Finite repetition threshold for large alphabets

Golnaz Badkobeh[a], Maxime Crochemore[a,b] and Michael Rao [c]
[a]King's College London, UK
[b]Université Paris-Est, France
[c]CNRS, Lyon, France

January 25, 2014

Abstract

We investigate the finite repetition threshold for k-letter alphabets, $k \geq 4$, that is the smallest number r for which there exists an infinite r^+-free word containing a finite number of r-powers. We show that there exists an infinite Dejean word on a 4-letter alphabet (*i.e.* a word without factors of exponent more than $\frac{7}{5}$) containing only two $\frac{7}{5}$-powers. For a 5-letter alphabet, we show that there exists an infinite Dejean word containing only 60 $\frac{5}{4}$-powers, and we conjecture that this number can be lowered to 45. Finally we show that the finite repetition threshold for k letters is equal to the repetition threshold for k letters, for every $k \geq 6$.

1 Introduction

Following the study of infinite words avoiding repetitions in relation to Dejean's statement on the repetition threshold of alphabets [5] we show that it is possible to impose more constraints on words. We are interested in infinite words whose maximal exponent of its finite factors does not exceed Dejean's threshold and that contain a finite number of factors having the maximal exponent. This introduces the notion of finite repetition threshold (see [2, 3]). Imposing this constraint is not possible on the binary alphabet whose finite repetition threshold is $\frac{7}{3}$ while the repetition threshold is 2 (see [13, 10]), but can be satisfied for the ternary alphabet [3]. We show here that the result also holds for larger alphabets. This confirms the intuition given by the growth rates of words having the smallest exponent according to their alphabet size (see [7, 14]).

Associated with the finite repetition threshold is the smallest number of factors of highest exponent that an infinite word can accommodate (see [6, 1]). We show here that there exists an infinite word on a 4-letter alphabet containing only two $\frac{7}{5}$-powers and no factor of exponent more than $\frac{7}{5}$. The only known proofs of the $\frac{7}{5}$ repetition threshold for 4 letters are due to Pansiot [9] and Rao [11]; both of their words contain 24 $\frac{7}{5}$-powers. On 5 letters, the proof of the $\frac{5}{4}$ threshold by Moulin-Ollagnier [8] provides a word with 360 $\frac{5}{4}$-powers of periods 4, 12 and 44. We show that this number can be reduced to 60 and conjecture that it can be lowered to 45, the smallest possible number.

Both results also provide in fact new proofs of the repetition thresholds for the corresponding alphabet sizes, 4 and 5. The question on the smallest number of factors of highest exponent in a Dejean word remains open for larger alphabets.

2 Preliminaries

We denote by Σ_k the set $\{1, 2, \ldots, k\}$ for $k \geq 2$. Let w be a word. We denote by $w[i]$ the i-th letter w, and by $w[i : j]$ (where $i \leq j$) the word $w[i]w[i+1]\ldots w[j]$. Two words w and w' are *conjugated* if there are u and v such that $w = uv$ and $w' = vu$. A *repetition* in w is a pair of words (p, e) where p is non-empty, e is a prefix of pe, and pe is a factor of w. The *period* of the repetition is $|p|$, and its *exponent* is $\frac{|pe|}{|p|}$. By abuse of notation, we sometimes identify the repetition (p, e) with the factor pe of w. A repetition (p, e) in w over the alphabet Σ_k is a *short repetition* if $|e| < k - 1$, otherwise it is a *kernel repetition*.

A word is *x-free* (resp. *x^+-free*) if it has no repetition of exponent at least x (resp. greater than x). A word is called an *x-power* if it is a repetition of exponent x.

The *repetition threshold* for k letters (or for a k-letter alphabet), denoted $\mathrm{RT}(k)$, is the infimum of maximum exponents of repetitions over all infinite words on a k-letter alphabet. The following was conjectured by Dejean [5] and finally proved by several authors (see [4, 11]).

$$\mathrm{RT}(k) = \begin{cases} \frac{7}{4} & k = 3 \\ \frac{7}{5} & k = 4 \\ \frac{k}{k-1} & k \geq 5 \text{ or } k = 2 \end{cases}$$

We say that an infinite word on a k-letter alphabet is a *Dejean word* if it is $\mathrm{RT}(k)^+$-free, and a factor is a *limit repetition* if its exponent is exactly $\mathrm{RT}(k)$.

The *finite repetition threshold* for k letters is the smallest number $\mathrm{FRT}(k)$ for which there exists an infinite $\mathrm{FRT}(k)^+$-free word containing a finite number of $\mathrm{RT}(k)$-powers (that is, it has a finite number of limit repetitions).

It is known that any infinite $\frac{7}{3}$-free infinite binary word contains an arbitrary number of squares [13, 10]. However, there exists an infinite binary word whose maximal exponent does not exceed $\frac{7}{3}$ and all of its squares have period length at most 7. In [2], the associated minimal number of squares that an infinite binary word can accommodate is given as follows: there exists an infinite binary word containing only 12 squares whose maximal exponent is $\frac{7}{3}$. The proof is based on a HD0L-system exploiting two special non-uniform morphisms, the first one on 6-letter alphabet and the second from 6 letters to binary. Furthermore, a simple construction of all binary words with only 11 squares whose maximal exponent is $\frac{7}{3}$ showed that this set is finite and that its longest element has length 116, which shows the minimality of 12.

This idea was extended and further studied in [3] on ternary words. The result is as follows: there exists an infinite ternary Dejean word containing only two $\frac{7}{4}$-powers. The proof is based on a 160-uniform morphism which translates any infinite Dejean word on 4 letters to an infinite Dejean word on 3 letters containing only two $\frac{7}{4}$-powers.

Pansiot proved that the repetition threshold for a 4-letter alphabet is $\frac{7}{5}$. In order to prove the result, Pansiot used a construction that codes a $\frac{k-1}{k-2}$-free word over alphabet Σ_k into a binary word. Let $k \geq 3$ and w be a $\frac{k-1}{k-2}$-free word over Σ_k, of length at least $k-1$. Then every factor of length $k-1$ consists of $k-1$ different letters. The *Pansiot code* of w is the binary word $\mathtt{P}_k(w)$ such that for all $i \in \{1, \dots, |w| - k + 1\}$ (for all $i \geq 1$ if w is infinite):

$$\mathtt{P}_k(w)[i] = \begin{cases} 0 & w[i+k-1] = w[i] \\ 1 & w[i+k-1] \notin \{w[i], \dots, w[i+k-2]\}. \end{cases}$$

Note that w is uniquely defined by $P_k(w)$ and $w[1 : k-1]$. One can define an inverse operation: for a binary word w, $\mathtt{M}_k(w)$ is the word on the alphabet Σ_k such that:

$$\mathtt{M}_k(w)[i] = \begin{cases} i & i < k \\ \mathtt{M}_k(w)[i-k+1] & i \geq k \text{ and } w[i-k+1] = 0 \\ \alpha & \text{otherwise} \end{cases}$$

where $\{\alpha\} = \Sigma_k \setminus \{\mathtt{M}_k(w)[i-k+1], \dots, \mathtt{M}_k(w)[i-1]\}$. Note that if $w[i] = i$ for every $i < k$, then $\mathtt{M}_k(P_k(w)) = w$.

We shall denote by $\mathbb{S}_k$ the *symmetric group* on k elements. Therefore the elements of this set are the permutations of the set $\Sigma_k = \{1, 2, \ldots, k\}$. We denote by Id_k the identity permutation of $\mathbb{S}_k$. We use cycle notation for permutations, that is $\sigma = (a_1 a_2 \ldots a_l)$ denotes the permutation such that $\sigma(a_i) = a_{i+1}$ (where the indices are taken modulo l). Let $\Psi : \Sigma^* \to \mathbb{S}_k$ be a morphism. A repetition (p, e) is a Ψ*-kernel repetition* if $p \in \ker(\Psi)$. Ψ-kernel repetitions were introduced by Moulin-Ollagnier [8], and generalized later by Rao [11].

Let $\varphi : \{0,1\}^* \to \mathbb{S}_k$ be the morphism such that $\varphi(0) = (1 \ldots k-1)$ and $\varphi(1) = (1 \ldots k)$. The following lemma by Moulin-Ollagnier gives a strong relation between kernel repetitions in a word on a k-letter alphabet and φ-kernel repetitions in its Pansiot code.

Lemma 1 ([8])**.** *Let w be a $\frac{k-1}{k-2}$-free word w on a k-letter alphabet. Then w has a kernel-repetition (p, e) if and only if $P_k(w)$ has a φ-kernel-repetition (p', e') with $|p'| = |p|$, $p'e' = P_k(pe)$ and $|e'| = |e| - k + 1$.*

Throughout this paper, in order to prove the existence of an infinite word complying with some properties, the following method is used. The main technique is to design two or more morphisms generating an appropriate infinite binary word and then translate that by the inverse of the Pansiot coding. One of the experimental techniques that we used consists of the following steps. We generate a long enough word satisfying the pre-defined constraints using a backtracking strategy, and we translate this word to a binary word by applying the Pansiot coding. Then, we search for its most repetitive motifs, and using selective elements of the set of motifs, we try to decode the word to find its pre-image according to the morphism defined by the motifs. If necessary, we iterate the previous step with the new word (pre-image of the first word). Backtracking is a general algorithm for finding all (or some) solutions to some computational problem; it incrementally builds candidates to the solutions, and abandons each partial candidate as soon as it determines it cannot possibly be completed to a valid solution.

3 Finite repetition threshold for 4-letter alphabets

Since the repetition threshold for a 4-letter alphabet is $\frac{7}{5}$, it suffices to show that there exists a $\frac{7}{5}^+$-free infinite word on Σ_4 with finitely many limit repetitions (that is $\frac{7}{5}$-powers). There are two proofs of Dejean's conjecture for 4-letter alphabets, by Pansiot [9] and Rao [11]. In both cases the number of limit repetitions contained in the infinite words is 24. This proves that the finite repetition threshold for 4 letters is $\frac{7}{5}$. In this section, we prove the following:

Theorem 1. *The finite repetition threshold for 4-letter alphabets is $\frac{7}{5}$ and the minimal number of $\frac{7}{5}$-powers is 2.*

A computer check shows that a word on a 4-letter alphabet for which the maximal exponent of factors is $\frac{7}{5}$ and that contains at most one limit repetition has maximal length 230. Then, to prove Theorem 1, we give a morphic word which is the Pansiot code of a Dejean word on 4 letters with only two limit repetitions. The correctness proof follows the plan and notations introduced in [11]. However, since the morphism φ' will be simpler here, we can make the proof self-contained. Informally, the idea is to prove that if the morphic word has a forbidden repetition with a long enough period, then it has a smaller forbidden repetition. Thus it remains to prove that the morphic word has no forbidden repetition with a period bounded by a constant, which can be done by a finite case analysis. Let:

$$f : \begin{cases} \mathtt{a} & \to \mathtt{abc} \\ \mathtt{b} & \to \mathtt{cda} \\ \mathtt{c} & \to \mathtt{adc} \\ \mathtt{d} & \to \mathtt{cba} \end{cases}$$

$$g : \begin{cases} \texttt{a} & \to \texttt{aacbbaaccbaabcabc} \\ \texttt{b} & \to \texttt{aacbacbaabbcaabbc} \\ \texttt{c} & \to \texttt{cbaaccbbaccabcabc} \\ \texttt{d} & \to \texttt{aacbaccaabbcaabbc} \end{cases}$$

$$h : \begin{cases} \texttt{a} \to & \texttt{10110101011011010110110101011010101101101010110110101010110101} \\ & \texttt{011011010101101101010110101011011010101} \\ \texttt{b} \to & \texttt{10110101011011010110110101011011010101101010110110101010110110} \\ & \texttt{101011010101101101010110110101011010101} \\ \texttt{c} \to & \texttt{10110101011011010110110101011011010101101101010110101010110110} \\ & \texttt{101011011010101101010110110101011011010.} \end{cases}$$

The rest of this section is devoted to the proof of the following theorem.

Theorem 2. *$w_0 = \mathtt{M}_4(h(g(f^\infty(\mathtt{a}))))$ is $\frac{7}{5}^+$-free and it contains only two $\frac{7}{5}$-powers:* $(3421432412, 3421)$ *and* $(1423412432, 1423)$.

Remark. *A computer check shows that the Pansiot code of every long enough $\frac{7}{5}^+$-free word on 4-letter alphabet with at most two limit repetitions contains $h(x)$ as factor, for an $x \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$. Moreover, every Pansiot code of a Dejean word with at most two limit repetitions starting with $h(x)$ (for $x \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$) must be followed by $h(y)$, for a $y \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$. Thus the morphism h in our construction is unavoidable,* i.e. *for every Dejean word w which proves Theorem 1, $P_4(w)$ must be the image by h of a ternary word (modulo the shift operation).*

The following properties derive from simple observations:

- f is 3-uniform, g is 17-uniform and h is 99-uniform. Thus $h \circ g$ is 1683-uniform. (A morphism $f : \Sigma^* \to \Sigma'^*$ is *l-uniform*, $l \in \mathbb{N}$, if for every $x \in \Sigma$, $|f(x)| = l$.)
- f, g, h and $h \circ g$ are comma-free. (A morphism $f : \Sigma^* \to \Sigma'^*$ is *comma-free* if whenever $f(xy) = uf(z)v$, then either $u = \epsilon$ or $v = \epsilon$, for every $x, y, z \in \Sigma$ and $u, v \in \Sigma'^*$.)
- The longest common prefix in $\{h \circ g(\mathtt{a}), h \circ g(\mathtt{b}), h \circ g(\mathtt{c}), h \circ g(\mathtt{d})\}$ has size 635 and the longest common suffix has size 990.
- For every $x \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, $\varphi(h(x)) = (13)$.

The last fact can be verified by a computer check (or by a tedious hand check). The notion of Ψ-kernel repetition is central in [8, 11]. However, the proof can be simplified here since $\varphi(h(x)) = (13)$ for every $x \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$ (which is not true for cases in [11]). Since g and h are uniform and of odd-size, for every $x \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}$, $\varphi(h(g(x))) = (13)$ and $\varphi(h(g(f(x)))) = (13)$. Let $\varphi' : \{0, 1, 2, 3\}^* \to \mathbb{S}_4$ such that $\varphi'(u) = (13)^{|u|}$. Thus (p, q) is a φ'-kernel repetition if (p, q) is a repetition, and $|p|$ is even. The following lemma gives a relation between φ-kernel repetitions in $w_1 = h(g(f^\infty(\mathtt{a})))$ and φ'-kernel repetitions in $w_2 = f^\infty(\mathtt{a})$.

Lemma 2. *Let (p_1, e_1) be a φ-kernel-repetition of w_1. If $|e_1| \geq 3365$, then w_2 has a φ'-kernel-repetition (p_2, e_2) with $|e_2| \geq \left\lceil \frac{|e_1| - 1625}{1683} \right\rceil$ and $|p_1| = 1683 \cdot |p_2|$.*

Proof. Suppose w.l.o.g. that (p_1, e_1) is a maximal repetition, *i.e.* there is no repetition (p'_1, e'_1) in w_1 such that $|p'_1| = |p_1|$ and $p_1 e_1$ is a proper factor of $p'_1 e'_1$. If $|e_1| \geq 3365 = 2 \cdot 1683 - 1$, then $h \circ g(\mathtt{a})$, $h \circ g(\mathtt{b})$ or $h \circ g(\mathtt{c})$ appears as a factor in e_1. Since $h \circ g$ is comma-free and 1683-uniform, $|p_1|$ is a multiple of 1683. Let $n \in \mathbb{N}$ such that $|p_1| = n \cdot 1683$. Then there is a factor $u = a_1 \ldots a_l$ in w_2 such that $h \circ g(u) = v p_1 e_1 v'$, v is a proper prefix of $h \circ g(a_1)$ and v' is a proper suffix of $h \circ g(a_l)$. Since (p_1, e_1) is a repetition of period $n \cdot 1683$, for every $n + 1 < i < l$,

$a_i = a_{i-n}$. Thus $(p_2, e_2) = (a_2 \ldots a_{n+1}, a_{n+2} \ldots a_{l-1})$ is a repetition in w_2 of period n. Moreover, $\varphi'(p_2) = \varphi(h \circ g(p_2)) = \mathrm{Id}_4$ since $\varphi(p_1) = \mathrm{Id}_k$, and p_1 is conjugate to $h \circ g(p_2)$. Since $p_1 e_1$ is maximal on the left, $|v| \geq 693$, and since $p_1 e_1$ is maximal on the right, $|v'| \geq 1048$. Thus $|e_1| - 1625 \leq 1683 \cdot |e_2|$, and w_2 has a φ'-kernel repetition (p_2, e_2) with $|e_2| \geq \left\lceil \frac{|e_1|-1625}{1683} \right\rceil$ and $|p_1| = 1683 \cdot |p_2|$. □

The proof of the following Lemma is similar, and is omitted.

Lemma 3. *If (p_2, e_2) is a φ'-kernel repetition of $w_2 = f^\infty(\mathtt{a})$ with $|e_2| \geq 5$, then w_2 has a φ'-kernel-repetition (p_2', e_2') with $|e_2'| \geq \left\lceil \frac{|e_2|-2}{3} \right\rceil$ and $|p_2| = 3 \cdot |p_2'|$.*

Lemma 4. *Suppose that w_2 has a φ'-kernel-repetition (p_2, e_2) with $|e_2| \geq 5$ and $\frac{|e_2|+1}{|p_2|} \geq \frac{2}{5}$. Then there exists a φ'-kernel-repetition (p_2', e_2') with $|p_2| = 3 \cdot |p_2'|$ and $\frac{|e_2'|+1}{|p_2'|} \geq \frac{2}{5}$.*

Proof. By Lemma 3,

$$\frac{2}{5} \leq \frac{|e_2| + 1}{|p_2|} \leq \frac{3 \cdot |e_2'| + 3}{3 \cdot |p_2'|} = \frac{|e_2'| + 1}{|p_2'|}.$$

□

The following fact can be verified by a computer check:

Fact 1. *There is no φ'-kernel-repetition (p_2, e_2) with $2 \leq |e_2| < 5$ and $\frac{|e_2|+1}{|p_2|} \geq \frac{2}{5}$ in w_2.*

Thus by Lemma 4:

Corollary 1. *There is no φ'-kernel-repetition (p_2, e_2) with $2 \leq |e_2|$ and $\frac{|e_2|+1}{|p_2|} \geq \frac{2}{5}$ in w_2.*

Lemma 5. *w_1 has no φ-kernel-repetition (p_1, e_1) with $|e_1| \geq 3 \cdot 1683$ and $\frac{|e_1|+3}{|p_1|} \geq \frac{2}{5}$.*

Proof. Suppose that w_1 has a φ-kernel-repetition (p_1, e_1) with $|e_1| \geq 3 \cdot 1683$ and $\frac{|e_1|+3}{|p_1|} \geq \frac{2}{5}$. By Lemma 2, w_2 has a φ'-kernel repetition (p_2, e_2) with $|e_2| \geq 2$ and

$$\frac{2}{5} \leq \frac{|e_1| + 3}{|p_1|} \leq \frac{1683 \cdot |e_2| + 1625 + 3}{1683 \cdot |p_2|} < \frac{|e_2| + 1}{|p_2|}.$$

By Corollary 1, w_2 has no such φ'-kernel repetition. Contradiction. □

By Lemma 5, if w_1 has a φ-kernel repetition (p_1, e_1) with $\frac{|p_1 e_1|+3}{|p_1|} \geq \frac{7}{5}$, then $|p_1| \leq \frac{5}{2}(|e_1| + 3) < \frac{5 \cdot (3 \cdot 1683 + 3)}{2}$, that is $|p_1| < 12630$. By Lemma 1, and since w_1 is the Pansiot code of w_0, w_0 has no repetition (p, e) with $|p| \geq 12633$ and $\frac{|pe|}{|p|} \geq \frac{7}{5}$. To complete the proof of Theorem 2, it suffices to show that for every repetition (p, e) in w_0 with $|p| < 12633$, either $\frac{|pe|}{|p|} < \frac{7}{5}$, or $\frac{|pe|}{|p|} = \frac{7}{5}$ and $(p, e) \in \{(3421432412, 3421), (1423412432, 1423)\}$. This fact has been verified by a computer check.

4 Finite repetition threshold for 5-letter alphabets

This section is devoted to the study of the minimal number of limit repetitions over all Dejean words on a 5-letter alphabet. Moulin-Ollagnier gave a proof of Dejean's conjecture for $k = 5$ (see [8]). Let:

$$m : \begin{cases} \mathtt{0} & \to \mathtt{010101101101010110110} \\ \mathtt{1} & \to \mathtt{101010101101101101101}. \end{cases}$$

Then $\mathtt{M}_5(m^\infty(\mathtt{0}))$ is $\frac{5}{4}^+$-free. We claim without proof that it contains 360 limit repetitions, of which a third have period 4, a third period 12 and the remaining have period 44. This proves that the finite repetition threshold for 5-letter alphabets is $\frac{5}{4}$. We show, with an explicit construction, that the number of limit repetitions can be lowered to 60, and we conjecture that the minimal number is 45. Most of the intermediate proofs are similar to those in Section 3, and are omitted. Let:

$$f : \begin{cases} \mathtt{a} & \to \mathtt{aaabbababbaaabbaabb} \\ \mathtt{b} & \to \mathtt{aabbbaababaabbbaabb} \end{cases}$$

$$g : \begin{cases} \mathtt{a} & \to \mathtt{aaaababbbbababaaaababbb} \\ \mathtt{b} & \to \mathtt{bbbbabaaaabababbbbabaaa} \end{cases}$$

$$h : \begin{cases} \mathtt{a} \to & \mathtt{110110101010110110101010110110101011011010101101101101010110} \\ & \mathtt{11011011010101011011010101101101010110110110101010110} \\ \mathtt{b} \to & \mathtt{110110101011011010101101101010101101101101101010110110110101} \\ & \mathtt{01101101010110110101010110110101010110110110101010110.} \end{cases}$$

Let $w_2 = f^\infty(\mathtt{a})$, $w_1 = h(g(w_2))$ and $w_0 = \mathtt{M}_5(w_1)$.

Theorem 3. *w_0 is a Dejean word on 5 letters, and it contains only 60 limit repetitions, all of which have period 4.*

The following properties will help with the proof of Theorem 3:

- f is 19-uniform, g is 29-uniform and h is 113-uniform. Thus $h \circ g$ is 3277-uniform.
- f, g, h and $h \circ g$ are comma-free.
- The longest common prefix in $\{h \circ g(\mathtt{a}), h \circ g(\mathtt{b})\}$ has size 11 and the longest common suffix has size 24.
- For every $x \in \{\mathtt{a}, \mathtt{b}\}$, $\varphi(h(x)) = (12)(354)$, thus for every $x \in \{\mathtt{a}, \mathtt{b}\}$, $\varphi(h(g(x)) = (12)(345)$ and $\varphi(h(g(f(x))) = (12)(345)$.

Let $\varphi' : \{0, 1, 2, 3, 4\}^* \to \mathbb{S}_5$ such that $\varphi'(u) = [(12)(345)]^{|u|}$. Thus (p, q) is a φ'-kernel repetition if and only if (p, q) is a repetition, and $|p|$ is divisible by 6.

Lemma 6. *Let (p_1, e_1) be a φ-kernel-repetition of $w_1 = h(g(f^\infty(\mathtt{a})))$. If $|e_1| \geq 6553$, then $w_2 = f^\infty(\mathtt{a})$ has a φ'-kernel-repetition (p_2, e_2) with $|e_2| \geq \left\lceil \frac{|e_1|-35}{3277} \right\rceil$ and $|p_1| = 3277 \cdot |p_2|$.*

Lemma 7. *If $|e_2| \geq 37$, then $w_2 = f^\infty(\mathtt{a})$ has a φ'-kernel-repetition (p'_2, e'_2) with $|e'_2| \geq \left\lceil \frac{|e_2|-8}{19} \right\rceil$ and $|p_2| = 19 \cdot |p'_2|$.*

Here, we adapt the same approach as in Section 3 (Lemma 4 and Fact 1) with the appropriate changes based on the size of the morphism f and the exponent $\frac{5}{4}$. The next corollary follows:

Corollary 2. *There is no φ'-kernel-repetition (p_2, e_2) with $6 \leq |e_2|$ and $\frac{|e_2|+1}{|p_2|} \geq \frac{1}{4}$ in w_2.*

Lemma 8. *w_1 has no φ-kernel-repetition (p_1, e_1) with $|e_1| \geq 6 \cdot 3277$ and $\frac{|e_1|+4}{|p_1|} \geq \frac{1}{4}$.*

The proof of Lemma 8 is similar to the proof of Lemma 5, and is a direct consequence of Lemma 6 and 7. By Lemma 8, if w_1 has a φ-kernel repetition (p_1, e_1) with $\frac{|p_1 e_1|+4}{|p_1|} \geq \frac{5}{4}$, then $|p_1| \leq \frac{4}{1}(|e_1| + 4) < 4 \cdot (6 \cdot 3277 + 4)$, that is $|p_1| < 78664$. By Lemma 1, and since w_1 is the Pansiot code of w_0, w_0 has no repetition (p, e) with $|p| \geq 78664$ and $\frac{|pe|}{|p|} \geq \frac{5}{4}$. A computer check showed that among every repetition (p, e) in w_0 of period at most 78664, none has an exponent greater than $\frac{5}{4}$. This proves that $\mathrm{FRT}(5) = \frac{5}{4}$. This check also reveals that there are only 60 limit repetitions (p, e) in w_0, and for every limit repetition, $|e| = 1$. This concludes the proof of Theorem 3.

To conclude this section, we give lower bounds on the number of limit repetitions for a Dejean word on 5 letters. The following facts have been verified by a computer check. A standard (and easily parallelizable) backtrack algorithm written in C++ took approximately 3 days (resp. 120 days) of single-core time on a 2.1GHz CPU to verify fact (a) (resp. fact (b)).

Fact 2.

(a) A $\frac{5}{4}^+$-free word on a 5-letter alphabet that contains at most 44 limit repetitions has size at most 4648.

(b) A $\frac{5}{4}^+$-free word on a 5-letter alphabet that contains at most 45 limit repetitions, and such that every limit repetition has period 4, has size at most 7331.

Thus the minimal number of limit repetitions over all Dejean words on 5 letters is between 45 and 60. Based on computer experiments, we conjecture the following.

Conjecture 1.

- *There exists an infinite Dejean word on a 5-letter alphabet with only 45 limit repetitions.*
- *There exists an infinite Dejean word on a 5-letter alphabet with only 46 limit repetitions, and such that every limit repetition has period 4.*

5 Finite repetition threshold for k-letter alphabets, $k \geq 6$

Looking at the existing proofs for Dejean's conjecture shows in fact $\mathrm{FRT}(k) = \mathrm{RT}(k)$ for $k \geq 6$, that is, known constructions of Dejean words have finitely many limit repetitions.

Lemma 9. *For every $5 \leq k \leq 11$, $\mathrm{FRT}(k) = \mathrm{RT}(k)$.*

Proof. Moulin-Ollagnier gave uniform morphisms h_k, for $5 \leq k \leq 11$, such that $\mathtt{M}_k(h_k^\infty(1))$ is a Dejean word on a k-letter alphabet [8]. We show that these Dejean words have finitely many limit repetitions. We fix a $5 \leq k \leq 11$, and let $h = h_k$. Let $u = |h(0)| = |h(1)|$, and let L be the longest common prefix of $h(0)$ and $h(1)$. Note that the last letters of $h(0)$ and $h(1)$ differ. Suppose that $\mathtt{M}_k(h^\infty(1))$ has infinitely many limit repetitions. Let $\mathcal{L}$ be the set of φ-kernel repetitions (p, e) in $h^\infty(1)$ with $\frac{|e|+k-1}{|p|} = \frac{1}{k-1}$, that is φ-kernel repetitions which correspond to a limit repetition. Since $\mathtt{M}_k(h^\infty(1))$ has infinitely many limit repetitions, $\mathcal{L}$ is also infinite. By [8, Corollary 3.20], there is a repetition $(p, e) \in \mathcal{L}$ and a $n > 0$ such that $(h^n(p), \mu^n(e)) \in \mathcal{L}$, where $\mu(w) = h(w)L$. Then:

$$\frac{|e| + k - 1}{|p|} = \frac{u^n \cdot |e| + |L| \cdot \sum_{i=0}^{n-1} u^i + k - 1}{u^n \cdot |p|}$$

which is satisfied when:

$$(u - 1) \cdot (k - 1) = |L|.$$

We have a contradiction, since $|L| \leq u - 1$ and $k \geq 5$. □

For $k \geq 12$ we use the following lemma.

Lemma 10. *Let $k \geq 5$. Let w_{TM} be the Prouhet-Thue-Morse word, that is $w_{TM} = \eta^\infty(0)$ where $\eta : 0 \to 01, 1 \to 10$. Let $w_1 = h(w_{TM})$ be a binary word such that:*

1. $w_0 = \mathtt{M}_k(w_1)$ *is a Dejean word on a k-letter alphabet.*

2. $h : \{0,1\}^* \to \{0,1\}^*$ *is n-uniform,*

3. *there exists* $\sigma \in \mathbb{S}_k$ *such that* $\varphi'(0)\varphi'(1) = \sigma\varphi'(0)\sigma^{-1}$ *and* $\varphi'(1)\varphi'(0) = \sigma\varphi'(1)\sigma^{-1}$*, where* $\varphi' : \{0,1\}^* \to \mathbb{S}_k$ *is the morphism such that* $\varphi'(0) = \varphi(h(0))$ *and* $\varphi'(1) = \varphi(h(1))$.

Then w_0 *has finitely many limit repetitions.*

Proof. Note that w_1 cannot contain arbitrarily large powers, otherwise w_0 would also contain arbitrarily large powers. We have $h(0) \neq h(1)$, since the Pansiot code of a Dejean word is not periodic. Thus we can suppose w.l.o.g. that the last letters of $h(0)$ and $h(1)$ differ, otherwise we replace $h(0)$ (resp. $h(1)$) by $uh(0)u^{-1}$ (resp. $uh(1)u^{-1}$), where u is the largest common suffix of $h(0)$ and $h(1)$. Moreover we can suppose w.l.o.g. that the last letter of $h(x)$ is x for $x \in \{0,1\}$, otherwise we exchange $h(0)$ and $h(1)$ (note the factor set of w_{TM} is closed under the complementation). Let L be the largest common prefix of $h(0)$ and $h(1)$, and let ℓ be the size of L.

Claim 1. *There is a* $B \in \mathbb{N}$ *such that for every* φ*-kernel repetition* (p,e) *in* w_1 *with* $|e| \geq B$*,* $|p|$ *is a multiple of* n*.*

Proof. Let $\overline{v}$, where v is a binary word, be the image of v by the morphism $0 \to 1, 1 \to 0$. Let $M = \{1 \leq i \leq n : h(0)[i] = h(1)[i]\}$, $N = \{1 \leq i \leq n : h(0)[i] = 0 \text{ and } h(1)[i] = 1\}$ and $N' = \{1 \leq i \leq n : h(0)[i] = 1 \text{ and } h(1)[i] = 0\}$. Note that $\{M, N, N'\}$ is a partition of $\{1, \ldots, n\}$, and since the last letter of $h(x)$ is x, we have $n \in N$. Suppose that the claim is false. Then there are arbitrarily large factors u and u' of w_{TM}, with $|u| = |u'|+1 \geq 4$, such that $vh(u')$ is a prefix of $h(u)$, where v is a non-empty proper suffix of $h(0)$ or $h(1)$. Since w_{TM} is cube-free, u', $u[1 : |u|-1]$ and $u[2 : |u|]$ contain 0 and 1 as factors. Thus for every $i \in M$, $i + |v| \in M \pmod{n}$, that is $M + |v| = M \pmod{n}$. Since $n \in N$, we have $|v| \notin M$. Since the last letter of $h(x)$ is x (for $x \in \{0,1\}$), u' is either a suffix of u or of $\overline{u}$, depending on whether $|v| \in N$ or $|v| \in N'$. If u' is a suffix of u, then $h(u')[i] = h(u')[i+n-|v|]$ for every $1 \leq i \leq n \cdot |u'| - n + |v|$, that is $h(u')$ is a repetition of period $n - |v|$. Suppose now that u' is a suffix of $\overline{u}$. Let $1 \leq i \leq n \cdot |u'| - 2(n-|v|)$. If $i \in M \pmod{n}$, $h(u')[i] = \overline{h(u')[i+n-|v|]} = h(u')[i+2(n-|v|)]$, since $\{i, i+n-|v|]\} \subseteq M \pmod{n}$. Otherwise $h(u')[i] = \overline{h(u')[i+n-|v|]} = h(u')[i+2(n-|v|)]$, since $\{i, i+n-|v|]\} \subseteq \{1 \ldots n\} \setminus M \pmod{n}$. Thus $h(u')$ is a repetition of period $2(n-|v|)$. In all cases, $h(u')$ is a repetition of period at most $2n$. Hence w_1 contains arbitrarily large powers, and we have a contradiction. ☐

Suppose that w_0 has infinitely many limit repetitions. Then w_1 has infinitely many φ-kernel repetitions (p,e) with $\frac{|e|+k-1}{|p|} = \frac{1}{k-1}$. By Claim 1, if e is long enough then $|p|$ is a multiple of n, and w_{TM} has a repetition (p', e') such that $n \cdot |p'| = |p|$, p is conjugated to $h(p')$ and $|e| = n \cdot |e'| + \ell$. Since (p,e) is a φ-kernel repetition, $\varphi(p) = \mathrm{Id}_k$ and $\varphi(h(p')) = \mathrm{Id}_k$. By condition (3), $\varphi'(p') = \mathrm{Id}_k$, and (p', e') is a φ'-kernel repetition of w_{TM}.

Thus w_{TM} has infinitely many φ'-kernel repetitions (p', e') with $\frac{n \cdot |e'|+\ell+k-1}{n \cdot |p'|} = \frac{1}{k-1}$. Let (p', e') be a φ'-kernel repetition in w_{TM} with $|e'| \geq 4$ and $\frac{n \cdot |e'|+\ell+k-1}{n \cdot |p'|} = \frac{1}{k-1}$. By [11, Corollary 9], w_{TM} has a φ'-kernel repetition (p'', e'') with $|p'| = 2 \cdot |p''|$ and $|e'| \leq 2 \cdot |e''|$. Thus w_1 has a φ-kernel repetition $(h(p''), h(e'')L)$, and w_0 has a kernel repetition of exponent $\frac{n \cdot |e''|+\ell+k-1}{n \cdot |p''|} > \frac{n \cdot |e'|+\ell+k-1}{n \cdot |p'|} = \frac{1}{k-1}$. We have a contradiction with the fact that w_0 is a Dejean word. ☐

We apply the previous lemma on constructions for $8 \leq k \leq 38$ in [11], or $k \geq 24$ in [12] to show that $\mathrm{FRT}(k) = \mathrm{RT}(k)$ for every $k \geq 8$.

We conclude with the following open questions.

Conjecture 2. *For every $k \geq 5$, there is a infinite Dejean word on k letters such that the only allowed limit repetitions have period $k-1$.*

Let $\mathrm{LR}(k)$, for $k \geq 3$, be the minimal number of limit repetitions over all Dejean words on k letters. Similarly let $\mathrm{LR}'(k)$, for $k \geq 5$, be the minimal number of limit repetitions over all Dejean words on k letters such that every limit repetition has period $k-1$. By the results of the present article, $\mathrm{LR}(k)$ is defined for every $k \geq 3$, and $\mathrm{LR}'(k)$ is defined if Conjecture 2 is true. We know that $\mathrm{LR}(3) = 2$ [3], $\mathrm{LR}(4) = 2$, $45 \leq \mathrm{LR}(5) \leq 60$ and $46 \leq \mathrm{LR}'(5) \leq 60$. It may be difficult to find the exact value of $\mathrm{LR}(k)$ or $\mathrm{LR}'(k)$ for any $k \geq 5$, but we can ask the following question.

Question 1. *Find a lower or an upper bound for* $\mathrm{LR}(k)$ *or* $\mathrm{LR}'(k)$*, $k \geq 5$.*

Conjecture 2 implies that $\mathrm{LR}(k) \leq \mathrm{LR}'(k) \leq k!$. On the other hand, limit repetitions cannot be avoided when $k \geq 5$ since every 0 in the Pansiot code leads to an occurrence of a limit repetition of period $k-1$. Thus $0 < \mathrm{LR}(k) \leq \mathrm{LR}'(k)$.

References

[1] Golnaz Badkobeh. Fewest repetitions vs maximal-exponent powers in infinite binary words. *Theoret. Comput. Sci.*, 412(48):6625–6633, 2011.

[2] Golnaz Badkobeh and Maxime Crochemore. Fewest repetitions in infinite binary words. *RAIRO - Theor. Inf. and Applic.*

[3] Golnaz Badkobeh and Maxime Crochemore. Finite-repetition threshold for infinite ternary words. In *WORDS*, pages 37–43, 2011.

[4] James D. Currie and Narad Rampersad. A proof of Dejean's conjecture. *Math. Comput.*, 80(274):1063–1070, 2011.

[5] Françoise Dejean. Sur un théorème de Thue. *J. Comb. Theory, Ser. A*, 13(1):90–99, 1972.

[6] Aviezri S. Fraenkel and Jamie Simpson. How many squares must a binary sequence contain? *Electr. J. Comb.*, 2, 1995.

[7] Juhani Karhumäki and Jeffrey Shallit. Polynomial versus exponential growth in repetition-free binary words. *J. Comb. Theory, Ser. A*, 105(2):335–347, 2004.

[8] Jean Moulin-Ollagnier. Proof of Dejean's conjecture for alphabets with 5, 6, 7, 8, 9, 10 and 11 letters. *Theoret. Comput. Sci.*, 95(2), 1992.

[9] Jean-Jacques Pansiot. A propos d'une conjecture de F. Dejean sur les répétitions dans les mots. In Josep Díaz, editor, *Automata, Languages and Programming, 10th Colloquium, Barcelona, Spain, July 18-22, 1983, Proceedings*, volume 154 of *Lecture Notes in Computer Science*, pages 585–596. Springer, 1983.

[10] Narad Rampersad, Jeffrey Shallit, and Ming Wei Wang. Avoiding large squares in infinite binary words. *Theor. Comput. Sci.*, 339(1):19–34, 2005.

[11] Michaël Rao. Last cases of Dejean's conjecture. *Theoret. Comput. Sci.*, 412(27):3010 – 3018, 2011.

[12] Michaël Rao and Elise Vaslet. Dejean words with frequency constraint. 2013. Manuscript.

[13] Jeffrey Shallit. Simultaneous avoidance of large squares and fractional powers in infinite binary words. *Intl. J. Found. Comput. Sci*, 15(2):317–327, 2004.

[14] Arseny M. Shur and Irina A. Gorbunova. On the growth rates of complexity of threshold languages. *RAIRO - Theor. Inf. and Applic.*, 44(1):175–192, 2010.