [7], is at least as hard as the boolean multiplication [10].

Combinatorics of partial words is much younger than their algorithmics; it began with the paper by Berstel and Boasson [2] and subsequent works [4,15]. These and other early papers focused on periodicity properties. The study of avoidability began with the paper [9], which focused mostly on cube-free partial words. A suitable definition of a square-free partial word was proposed in [8] and independently in [5]; in both these papers the existence of infinite ternary square-free partial words with infinite number of wildcards was proved. In fact, it was demonstrated that wildcards in such a word can have nonzero density: the construction from [8] gives a word with the density 1/39. Thus, a natural question arises: *what is the maximum possible density of wildcards in an infinite ternary square-free partial word?* A related question about the minimum $k$ such that any factor of length $k$ of some infinite ternary square-free partial word contains

a wildcard was answered in [3]; the answer is $k = 7$. Moreover, a thorough analysis of the word from [3] shows that the density of wildcards in it is 7/39.

We study the density of wildcards in a more general context. Every square-free partial word can be obtained by taking a square-free word and replacing some of its letters with wildcards. Thus, for infinite square-free words we have a natural characteristic which we call *flexibility*: the maximum density of the set of positions, in which letters can be simultaneously replaced by wildcards preserving square-freeness (we refer to such sets as *wildcard sets*). Our main result is the exact value of the maximal flexibility of infinite ternary square-free words: 3/16. First, we prove that not only density, but even the upper density of a wildcard set for an infinite ternary square-free word cannot exceed 3/16. Second, we construct a square-free word $\mathsf{G}$ (which probably never appeared in the study of square-free words before) with flexibility 3/16. Moreover, the wildcard set for $\mathsf{G}$ is periodic with period 16. Additional results include the flexibility of the Arshon word (1/9) and the Dejean word (2/19), and also a series of "rigid" square-free words, which have no room for wildcards at all. Our technique is based on the encoding of ternary square-free words by the walks in the weighted $K_{33}$ graph. This encoding was proposed by the second author [14] and proved useful in solving different problems on ternary square-free words [11–13].

The further text consists of preliminary Sect. 2, technical Sect. 3, and proofs of the main results in Sect. 4.

## 2   Preliminaries

**Definitions and Notation.** We study words over the ternary alphabet $\Sigma = \{a, b, c\}$; by default, the letters $x, y, z$ denote variable symbols from $\Sigma$. Finite (infinite) words over $\Sigma$ are treated as functions $w : \{1, \ldots, n\} \rightarrow \Sigma$ (resp., $w : \mathbb{N} \rightarrow \Sigma$); the numbers from the domain of such a function are *positions* (in $w$). In this setting *partial words* are partial functions; the *wildcard* symbol $\diamond$ is used to fill undefined positions.

Standard notions of factor, prefix, and suffix are used for both words and partial words. We write $\lambda$ for the empty word, $|w|$ for the length of $w$, $w[i]$ for the $i$th letter of $w$ and $w[i..j]$ for the factor of $w$ occupying the positions $i, i+1, \ldots, j$. A factor $v = w[i..j]$ is referred to as the *occurrence* of $v$ in $w$ at the $i$th position. Two partial words $u$ and $v$ *match* if $|u| = |v|$ and for each $i = 1, \ldots, |u|$ either $u[i] = v[i]$ or at least one of $u[i], v[i]$ is a wildcard.

A finite word $w$ has *period* $p < |w|$ if $w[1..|w|-p] = w[p+1..w]$. The *exponent* $\exp(w)$ of $w$ is the ratio between its length and its minimal period. The *local exponent* $\text{lexp}(w)$ of a finite or infinite word $w$ is the supremum of the exponents of the finite factors of $w$. The *extension* of a factor $v = w[i..j]$ is the factor $u = w[i'..j']$ such that $i' \leq i$, $j' \geq j$, $u$ has period $|v|$ but the factors $w[i'..j'+1]$, $w[i'-1..j']$ has not.

A *square* is a nonempty word of the form $uu$. A word is *square-free* if it has no squares as factors. The set of ternary square-free words (both finite and infinite) is denoted by $\mathsf{SF}$. A *partial square* is a word of the form $uu'$ such that

$u$ matches $u'$. A (partial or not) square is called *p-square* if $|u| = p$. Partial 1-squares of the form $\diamond x$ or $x \diamond$ occur in every partial word $u$ such that $|u| > 1$ and $u$ contains a wildcard, so we regard them as trivial. A partial word containing no 1-squares and no *partial p-squares* for any $p > 1$, is called *square-free*. Ternary square-free infinite partial words exist [5,8]; we write PSF for the partial counterpart of SF.

Words of the form $uv$ and $vu$ are *conjugates*; conjugacy is an equivalence relation. Linking up the ends of a finite word, we obtain a *circular word*. A circular word represents a conjugacy class in an obvious way. The factors of a circular word are just words, so one can speak about square-free circular words.

Let $P \subset \mathbb{N}$ and $d_n = |P \cap \{1, \ldots, n\}|/n$. Then the *density* of $P$ is the limit $d = \lim_{n \to \infty} d_n$ if it exists; otherwise, we speak about *upper* and *lower* density, meaning $\limsup d_n$ and $\liminf d_n$, respectively.

**Basic Properties.** The following basic property of partial words is important.

**Lemma 1.** *If partial words $u$ and $v$ match and $u'$ is obtained from $u$ by replacing a letter with a wildcard, then $u'$ and $v$ match.*

**Lemma 2.** *Let a partial word $u$ be square-free and $u[i] = \diamond$. The partial word $u'$ obtained from $u$ by replacing $u[i]$ with a letter distinct from the adjacent letters is square-free.*

*Proof.* Assume that $u'$ contains a square. It is not a 1-square by construction and it contains $u'[i]$. Then $u$ must contain a square at the same position by Lemma 1. This contradicts the square-freeness of $u$. □

**Proposition 3.** *Every finite or infinite square-free partial word matches some square-free word. In the case of ternary alphabet, such a matching word is unique up to the first and the last letter.*

*Proof.* The existence of a matching square-free word follows by repeated application of Lemma 2. Further, a square-free partial word has no factors of the form $x \diamond x$, because such a factor forms a 2-square with any subsequent/preceding symbol. Thus, the letters adjacent to a wildcard are distinct. If the alphabet is ternary, there is only one possibility to replace this wildcard. □

Due to Proposition 3, any element of PSF can be seen as a word from SF in which the letters in some positions are replaced by wildcards. Let $u \in$ SF, $P \subset \mathbb{N}$. We denote by $u_P$ the partial word obtained by replacing the letters in $u$ at the positions from $P$ by wildcards. We call $P$ a *wildcard set for $u$* if $u_P \in$ PSF. The maximum density of a wildcard set for $u$ is a natural combinatorial characteristic of $u$; we call it *flexibility*[1]. The original question about the density of wildcards can be reformulated as

---

[1] As mentioned above, some sets do not have density; to avoid additional notions we postulate that upper (lower) bounds on flexibility should work for upper (resp., lower) densities of the corresponding wildcard sets.

– *What is the maximum flexibility of an infinite ternary square-free word?*

We give an upper bound in Sect. 4.1 and a matching lower bound in Sect. 4.2. Another natural question is about words of zero flexibility. We say that an infinite word $u \in \mathsf{SF}$ is *rigid* (resp., *almost rigid*) if it has no nonempty wildcard sets (resp., only finite wildcard sets). In Sect. 3.1, we characterize a class of almost rigid words and find a series of rigid words in it.

**Codewalks.** The representation of ternary square-free words described in this subsection was proposed in [14]. These words contain three-letter factors of the form $xyx$, called *jumps* (of one letter over another). Jumps occur quite often: if $u[i..i+2]$ is a jump in $u \in \mathsf{SF}$, then the next jump in $u$ occurs at one of the positions $i+2$, $i+3$, $i+4$. (A jump at position $i+1$ produces a 2-square at position $i$, while no jump up to position $i+5$ leads to a 3-square at position $i+1$.) Moreover, a jump in a word can be uniquely reconstructed from the previous (or the next) jump and the distance between them. Indeed, let $u[i..i+2] = xyx$. If the next jump is in the position $i+2$ (resp., $i+3$, $i+4$), then $u[i..i+4] = xyxzx$ (resp., $u[i..i+5] = xyxzyz$, $u[i..i+6] = xyxzyxy$). Thus,

($\star$) a word $u \in \mathsf{SF}$ can be uniquely reconstructed from the following information: the leftmost jump, its position, the sequence of distances between successive jumps, the number of positions after the last jump (for finite words).

The property ($\star$) allows one to encode square-free words by walks in the weighted $K_{33}$ graph shown in Fig. 1. A word $u \in \mathsf{SF}$ is encoded by the walk visiting the vertices in the order in which jumps occur when reading $u$ left to right. If the leftmost jump occurs in $u$ at position $i > 1$, then we add the edge of length $i-1$ to the beginning of the walk; note that in this case the walk begins with an edge, not a vertex. A symmetric procedure applies to the end of $u$ if $u$ is finite. By ($\star$), we can omit the vertices (except for the first one), keeping just the lengths of edges and marking the "hanging" edges in the beginning and/or the end. Due to symmetry, we can omit even the first vertex, retaining all information about $u$ up to renaming the letters. For example, $u = abcbabcacbacabc$ has the jumps (left to right) $bcb$, $bab$, $cac$, $aca$ and is encoded, according to Fig. 1, by $\underline{1}123\underline{2}$.



**Fig. 1.** The graph of jumps in ternary square-free words. Vertices are jumps; two jumps that can follow each other in a square-free word are connected by an edge of length $i$, where $i$ is the number of positions between the starting positions of these jumps. Due to symmetry, the graph is undirected.

Such a code is called *codewalk* and denoted by $\mathsf{cwk}(u)$. The codewalk $\underline{1}123\underline{2}$ is decoded by any word $xyzyxyzxzyxzyxzxyz$, where $\{x, y, z\} = \{a, b, c\}$. Note that the choice of decoding does not affect the properties concerning periods and squares. A codewalk is *closed* if it marks a closed walk without
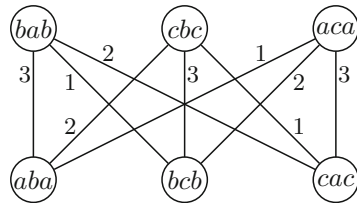
hanging edges in $K_{3,3}$; e.g., 121212 is closed and 1212 is not. For a codewalk $w$ without hanging edges, its *literal length* $\ell(w)$ is the distance between the positions of the last and the first jumps in the decoded word; it can be computed by adding $|w|$ to the sum of digits of $w$. Note that if a codewalk $wv = \mathsf{cwk}(u)$ has period $|w|$ and $w$ is closed, then $u$ has period $\ell(w)$.

Clearly, not all walks in the weighted $K_{33}$ graph encode square-free words.

*Remark 4.* The codewalk 11 is decoded by a word of the form $xyxzxyx$, which is square-free but cannot be extended to a square-free word by any letter. This property is shared by the codewalk 333 and, moreover, by any codewalk of the form $v3v$, where $v3$ is closed. The codewalks of the form $vxv$, where $x \in \{1, 2\}$ and $vx$ is closed, encode words containing squares. The codewalks 223 and 322 decode to square-free words that cannot be extended by any letter to the left (resp., to the right).

A sufficient condition for square-freeness was proved in [14].

**Lemma 5.** *A codewalk having (a) no factors* 11, 222, 223, 322, 333, *and (b) no factors of the form* $vxyv$, *where* $x, y$ *are symbols and the codewalk* $vxy$ *is closed, encodes a square-free word.*

## 3   White and Black Positions

Assume that $u \in \mathsf{SF}$ is fixed. We say that a position $i$ is *white* if it belongs to some wildcard set for $u$ and *black* otherwise. Usually, the set of *all* white positions is not a wildcard set for $u$, because some white positions "interact" in the sense that a wildcard set can contain some of them but not all of them; the simplest interactions are considered in Lemma 10 below.

We start the study of white and black positions with the following criterion.

**Proposition 6.** *Given a fixed* $u \in \mathsf{SF}$, *a position* $i$ *is black if for some factor* $vxv$ *of* $u$, *where* $v \neq \lambda$ *and* $x \in \Sigma$, $i$ *is either the position of* $x$, *or the position preceding the left* $v$, *or the position following the right* $v$. *Otherwise,* $i$ *is white.*

*Proof.* The set of wildcard sets for $u$ is closed downwards by Lemma 1. Then a position $i$ is white if and only if $\{i\}$ is a wildcard set for $u$. Placing a wildcard in a position described in the conditions of the proposition gives us a $(|v|+1)$-square, so all such positions are black. Conversely, let some position $i$ be black and $x = u[i]$. Since $\{i\}$ is not a wildcard set, $u$ has a factor of the form $vxwvyw$, turning into a square when $x$ is replaced by a wildcard. If either $v$ or $w$ is empty, the position of $x$ satisfies the conditions of the proposition. Otherwise, note that $x \neq y$ and $w[1] = v[|v|] = z$, where $z \neq x$, $z \neq y$ since $u$ is square-free. Hence $u[i-1..i+1] = zxz$, and the position of $x$ satisfies the conditions again.     □

*Example 7.* The word $u = abcbabcacb$ has two white positions: 2 and 9. All other positions are black by Proposition 6; e.g., the factor $u[2..4] = bcb$ makes black the positions $1, 3,$ and $5$, while 4 and 8 are black due to $u[1..7] = abc\,b\,abc$. If we consider $u' = u[2..10]$ instead of $u$, the position of the second $b$ (now position 3) will be white. More generally,

($*$) in some words with the prefix $xyxzx$ the position 3 is white.

The distribution of white positions in infinite square-free words (and in their finite factors) is densely related to jumps.

**Proposition 8.** *Let $u \in \mathsf{SF}$ be an infinite word.*
*(1) Every white position in $u$ is either the first or the last position of a jump.*
*(2) Modulo the single exclusion ($*$), every jump in $u$ contains at most one white position, and every white position belongs to exactly one jump.*

*Proof.* The middle position of a jump and a position adjacent to a jump are black by Proposition 6. Since two consecutive jumps are separated by at most one position, statement 1 is proved.

Consider two consecutive jumps in $u$. Depending on the length of the edge between them, they form a factor $v_1 = xyxzx$, $v_2 = xyxzyz$, or $v_3 = xyxzyxy$ (see Fig. 1). The position of the middle $x$ in $v_1$ is black if $v_1$ is not a prefix of $u$ (cf. Example 7); if $v_1$ is a prefix, then we have the exclusion ($*$). Applying Proposition 6 to $v_2$ and $v_3$ we see that all positions except for 1, 6 in $v_2$ and for 3, 5 in $v_3$ are black. Statement 2 now follows.            □

So, the further study of the distribution of the white positions in a word should clarify which jumps contain white positions and which do not. From the proof of Proposition 8 we have the following basic picture (see Fig. 2). We call the positions of question marks *potentially white*.



**Fig. 2.** Black and potentially white positions in the pairs of consecutive jumps. The jumps in the left (resp., middle, right) picture are connected in the $K_{33}$ graph by the edge of length 1 (resp., 2, 3).

Our main interest is in the asymptotic distribution of white positions in infinite words. So we pay little attention to special cases concerning prefixes of these words (like ($*$) and "almost squares" described in Remark 4). We call a factor of a codewalk *regular* if it is not its prefix. The following three lemmas form the basis for the proof of our main results.

**Lemma 9.** *A jump in $u \in \mathsf{SF}$ contains no white position if it is located in the place indicated by a dot in any of the following regular factors of $\mathsf{cwk}(u)$:*

$$1.2,\ 2.1,\ 2.2,\ 3.3,\ 13.,\ .31,\ .1221.,\ .2332., \tag{1}$$

$$.1212.321,\ 123.2121.,\ .13132312.,\ .21323131.,\ 323.1321.,\ .1231.323 \tag{2}$$

*Proof.* For the first four factors in (1), it is enough to look at Fig. 2. In the jump that follows 1 in the codewalk, the last position is potentially white; but if the next edge has length 2, this position is black. The same argument works for $21, 22$ and $33$.

For the remaining factors we use Proposition 6. Decoding each of these factors (together with its unique extension if necessary), we get a factor of $u$ of the form $vxv$; for convenience, the $v$'s are overlined:

$\underline{1}13 \;\; \to \overline{zx\dot{y}x\overline{zxy}}\boldsymbol{z}xz$

$31\underline{1} \;\; \to xy\boldsymbol{x}\overline{zyxyzy}\overline{x}$

$1221 \to \boldsymbol{x}\overline{yxzx\dot{y}zyxzxy}\boldsymbol{x}$

$2332 \to \boldsymbol{x}\overline{yxzyzxyzyxzyzxy}\boldsymbol{x}$

$1212321\underline{2} \to \boldsymbol{x}\overline{yxzxyzyxyzxz}\boldsymbol{z}\overline{yxzxyzyxyzx}$

$\underline{2}1232121 \to \overline{xzyxyzyxzxy}\boldsymbol{z}\overline{xzyxyzyxzxy}\boldsymbol{x}$

$13132312 \to \boldsymbol{x}\overline{yxzxyzxzyzxyzyxzxyzxzyzxy}\boldsymbol{x}$

$21323131 \to \boldsymbol{x}\overline{yxzyzxzyxzxyzyxzyzxzyxzxy}\boldsymbol{x}$

$\underline{1}3231321 \to \overline{xzyzxyzyxzxyz}\boldsymbol{z}\overline{xzyzxyzyxzxy}\boldsymbol{x}$

$12313231\underline{} \to \boldsymbol{x}\overline{yxzxyzyxzyzxz}\boldsymbol{z}\overline{yxzxyzyxzyzx}$

In the jumps indicated by dots, the potentially white positions are those with boldface letters (see Fig. 2); all these positions are black by Proposition 6. $\square$

**Lemma 10.** *For every $u \in \mathsf{SF}$ and every $i > 1$, a wildcard set for $u$ cannot contain simultaneously $i$ and $i+5$, or $i$ and $i+6$.*

*Proof.* Assume that both positions $i$ and $i+5$ are white (otherwise, there is nothing to prove). Examining the location of white positions in jumps (Fig. 2), we see the only possibility: these positions are located in consecutive jumps connected by an edge of length 2 (the middle picture). This 2 in the codewalk gives a factor of the form $zxyxzyzx$ (note that $i > 1$, so the initial $z$ exists). Placing wildcards in both white positions, we get a partial 4-square $z\diamond yx\, zy\diamond x$.

A slightly longer analysis shows that the white positions $i$ and $i+6$ are always in the jumps connected by the codewalk 33: they cannot be connected by 11 by Remark 4 and by 13 or 31 by Lemma 9. The codewalk 33 gives us a factor of the form $xyxzyxyzxyx$. If we place wildcards in both white positions, we get a partial word starting with a partial 5-square $xy\diamond zy\, xyz\diamond y$. $\square$

**Lemma 11.** *Let $P$ be a wildcard set for an infinite word $u \in \mathsf{SF}$, $j$, $j+k \in P$ and $j+1, \ldots, j+k-1 \notin P$. Then (1) $k \notin \{1, 3, 5, 6, 8\}$, (2) if $k = 2$ (resp., $k = 4$; $k = 7$) then $j$, $j+k$ are located in jumps connected by 3 (resp., by 1; by one of the paths 12, 21, 23, or 32) in $\mathsf{cwk}(u)$.*

*Proof.* The statement readily follows from Proposition 8 and Fig. 2 for $k = 1, 2, 3, 4$ and from Lemma 10 for $k = 5, 6$. Further, it is easy to check that the distance between potentially white positions in jumps connected by a codewalk of length 3 is at least 9. Thus, for $k = 7, 8$ the jumps containing $j$ and $j+k$ are connected by a two-edge codewalk. This codewalk is not 11 by Remark 4, not 13 or 31 by Lemma 9, and not 33 by Lemma 10. If it equals $12, 21, 23$, or $32$, we have $k = 7$. If it equals 22, we have $k = 8$. But 22 is followed by 1 in $\mathsf{cwk}(u)$, which means that the position $j+k$ is black by Lemma 9, contradicting the condition $j+k \in P$. Hence, $k \neq 8$. The lemma is proved. $\square$

### 3.1 Rigid Words

Lemma 9 implies the following result about almost rigid words.

**Proposition 12.** *Let $u \in \mathsf{SF}$ be an infinite word. If $\mathsf{cwk}(u)$ contains finitely many 3's, then $u$ is almost rigid.*

It is not a priori clear whether a word required in Proposition 12 (or, equivalently, an infinite word $u \in \mathsf{SF}$ such that $\mathsf{cwk}(u)$ contains no 3's) exists. Fortunately, this is the case. Consider the alphabet $\{1, 2\}$ and take the *Fibonacci word* $\mathsf{F}$ which is the fixed point of the morphism $\phi$:

$$\phi(2) = 21, \ \phi(1) = 2; \qquad \mathsf{F} = 2122121221221 \cdots$$

The *1-2-bonacci word* is the ternary word $\mathsf{F}_{12} = ab \cdots$ such that $\mathsf{cwk}(\mathsf{F}_{12}) = \mathsf{F}$. As was proved in [13], $\mathrm{lexp}(\mathsf{F}_{12}) = 11/6$. This means the existence of an infinite set of rigid square-free words, as the next proposition shows.

**Proposition 13.** *Any suffix $u$ of the 1-2-bonacci word such that $\mathsf{cwk}(u) = \underline{1}1221 \cdots$ is a rigid square-free word.*

*Proof.* By Proposition 8(2), every jump in $u$ has at most one white position: since $u = xyz \cdots$, we avoid the case $(*)$. By Lemma 9, all positions in the jumps are black. The first position of $u$ precedes a jump, so it is also black.     □

## 4   Proofs of Main Results

### 4.1 Upper Bound on Flexibility

**Theorem 14.** *The flexibility of any infinite word $w \in \mathsf{SF}$ is at most 3/16.*

*Proof.* Let $P$ be any infinite wildcard set for $w$. We aim at building a factorization $w = u_0 u_1 \cdots u_n \cdots$ such that the lengths of all factors $u_i$ are bounded and $u_i$ contains $p_i \leq \frac{3|u_i|}{16}$ positions from $P$ for any $i > 0$. The existence of such a factorization implies the upper bound 3/16 on the upper density of $P$, thus proving the theorem.

We factorize $w$ greedily from left to right, checking that each factor $u_i$, $i > 0$, satisfies the conditions $|u_i| \leq 22$, $p_i \leq \frac{3|u_i|}{16}$, and begins with a position from $P$ whenever $p_i > 0$. To define the position of $u_1$, consider the third from the left jump in $w$. By Proposition 8, it contains at most one position from $P$. If it contains a position from $P$, $u_1$ begins at this position; otherwise, it begins at any position of this jump. Now assume that all factors up to $u_{i-1}$ are built and $u_i$ begins at the $j$th position of $w$. We should define $k = |u_i|$. Let $l < l' < l'' < l'''$ be the first four positions from $P$ on the right of $j$. If $j \notin P$, then put $k = \min\{22, l-j\}$. If $j \in P$, the choice of $k$ depends on the distances between $j, l, l', l'', l'''$ and is described in Table 1. In all possible cases $u_i$ satisfies the prescribed conditions; the desired factorization is thus constructed.     □

**Table 1.** Choosing the length of $u_i$ in the proof of Theorem 14. The possible distances between consecutive positions from $P$ as well as the corresponding fragments of $\mathsf{cwk}(w)$ are taken from Lemma 11. For impossible sets of distances, the contradictions are given.

| $l-j$ | $l'-l$ | $l''-l'$ | $l'''-l''$ | Fragment of $\mathsf{cwk}(w)$ | $k$ | $p_i$ | $\frac{3k}{16}$ |
|---|---|---|---|---|---|---|---|
| $\geq 7$ | any | any | any | irrelevant | $\min\{22, l-j\}$ | 1 | $\geq \frac{21}{16}$ |
| 4 | $\geq 7$ | any | any | irrelevant | $\min\{22, l'-j\}$ | 2 | $\geq \frac{33}{16}$ |
| 4 | 4 | any | any | $\begin{smallmatrix} j & l & l' \\ |1|1| \end{smallmatrix}$ | $w \notin \mathsf{SF}$ by Remark 4 | | |
| 4 | 2 | any | any | $\begin{smallmatrix} j & l & l' \\ |1|3| \end{smallmatrix}$ | $l' \notin P$ by Lemma 9 | | |
| 2 | $\geq 9$ | any | any | irrelevant | $\min\{22, l'-j\}$ | 2 | $\geq \frac{33}{16}$ |
| 2 | 7 | $\geq 7$ | any | irrelevant | $\min\{22, l''-j\}$ | 3 | $\geq 3$ |
| 2 | 7 | 4 | $\geq 9$ | irrelevant | 22 | 4 | $\frac{66}{16}$ |
| 2 | 7 | 4 | 7 | $\begin{smallmatrix} j & l & l' & l'' & l''' \\ |3|R|1|Q| \end{smallmatrix}$ | $R \in \{12, 32\}$: $l' \notin P$ by Lemma 9 | | |
| | | | | | $Q \in \{21, 23\}$: $l'' \notin P$ by Lemma 9 | | |
| | | | | | $R = 21$ or $Q = 12$: $w \notin \mathsf{SF}$ by Remark 4 | | |
| | | | | | $R = 23$, $Q = 32$: $l' \notin P$ by Lemma 9(2) | | |
| 2 | 7 | 4 | 4 | $\begin{smallmatrix} j & l & l' & l'' & l''' \\ |3|R|1|1| \end{smallmatrix}$ | $w \notin \mathsf{SF}$ by Remark 4 | | |
| 2 | 7 | 4 | 2 | $\begin{smallmatrix} j & l & l' & l'' & l''' \\ |3|R|1|3| \end{smallmatrix}$ | $l''' \notin P$ by Lemma 9 | | |
| 2 | 7 | 2 | any | $\begin{smallmatrix} j & l & l' & l'' \\ |3|R|3| \end{smallmatrix}$ | $j$ [resp. $l; l'; l''$] $\notin P$ by Lemma 9 | | |
| | | | | | for $R = 12$ [resp. 32; 23; 21] | | |
| 2 | 4 | any | any | $\begin{smallmatrix} j & l & l' \\ |3|1| \end{smallmatrix}$ | $j \notin P$ by Lemma 9 | | |
| 2 | 2 | any | any | $\begin{smallmatrix} j & l & l' \\ |3|3| \end{smallmatrix}$ | $l \notin P$ by Lemma 9 | | |

## 4.2 Word of Maximal Flexibility

Return to the 1-2-bonacci word from Sect. 3.1 and consider the codewalk $\mathsf{H} = \eta(\mathsf{F}_{12})$, where the morphism $\eta : \Sigma^* \to \{1, 2, 3\}^*$ is defined by

$$\eta(a) = 1232\,132323$$
$$\eta(b) = 1232\,13232\,132323$$
$$\eta(c) = 1232\,132323\,12323$$

**Theorem 15.** *The word* $\mathsf{G} = ab \cdots$ *with the codewalk* $\mathsf{H}$ *is square-free and has flexibility 3/16.*

*Proof.* To prove square-freeness of $\mathsf{G}$, it suffices to show that $\mathsf{H}$ satisfies the conditions of Lemma 5. The condition (a) is obviously satisfied; let us check (b). Note that any conjugate of a closed codewalk is closed; hence if $g$ is a codewalk with period $p$ and some factor of $g$ of length $p$ is a closed codewalk, then all such factors are closed. According to the condition (b), our aim is to prove that for any closed codewalk $h$ its extension in $\mathsf{H}$, denoted below by $g$, has length $< 2|h| - 2$. Since $K_{3,3}$ is bipartite, closed codewalks have even lengths.

It is easy to check (b) for closed walks of length 4 (1232 and 1323) and 6 (no such codewalks in H). So let $|h| \geq 8$. Assume to the contrary that $|g| \geq 2|h| - 2$. Then $g$ has a factor of length $|h|$ beginning with 1. So we assume w.l.o.g. that $h$ begins with 1. We call the codewalks $1232, 12323, 13232, 132323$ *miniblocks*; they constitute the blocks $\eta(a), \eta(b), \eta(c)$. Let $h = u_1 \cdots u_n u_{n+1}$, where $u_1, \ldots, u_n$ are miniblocks, while $u_{n+1}$ is a prefix of a miniblock but not a miniblock itself. Then $h$ is followed in H by a symbol distinct from $1 = h[1]$. Hence $g$ ends at the same position in H as $h$. On the other hand, $g$ extends $h$ to the left by less than $|u_{n+1}| < 6$ symbols. Since $|h| \geq 8$, this contradicts our assumption on $|g|$. Therefore, $h$ is a product of miniblocks.

To know which codewalks are closed, we partition them into six *types*: $\lambda$, $1, 2, 3, 12,$ and $13$. A codewalk $u$ has type $t = \mathsf{type}(u)$ if the paths in $K_{3,3}$ with a common starting point and labels $u$ and $t$, respectively, have a common endpoint. In particular, the codewalks of type $\lambda$, and only they, are closed. The concatenation of codewalks has the same type as the concatenation of their types, and the latter can be easily computed by Fig. 1. For miniblocks, one has $\mathsf{type}(1232) = \lambda$, $\mathsf{type}(12323) = 3$, $\mathsf{type}(13232) = 2$, $\mathsf{type}(132323) = 12$.

A direct check of types of concatenations of miniblocks shows that $h$, which is closed, is long enough; in particular, $h$ contains at least two occurrences of the miniblock 1232, each followed by 1 (not by 3). Then some factor of $g$ of length $|h|$ starts with the leftmost factor 12321 in $g$ (otherwise $g$ would be too short); we assume w.l.o.g. that $h$ begins with this leftmost 12321. Then $h = \eta(w)u$, where $w \neq \lambda$ and $u$ is a proper prefix of a block, but not a block. If $u$ is nonempty, $h$ is not followed by 12321 in H, so $g$ extends $h$ to the right by at most four symbols. At the same time, $g$ extends $h$ to the left by less than $|u|$ symbols. This contradicts our assumption on $|g|$. Therefore, $h = \eta(w)$ is a product of blocks. Note that $\mathsf{type}(\eta(a)) = 12$, $\mathsf{type}(\eta(b)) = 3$, $\mathsf{type}(\eta(c)) = 2$. Since $h$ is closed, $|w| \geq 3$; e.g., $\mathsf{type}(\eta(abc)) = \lambda$. Let $w = x_1 \cdots x_n$.

By the choice of $h$ and square-freeness of $\mathsf{F}_{12}$, the extension of $w$ in $\mathsf{F}_{12}$ equals $x_1 \cdots x_n x_1 \cdots x_{n-i}$ for some $i \geq 1$. Hence this extension occurs in $\mathsf{F}_{12}$ inside the factor $\hat{w} = \bar{x} x_1 \cdots x_n x_1 \cdots x_{n-i} \hat{x}$, where $\bar{x} \neq x_1, x_n$; $\hat{x} \neq x_{n-i}, x_{n-i+1}$. Then $|g| = 2|h| - |\eta(x_{n-i+1} \cdots x_n)| + M + N$, where $M$ is the length of the common suffix of $\eta(\bar{x})$ and $\eta(x_n)$, while $N$ is the length of the common prefix of $\eta(\hat{x} \cdots)$ and $\eta(x_{n-i+1} \cdots)$. One has $M = 9$ for the pair $(a, b)$ and $M = 4$ otherwise; $N = 14$ for the pair $(a, c)$ and $N = 9$ otherwise. If $i > 2$, then clearly $|g| < 2|h| - 2$; hence $i \leq 2$.

Case $i = 1$. Since $\mathsf{lexp}(\mathsf{F}_{12}) = 11/6$, we have $|w| \leq 6$. Note that $\mathsf{F}_{12}$ has no factors of the form $xyzxy$, because $\mathsf{F}$ contains no 3's (this is why we have chosen $\mathsf{F}_{12}$ as the argument of $\eta$); hence $|w| \geq 4$. It is easy to check that the only candidates to $w$ have the form $xyzy$ or $xyzyxz$, but then $\eta(w)$ is not closed for any values of $x, y,$ and $z$.

Case $i = 2$. We have $M = 9$, $N = 14$, $|\eta(x_{n-1}x_n)| = 25$, and, respectively, $|g| = 2|h| - 2$. Then $\bar{x} \in \{a, b\}$, $\hat{x} \in \{a, c\}$, $a \in \{x_{n-1}, x_n\}$. By square-freeness of $\mathsf{F}_{12}$ we get

$$\hat{w} = b\,c\,\underline{u}\,bca\,c\,u\,b\,a \quad \text{or} \quad \hat{w} = a\,\underline{cu}\,bab\,c\,u\,b\,c,$$

where $w$ is underlined, $u \in \Sigma^*$. We have $\exp(\hat{w}) \leq 11/6$ and then $|w| \leq 12$. It appears that the only possibility for $u$, giving a square-free word $\hat{w}$ such that $\mathsf{cwk}(\hat{w})$ contains no 3's, is $u = ba$ (resp., $u = ca$) for the left (resp., right) case. But then $\eta(w)$ is not closed. This finishes the proof of condition (b); so $\mathsf{G}$ is square-free by Lemma 5.

To find the flexibility of $\mathsf{G}$, we need a technical lemma.

**Lemma 16.** *All factors of $\mathsf{G}$ of the forms $vxv$ and $vxyv$ have periods $\leq 10$.*

*Proof.* Consider a factor $w = vuv$ of $\mathsf{G}$ with the minimal period $p = |vu| \geq 11$ and minimal possible $|u|$. Then $w$ is the extension of $vu$. Aiming at a contradiction, assume that $|u| \leq 2$. Consider the codewalk $w'$ starting at the leftmost jump and ending at the rightmost jump in $w$. Since $v$ is long enough to contain at least two jumps, the walk between the leftmost and the rightmost jump in $v$ repeats twice, so $w' = v'u'v'$, where $v'u'$ is closed and $\ell(v'u') = p$. Now compute $|w|$. By the definition of extension, the left and right $v$ in $w$ are preceded (and also followed) in $\mathsf{G}$ by different letters. Hence the first (resp., last) jump in $w$ is preceded (resp., followed) by at most two letters. The length of the last jump is 3, and the remaining part of $w$ has length $\ell(v'u'v')$. Thus, $|w| \leq 2p - \ell(u') + 7$. Since $w'$ is a factor of $\mathsf{H}$, we know that $|u'| \geq 3$ by condition (b). Then $\ell(u') \geq 9$, where the equality takes place for $u' = 123, 132, 213, 231, 312, 321$. If $\ell(u') \geq 10$, we obtain $|w| \leq 2p - 3$ and $|u| \geq 3$, contradicting our assumption. Let $\ell(u') = 9$. Then $v'[|v'|] \in \{2, 3\}$. Since $\mathsf{H}$ has no factors 22 and 33, either $u[1]$ or the symbol following $w'$ in $\mathsf{H}$ is 1. Hence the last jump in $w$ is followed by just one letter, not two; we again have $|w| \leq 2p - 3$, contradicting the assumption $|u| \leq 2$.     □

By Lemma 16 and Proposition 6, the potentially white positions in all jumps in $\mathsf{G}$, except for those described in Lemma 9(1), are white. The set $P$ of all white positions in $\mathsf{G}$ is periodic with period 16 and has density $3/16$. Indeed, white positions in the jumps connected by 3 (resp., $12, 23, 21, 32$) in the codewalk, are at distance 2 (resp., 7) by Lemma 11:

$$\mathsf{G} = 12{}^|3{}^|21{}^|32{}^|3{}^|23{}^|12{}^|3{}^|21{}^|32{}^|3{}^|21{}^|32{}^|3{}^|23{}^|12{}^|3{}^|21{}^|32{}^|3{}^|23{}^|12{}^|3{}^|23{}^| \cdots$$

It remains to prove that $P$ is a wildcard set. Assume to the contrary that the partial word $\mathsf{G}_P$ contains a partial square $vv'$, $m = |v|$. A direct check shows that $m > 10$. We transform $vv'$, whenever possible, replacing back wildcards with the letters from the same positions of $\mathsf{G}$. The replacement rules are as follows. Consider all pairs $(v[i], v'[i])$. If one symbol is $x$ and the other is a wildcard obtained from $x$, replace the wildcard; if both are wildcards obtained from the same $x$, replace both; if both are wildcards obtained from different letters, replace one of them. Let $uu'$ be the resulting partial square. At least one of the words $u[2..m-1]$, $u'[2..m-1]$ contains a wildcard; otherwise $\mathsf{G}$ has the factor $u[2..m-1]u[m]u'[1]u[2..m-1]$ with period $m$, contradicting Lemma 16. W.l.o.g., $u[i]$ is a wildcard obtained from a letter $z$ and $1 < i < m$. Then $u'[i] = y \neq z$. One of the letters $u[i-1], u[i+1]$ is $y$; the corresponding letter of $u'$ cannot be $y$, hence it is a wildcard. Thus, the wildcard $u[i]$ matches a letter adjacent to

another wildcard. This condition is quite restrictive. A case analysis shows that
only short factors of $\mathsf{G}_P$ match under this condition, and the square $uu'$ cannot
exist. This finishes the proof of square-freeness of $\mathsf{G}_P$ and thus of the theorem.

Some details of the case analysis follow. Up to symmetry, there are two
possible fragments of $\mathsf{G}$ that can contain the position of $u[i]$ (this position is
indicated by a wildcard replacing $z$):

$2^|32$:     $\underline{x\ y\ x\ z\ y\ \diamond\ x\ \boldsymbol{y}\ z\ y\ x\ z\ x}$
$\qquad\qquad\boldsymbol{y}\ z\ \boldsymbol{x}\ y\ x\ z$
$\qquad\qquad\quad\boldsymbol{z}\ y\ x\ z$
$\qquad\quad y\ x\ z\ \boldsymbol{x}\ y\ \boldsymbol{z}\ x\ z\ y\ x$
$\qquad\qquad x\ z\ y\ x\ \boldsymbol{y}\ z\ \boldsymbol{x}\ y\ x\ z$
$\qquad\qquad\quad z\ y\ x\ \boldsymbol{z}\ y\ z$
$\qquad\qquad\qquad\boldsymbol{z}\ x\ \boldsymbol{y}$

$21^|3$:     $\underline{z\ \boldsymbol{x}\ y\ x\ z\ y\ z\ x\ \diamond\ y\ x\ z\ x}$
$\qquad\qquad z\ y\ x\ z\ y\ z\ x\ y\ \boldsymbol{x}$
$\qquad\qquad\qquad z\ x\ y\ \boldsymbol{z}\ x\ z$
$\qquad\qquad\qquad\quad z\ y\ \boldsymbol{x}$
$\qquad\qquad\qquad\quad z\ \boldsymbol{y}\ x\ y$
$\qquad\qquad\qquad\quad \boldsymbol{z}\ x\ y$
$\qquad\qquad\qquad y\ z\ \boldsymbol{y}\ x\ \boldsymbol{z}$

Below each line, the maximal factors of $\mathsf{G}$ that match the top word are given;
here all boldface symbols occupy white positions and can be replaced by wild-
cards. These factors are computed to satisfy both Fig. 2 and the structure of
the codewalk $\mathsf{H}$. For example, the factor $w = yxz\boldsymbol{x}y\boldsymbol{z}xzyx$ in the left column
contains two white positions at distance 2, so the corresponding jumps are sur-
rounded by 2's in $\mathsf{H}$: $2^|3^|2$. Then $w$ is preceded by $z$ and followed by $y$, the letters
mismatching their counterparts in the top word.                                    □

### 4.3   Morphic and Substitutional Flexible Words

The flexibility of well-known square-free words is of certain interest. The proofs
of the next results are similar to Theorem 15 and omitted due to space con-
straints. The result of Theorem 17 is the best we have for purely morphic words.

**Theorem 17.** *The Dejean word [6] has flexibility 2/19.*

**Theorem 18.** *The Arshon word [1] has flexibility 1/9.*

## 5   Conclusion and Open Problems

The problem of finding the maximum density of wildcards in a ternary infinite
square-free partial word can be conveniently reformulated in terms of placing
wildcards at some positions in square-free words. We developed a technique to
find the appropriate sets of positions (wildcard sets) and define flexibility of a
square-free word as the maximum density of its wildcard set. We proved that
the maximum flexibility of a ternary square-free word is 3/16. Besides that, we
proved the existence of rigid words, having no positions for wildcards at all. Two
open problems can direct further development of this topic:

1. What is the maximum flexibility of a morphic/purely morphic ternary square-
   free word?
2. What is the minimum local exponent of a rigid word?

# References

1. Arshon, S.E.: Proof of the existence of asymmetric infinite sequences. Mat. Sbornik **2**, 769–779 (1937). in Russian, with French abstract
2. Berstel, J., Boasson, L.: Partial words and a theorem of Fine and Wilf. Theoret. Comput. Sci. **218**, 135–141 (1999)
3. Blanchet-Sadri, F., Black, K., Zemke, A.: Unary pattern avoidance in partial words dense with holes. In: Dediu, A.-H., Inenaga, S., Martín-Vide, C. (eds.) LATA 2011. LNCS, vol. 6638, pp. 155–166. Springer, Heidelberg (2011)
4. Blanchet-Sadri, F., Hegstrom, R.A.: Partial words and a theorem of Fine and Wilf revisited. Theor. Comput. Sci. **270**(1–2), 401–419 (2002)
5. Blanchet-Sadri, F., Mercaş, R., Scott, G.: A generalization of Thue freeness for partial words. Theoret. Comput. Sci. **410**, 793–800 (2009)
6. Dejean, F.: Sur un théorème de Thue. J. Combin. Theory. Ser. A **13**, 90–99 (1972)
7. Fischer, M., Paterson, M.: String matching and other products. SIAM-AMS Proc. **7**, 113–125 (1974)
8. Halava, V., Harju, T., Kärki, T.: Square-free partial words. Inform. Process. Lett. **108**(5), 290–292 (2008)
9. Manea, F., Mercaş, R.: Freeness of partial words. Theoret. Comput. Sci. **389**(1–2), 265–277 (2007)
10. Muthukrishnan, S., Ramesh, H.: String matching under a general matching relation. In: Shyamasundar, R.K. (ed.) FSTTCS 1992. LNCS, vol. 652, pp. 356–367. Springer, Heidelberg (1992)
11. Petrova, E.A., Shur, A.M.: Constructing premaximal ternary square-free words of any level. In: Rovan, B., Sassone, V., Widmayer, P. (eds.) MFCS 2012. LNCS, vol. 7464, pp. 752–763. Springer, Heidelberg (2012)
12. Petrova, E.A., Shur, A.M.: On the tree of ternary square-free words. In: Manea, F., Nowotka, D. (eds.) WORDS 2015. LNCS, vol. 9304, pp. 223–236. Springer, Heidelberg (2015)
13. Petrova, E.A.: Avoiding letter patterns in ternary square-free words. Electr. J. Comb. **23**(1), P1.18 (2016)
14. Shur, A.M.: On ternary square-free circular words. Electronic J. Combinatorics **17**, R140 (2010)
15. Shur, A.M., Konovalova, Y.V.: On the periods of partial words. In: Sgall, J., Pultr, A., Kolman, P. (eds.) MFCS 2001. LNCS, vol. 2136, pp. 657–665. Springer, Heidelberg (2001)