Discrete Mathematics and Theoretical Computer Science
vol. 27:3 #13 (2025)

Words avoiding the morphic images of most of their factors

Pascal Ochem Matthieu Rosenfeld*

LIRMM, CNRS, Université de Montpellier, Montpellier, France.

revisions 23rd June 2025, 8th Sep. 2025, 14th Sep. 2025; accepted 17th Sep. 2025.

We say that a finite factor f of a word w is *imaged* if there exists a non-erasing morphism m, distinct from the identity, such that w contains $m(f)$. We show that every infinite word contains an imaged factor of length at least 6 and that 6 is best possible. We show that every infinite binary word contains at least 36 distinct imaged factors and that 36 is best possible.

Keywords: combinatorics on words

1 Introduction

Following the construction by Fraenkel and Simpson [2] of an infinite binary word containing only the squares `00`, `11`, and `0101`, some results in the literature have shown that infinite binary words can also contain finitely many distinct regularities that are more general than squares. In particular, the following result considers generalizations of squares such that one half of the pseudosquare is allowed to be a morphic image of the other half, for some non-erasing morphism.

Theorem 1. *[4] There exists an infinite binary word that avoids all factors of the form $fm(f)$ and $m(f)f$, for all non-erasing binary morphisms m, with $|f| \geqslant 5$.*

Notice that classical squares are indeed forbidden in Theorem 1 when m is the identity. Notice also that m is required to be non-erasing to prevent that $m(f)$ is empty.

In this paper, we investigate whether we can remove the constraint that f and $m(f)$ are adjacent. That is, does there exist an infinite word w such that f and $m(f)$ are both factors of w only for finitely many f's? The answer is obviously negative because m can be the identity, that is, the morphism such that $m(i) = i$ for every letter i in the alphabet of w. This implies that for every $n \geqslant 1$, the prefix p_n of length n of w and the word $m(p_n) = p_n$ are both factors of w. To get rid of such a triviality, we require that m is not the identity. Thus, we say that a morphism is *admissible* if it is non-erasing and distinct from the identity. Then we say that a finite factor f of a word w is *imaged* if there exists an admissible morphism m such that w contains $m(f)$.

*This work is supported by the ANR project CADO (ANR-24-CE48-3758-01).

ISSN 1365–8050

Remark 2.

(i) If f is imaged in w, then every factor of f is imaged in w.

(ii) If f is a factor of w that does not contain every letter of w, then f is imaged in w.

(iii) In particular, the empty word ε is imaged in every infinite word.

To prove Remark 2.(ii), let f be a factor of w that does not contain the letter a of w. Let m be the admissible morphism such that $m(a) = aa$ and $m(i) = i$ for every letter i of w other than a. Then $m(f) = f$ even if m is not the identity. So f is imaged since w contains f and $m(f)$.

Our results consider the length and the number of imaged factors that are unavoidable in infinite binary words. First, we prove that long imaged factors can be avoided.

Theorem 3. *There exist exponentially many binary words that avoid imaged factors of length 7.*

Then we prove that the length 7 is sharp in the previous result.

Theorem 4. *Every infinite word over a finite alphabet contains an imaged factor of length 6.*

This naturally leads to wonder what is the minimum number of imaged factor in an infinite binary word. We also obtain a precise answer in this case.

Theorem 5. *There exist exponentially many binary words that contain at most 36 imaged factors.*

Theorem 6. *Every infinite binary word contains at least 36 imaged factors.*

Now let us give useful standard definitions. A *repetition* w is a finite prefix of u^ω such that u is a non-empty finite word and $|w| > |u|$. The *period* of w is $|u|$ and the *exponent* of w is $\frac{|w|}{|u|}$. A word is (β^+, n)-free if it contains no repetition u^e with $|u| \geqslant n$ and $e > \beta$. A word is β^+-free if it is $(\beta^+, 1)$-free. A morphism $f : \Sigma^* \to \Delta^*$ is *q-uniform* if $|f(a)| = q$ for every $a \in \Sigma$, and f is *synchronizing* if for all $a, b, c \in \Sigma$ and $u, v \in \Delta^*$, such that $f(ab) = uf(c)v$, then either $u = \varepsilon$ and $a = c$, or $v = \varepsilon$ and $b = c$.

The proofs of Theorems 3 and 5 use the following lemma.

Lemma 7. *[5] Let $\alpha, \beta \in \mathbb{Q}$, $1 < \alpha < \beta < 2$ and $n \in \mathbb{N}^*$. Let $h : \Sigma_s^* \to \Sigma_e^*$ be a synchronizing q-uniform morphism (with $q \geqslant 1$). If $h(w)$ is (β^+, n)-free for every α^+-free word w such that $|w| < \max\left(\frac{2\beta}{\beta-\alpha}, \frac{2(q-1)(2\beta-1)}{q(\beta-1)}\right)$, then $h(t)$ is (β^+, n)-free for every (finite or infinite) α^+-free word t.*

To obtain Theorems 3 and 5, we use Lemma 7 such that the pre-images are ternary $\frac{7}{4}^+$-free words. Since we know that there exist exponentially many ternary $\frac{7}{4}^+$-free words [5], this proves that there exist indeed exponentially many of the binary words considered in Theorems 3 and 5.

We provide in the associated git repository a program that verifies if Lemma 7 can be applied to a given morphism [6].

2 Avoiding imaged factors of length 7

Now we prove Theorem 3. Let w be the image of any ternary $\frac{7}{4}^+$-free word by the following 37-uniform morphism.

$$\begin{aligned}
\texttt{0} &\to \texttt{0001110101001100011101000110101001101}\\
\texttt{1} &\to \texttt{0001110101000110100110001110100110101}\\
\texttt{2} &\to \texttt{0001110100110001110101000110101001101}
\end{aligned}$$

This morphism was found using the method described in [5]. The following properties of w can be checked by standard techniques and Lemma 7.

(a) w contains no factor in $F = \{\texttt{0000}, \texttt{1111}, \texttt{0010}, \texttt{1011}, \texttt{010101}\}$.

(b) w is $\left(\frac{289}{148}^+, 3\right)$-free and contains no square other than $\texttt{00}$, $\texttt{11}$, $\texttt{0101}$, and $\texttt{1010}$.

(c) w does not contain both a factor f of length 7 and $\overline{f}$ (where $\overline{f}$ is the bit-complement of f, that is, the image of f by $\texttt{0} \to \texttt{1}$; $\texttt{1} \to \texttt{0}$).

(d) Every factor f of length 7 of w contains at least one of the following:

- the factor $\texttt{0101}$,
- the factor $\texttt{1010}$,
- both the factors $\texttt{00}$ and $\texttt{11}$.

Let us show that w does not contain two distinct factors f and $m(f)$ such that $|f| = 7$ and m is a non-erasing morphism. To write a morphism m, we use the notation $m(\texttt{0})/m(\texttt{1})$.

By (d), both letters $\texttt{0}$ and $\texttt{1}$ are contained in a square of f. The m-images of these squares are constrained by (b), so that $|m(\texttt{0})| \leqslant 2$ and $|m(\texttt{1})| \leqslant 2$. Here are further restrictions on m:

- The identity morphism $\texttt{0}/\texttt{1}$ is ruled out by definition.
- The bit-complement morphism $\texttt{1}/\texttt{0}$ is ruled out by (c).
- $m(\texttt{0}) = m(\texttt{1})$ is ruled out since otherwise $m(f) = (m(\texttt{0}))^7$ would contain $\texttt{0000}$, $\texttt{1111}$, or $\texttt{010101}$, contradicting (a)).

In particular, $|m(\texttt{0})| + |m(\texttt{1})| \geqslant 3$. So f contains neither $\texttt{0101}$ nor $\texttt{1010}$, since otherwise $m(f)$ would contain a square with period at least 3, contradicting (b). By (d), f contains both $\texttt{00}$ and $\texttt{11}$. This implies one more restriction on m:

- $\{m(\texttt{0}), m(\texttt{1})\} \cap \{\texttt{00}, \texttt{11}\} = \emptyset$ since otherwise $m(f)$ would contain $\texttt{0000}$ or $\texttt{1111}$.

So m is in $S_m = \{\texttt{0/01, 0/10, 1/01, 1/10, 01/0, 01/1, 10/0, 10/1, 01/10, 10/01}\}$ and f is in $S_f = \{\texttt{0001101}, \texttt{0001110}, \texttt{0011000}, \texttt{0011101}, \texttt{0100011}, \texttt{0100110}, \texttt{0110001}, \texttt{1000110}, \texttt{1000111}, \texttt{1001100}, \texttt{1001101}, \texttt{1100011}\}$. Finally, we check that for every $m \in S_m$ and every $f \in S_f$, the image $m(f)$ contains a factor in F, and thus is not a factor of w.

3 Imaged factors of length 6 are unavoidable

This section is devoted to Theorem 4. Suppose that Theorem 4 is false, that is, there exists an infinite word w over a finite alphabet that does not contain two factors f and $m(f)$ such that $|f| = 6$ and m is an admissible morphism. Notice that the conditions on f and $m(f)$ define a factorial language. By a standard argument, we can assume that w is recurrent, that is, all its finite factors appear infinitely often.

Lemma 8. *Every factor of length 6 of w contains every letter of w at least twice.*

Proof: By remark 2.(ii), every factor of length 6 of w contains every letter of w at least once. Now, we rule out that a factor f of length 6 of w contains a letter a exactly once. We write $f = pas$ where p and s are possibly empty words that do not contain a. Since w is recurrent, it contains a factor $fvf = pasvpas$ for some word v. Notice that $pasvpas = m(pas)$ where m is the admissible morphism such that $m(a) = asvpa$ and $m(i) = i$ for every other letter i. This is a contradiction since w contains both f and $m(f)$. □

Now we consider the size s of the alphabet of w. Lemma 8 implies $s \leqslant 3$. We can rule out $s = 1$ since $\mathtt{0}^{12}$ is the image of $\mathtt{0}^6$ by the admissible morphism $\mathtt{0} \mapsto \mathtt{00}$. If $s = 3$, then Lemma 8 implies that every factor of w of length 6 contains each of the three letters exactly twice. To ensure this property of the factors $w_k w_{k+1} w_{k+2} w_{k+3} w_{k+4} w_{k+5}$ and $w_{k+1} w_{k+2} w_{k+3} w_{k+4} w_{k+5} w_{k+6}$, we must have $w_k = w_{k+6}$. So w is periodic with period 6. Let f be any factor of w of length 6. Then w contains $f^6 = m(f)$, where m is the admissible morphism such that $m(\mathtt{0}) = m(\mathtt{1}) = m(\mathtt{2}) = f$. This contradiction rules out $s = 3$.

There remains the case $s = 2$, that is, w is a binary word. Let us show that w does not contain 0000. By Lemma 8, 0000 can only extend to 11000011. However $\mathtt{11000011} = m(\mathtt{100001})$ with $m(\mathtt{0}) = \mathtt{0}$ and $m(\mathtt{1}) = \mathtt{11}$. So w avoids 0000, and by symmetry, w avoids 1111. Now, we describe the computation that checks that no such infinite binary word w exists. It is a backtracking algorithm which backtracks in the following situations:

- The suffix of the current word is 0000 or 1111.
- The current word contains the complement $\overline{s}$ of its suffix s of length 6.
- For $f \in \{\mathtt{010101}, \mathtt{001100}, \mathtt{001001}, \mathtt{011011}, \mathtt{001010}, \mathtt{010100}, \mathtt{011101}, \mathtt{010001}, \mathtt{011100}, \mathtt{001110}, \mathtt{011000}, \mathtt{000110}, \mathtt{010111}, \mathtt{000101}, \mathtt{010110}, \mathtt{011010}\}$, the current word contains either f or $\overline{f}$, and also contains as a suffix an image $m(f)$ such that $|m(\mathtt{0})| + |m(\mathtt{1})| \geqslant 3$.

This backtrack finishes[(i)], which implies that every binary word contains an imaged factor of length 6. This concludes the proof of Theorem 4.

[(i)] The program we used is available in the associated git repository [6].

4 Binary words with only 36 imaged factors

Let us prove Theorem 5. Let w be the image of any ternary $\frac{7}{4}^+$-free word by the following 342-uniform morphism.

$0 \to p$1100011000011000011000110001100001100011000110000110000110001
100011000011000110001100001100011000110000110000110001100011000011
000110001100001100001100011000110000110001100011000011000011000110
00110000110000110001100011000011000110001100001100001100011000110

$1 \to p$0110001100011000011000011000110001100001100011000110000110000
110001100011000011000110001100001100011000110000110000110001100011
000011000110001100001100001100011000110000110001100011000011000110
00110000110000110001100011000011000110001100001100001100011000110
$2 \to p$0110001100011000011000011000110001100001100011000110000110000
110001100011000011000011000110001100001100011000110000110000110001
100011000011000110001100001100001100011000110000110001100011000011
00011000110000110000110001100011000011000110001100001100001100011

where $p =$ 00011000011000110001100001100011000110000110000110001100011000
110000110001100011000011000.

The following properties can be checked by standard techniques and Lemma 7.

(a) w avoids the factors in $F = \{010, 101, 111, 1001, 00000\}$.

(b) w is $\left(\frac{1321}{684}^+, 245\right)$-free.

First, we consider the set $T = \{0000110, 0110000, 0110001, 1000011, 1000110, 1100001, 00011000\}$ of factors of w and we show that these factors are not imaged in w. Suppose for contradiction that f is a word in T and that m is an admissible morphism such that $m(f)$ is a factor of w. Notice that 00 and 11 are factors of f, so that $|m(0)| \leqslant 244$ and $|m(1)| \leqslant 244$ by (b). Then a computer check shows that f is not imaged in w.

By (a), the factors in $T' = F \cup T = \{010, 101, 111, 1001, 00000, 0000110, 0110000, 0110001, 1000011, 1000110, 1100001, 00011000\}$ are not imaged in w.

Let $I = \{\varepsilon, 0, 00, 000, 0000, 00001, 000011, 0001, 00011, 000110, 0001100, 001, 0011, 00110, 001100, 0011000, 01, 011, 0110, 01100, 011000, 1, 10, 100, 1000, 10000, 100001, 10001, 100011, 11, 110, 1100, 11000, 110000, 110001, 1100011\}$ be the set of words that avoid T'. Then every imaged factor of w is in I. Thus w contains at most $|I| = 36$ imaged factors.

Finally, Theorem 6 can be verified with an exhaustive search. If there exists a right-infinite word z with at most 35 imaged factors, then z contains 001 or 110, since otherwise z would contain $(01)^\omega$, 0^ω, or 1^ω as a suffix. So without loss of generality, we assume that 001 is a prefix of z. Then, we perform depth-first exploration of the binary words with prefix `001` using this straightforward procedure: we backtrack if the current word w contains at least 36 imaged factors, and we try to extend w otherwise. Since this backtracking program ends, there are no infinite binary words with at most 35 imaged factors. To count the number of imaged factors faster, we pre-compute the set A of all the factors of w and the set S_{uu} (resp. C_{uuu}) of the factors u of w such that uu (resp. uuu) is also a factor of w. Now, if f is a factor of w containing the factor `00` and we look for an image $m(f)$ that is also a factor of w, then we know that $m(\texttt{0}) \in S_{uu}$. Since S_{uu} is much smaller than A, the search for m is faster. We provide in the associated git repository a program that implements this exhaustive search [6] and runs in a few minutes on an ordinary laptop.

5 Concluding remarks

In this paper, we have considered infinite words that aim at limiting the amount of imaged factors. This rules out pure morphic words (i.e., fixed points of an admissible morphism), since every factor of a pure morphic word is imaged by its defining morphism. However, these words can be morphic. Indeed, our words are constructed as morphic images of arbitrary ternary $\frac{7}{4}^+$-free words and Dejean [1] constructed a pure morphic ternary $\frac{7}{4}^+$-free word.

One may consider a more extreme version of the notion in this paper by allowing f to be any binary word, not necessarily a factor of the considered infinite word w itself. We notice that every binary word, and in particular every factor of w, is the complement of a binary word. So again, to avoid trivialities, we have to put another constraint on the morphisms. The *complentary morphism* is $\texttt{0} \to \texttt{1}$; $\texttt{1} \to \texttt{0}$. A morphism is *neat* if it is non-erasing, distinct from the identity, and distinct from the complementary morphism. Given an infinite binary word w, we say that a finite binary word b is *realized* in w if there exists a neat morphism m such that $m(b)$ is a factor of w. Thus, b can be realized in w even if b is not a factor of w. We also notice that if b is realized in w, then every factor of b is realized in w.

Consider any infinite binary word w that contains no squares other than `00`, `11`, and `0101` [2, 3]. In particular, w contains no 4-power. We can quickly check that there exist only finitely many words that are realized in w. Consider a square u with period at least 2, that is, $u = vv$ with $p = |v| \geqslant 2$. Suppose that u is realized in w, that is, there exists a neat morphism m such that $m(u)$ is a factor of w. First, we rule out the case where v does not contain both `0` and `1`, since u and $m(u)$ would contain a 4-power. Similarly, we rule out the case where m is of the form $\texttt{0} \to x; \texttt{1} \to x$, since $m(u) = x^{2p}$ would contain x^4. Since m is neat, this implies that $|m(\texttt{0})| + |m(\texttt{1})| \geqslant 3$ and thus $|m(v)| \geqslant 3$. Therefore, w contains $m(u) = m(v)m(v)$, which is a square with period at least 3. This contradiction shows that words realized in w do not contain a square with period at least 2. It is well known that the length of a binary word with no square with period at least 2 is at most 18. Therefore, there are finitely many words realized in w.

Similarly to the results in this paper, we can ask for infinite binary words that optimally bound the length and the number of realized words.

References

[1] F. Dejean. Sur un théorème de thue. *J. Combin. Theory Ser. A*, 13:90–99, 1972.

[2] A. Fraenkel and R. Simpson. How many squares must a binary sequence contain? *Electron. J. Combinatorics.*, 2, 1995.

[3] D. Gabric and J. Shallit. The simplest binary word with only three squares. *RAIRO - Theoret. Informatics Appl.*, 55, 2021.

[4] T. Ng, P. Ochem, N. Rampersad, and J. Shallit. New results on pseudosquare avoidance. *WORDS 2019*, pages 264–274, 2019.

[5] P. Ochem. A generator of morphisms for infinite words. *RAIRO - Theoret. Informatics Appl.*, 40:427–441, 2006.

[6] P. Ochem and M. Rosenfeld. Words with few imaged factors. `https://gite.lirmm.fr/mrosenfeld/words-with-few-imaged-factors`, 2025.