REPETITION THRESHOLDS ALONG ARITHMETIC PROGRESSIONS: OVERLAP-FREE BINARY WORDS AND THEIR DECIMATIONS

KOREA SUPERINTELLIGENCE LABS

Abstract. For an infinite word $w = w_0w_1w_2\cdots$ over a k-letter alphabet and an integer $p \geq 1$, let $\langle w\rangle_p = w_0w_pw_{2p}\cdots$ be the decimation of w modulo p, and let $\mathrm{RT}(k,p)$ be the infimum of the rationals α for which some infinite k-ary word w has both w and $\langle w\rangle_p$ free of factors of exponent exceeding α. At $p = 1$ this is Dejean's repetition threshold $\mathrm{RT}(k)$.

Our main theorem is that for every p with $1 \leq p \leq 40$ there is an infinite binary overlap-free word whose decimation modulo p is again overlap-free *if and only if* p is a power of two. The statement quantifies over *all* infinite binary overlap-free words, an uncountable set. Brown, Rampersad, Shallit and Vasiga proved the corresponding statement for the single word $\mathbf{t}$: the linear subsequence $(\mathbf{t}_{in+a})_{n\geq 0}$ is overlap-free exactly when i is a power of two. Our negative half is instead an exhaustion of the whole binary overlap-free language at each of the 34 non-powers of two $p \leq 40$, and it delivers for each such p the exact length of the longest binary overlap-free word whose decimation modulo p is overlap-free — 34 values, from 18 at $p = 3$ to 766 at $p = 38$ — which no statement about $\mathbf{t}$ alone can produce. The positive half is uniform: $\langle\mathbf{t}\rangle_{2^j} = \mathbf{t}$ for every j, with no bound on j.

We further prove $\mathrm{RT}(2,3) \geq 5/2$; $\mathrm{RT}(2,p) \geq 7/3$ for each of the seven moduli $p \in \{5,6,7,9,10,11,12\}$; and $\mathrm{RT}(3,4) \geq 11/6$. Each of these bounds on an infimum over a continuum of exponents comes from a *single* exhaustive search, through a monotonicity lever that converts one search at a non-strict threshold into a bound valid on a whole half-line of exponents. Finally we record the exact sense in which Harju's theorem is sharp: $\mathrm{RT}(3,2) = 2$, the infimum being attained just above 2 and not at 2. Every theorem, proposition and lemma stated below is machine-checked in Lean 4.

1. Introduction

Words, exponents and thresholds. Fix $k \geq 2$ and write $\Sigma_k = \{0,1,\ldots,k-1\}$. A finite word is an element of Σ_k^* and an infinite word is a map $w\colon \mathbb{N} \to \Sigma_k$, written $w = w_0w_1w_2\cdots$; a *factor* of w is a block $w_iw_{i+1}\cdots w_{j-1}$ of consecutive letters. A nonempty word u has *period* $q \geq 1$ when $u_i = u_{i+q}$ for every i with $0 \leq i < i+q < |u|$, and the *exponent* of u relative to that period is $|u|/q$. For a rational $\alpha > 1$ we say that a finite or infinite word w is

- α^+*-power-free* when no factor u of w has a period $q \geq 1$ with $|u| > \alpha q$, and
- α*-power-free* when no factor u of w has a period $q \geq 1$ with $|u| \geq \alpha q$.

With $\alpha = 2$ the second notion is *square-freeness* (no factor uu with u nonempty) and the first is *overlap-freeness* (no factor $axaxa$ with a a letter and x a word).

Dejean's *repetition threshold* is

$$\mathrm{RT}(k) \;=\; \inf\{\alpha :\ \text{some infinite word over } \Sigma_k \text{ is } \alpha^+\text{-power-free}\},$$

and its values are known: $\mathrm{RT}(2) = 2$ by Thue [16], $\mathrm{RT}(3) = 7/4$ by Dejean [4], $\mathrm{RT}(4) = 7/5$ by Pansiot [12], and $\mathrm{RT}(k) = k/(k-1)$ for $k \geq 5$, this last being Dejean's conjecture, whose remaining cases were settled by Rao [13] and, independently, by Currie and Rampersad [3].

The decimation transform. For an infinite word w and an integer $p \geq 1$, the *decimation of w modulo p* is the infinite word

$$\langle w \rangle_p \;=\; w_0\, w_p\, w_{2p}\, w_{3p} \cdots ,$$

the subsequence of w read along the arithmetic progression of difference p starting at 0; the same notation is used for a finite word, where $\langle v \rangle_p$ is the finite word of the letters of v at positions divisible by p. Harju [6] proved that for every $p \geq 3$ there is an infinite ternary square-free word whose decimation modulo p is square-free, and that for $p = 2$ there is none. Delépine, Ochem and Rosenfeld [5] take up the case of two simultaneous moduli, and a companion to the present paper [9] settles a number of their cells, including the pair $(3, 9)$ that their Proposition 28 lists as negative. All of this work keeps the alphabet ternary and the exponent equal to 2; the Dejean literature, conversely, varies the alphabet and the exponent and never leaves the identity transform.

This paper crosses the two axes. Define

(1)
$$\mathrm{RT}(k, p) \;=\; \inf\bigl\{\alpha : \text{ some infinite } w \in \Sigma_k^{\omega} \text{ has } w \text{ and } \langle w \rangle_p \text{ both } \alpha^{+}\text{-power-free}\bigr\}.$$

Since $\langle w \rangle_1 = w$, we have $\mathrm{RT}(k, 1) = \mathrm{RT}(k)$; and since the second condition only adds constraints, $\mathrm{RT}(k, p) \geq \mathrm{RT}(k)$ for every p. The question this paper asks is quantitative: *how much does the decimation constraint cost over Dejean's threshold, and for which moduli does it cost nothing?*

Results. The main theorem answers the second half of the question completely in the binary case, for $p \leq 40$.

Theorem 1.1 (Theorem 4.1)**.** *Let $1 \leq p \leq 40$. There exists an infinite binary overlap-free word w such that $\langle w \rangle_p$ is overlap-free if and only if p is a power of two.*

The positive direction holds without any bound on p: the Thue–Morse word $\mathbf{t}$ satisfies $\mathbf{t}_{2n} = \mathbf{t}_n$, hence $\langle \mathbf{t} \rangle_{2^j} = \mathbf{t}$ for every $j \geq 0$, and $\mathbf{t}$ is overlap-free by Thue's theorem. The bound $p \leq 40$ constrains only the negative direction, which is 34 exhaustive searches, one for each non-power of two $p \leq 40$. Each search terminates, and each terminating search yields more than the non-existence statement: it yields the exact maximum length of a finite witness.

Theorem 1.2 (Theorem 4.2)**.** *For each of the 34 integers $p \leq 40$ that are not powers of two, the length of the longest binary overlap-free word whose decimation modulo p is overlap-free is the value $M(p)$ recorded in Table 1; both the exhaustion at length $M(p) + 1$ and a witness of length $M(p)$ are proved.*

The comparison a reader who knows the subject will ask for is with Brown, Rampersad, Shallit and Vasiga [1, Theorem 5.3]: for integers $i \geq 1$ and $a \geq 0$, the sequence $(\mathbf{t}_{in+a})_{n \geq 0}$ is overlap-free if and only if i is a power of two. Their statement is unbounded in i and allows an arbitrary offset a, but it is about the *single* word $\mathbf{t}$. Theorem 4.1 instead quantifies over *all* infinite binary overlap-free words, an uncountable set, and no statement about $\mathbf{t}$ alone implies it: for a given p one must

rule out every infinite overlap-free word, not one distinguished sequence. The finite maxima of Table 1 are likewise out of reach of a statement about $\mathbf{t}$, which produces no finite maximum at all. The two results are bounded in different directions — theirs in the class of words, ours in the modulus — and neither contains the other. Nearest in spirit on the other side is the work of Kao, Rampersad, Shallit and Silva [7], who ask for words avoiding overlaps simultaneously in *all* progressions of odd difference and at every starting position; at an odd modulus that requirement subsumes ours, and their binary maximum is 10, a value that occurs nowhere in Table 1.

Turning to the threshold itself, we prove one two-sided value beyond the powers of two, and four lower bounds.

Theorem 1.3 (Proposition 5.4)**.** $\mathrm{RT}(3,2) = 2$*, and the infimum in* (1) *is attained just above* 2 *and not at* 2*: some infinite ternary word is overlap-free with overlap-free decimation modulo* 2*, and no infinite ternary word is square-free with square-free decimation modulo* 2*.*

This is the precise sense in which Harju's negative result at $p = 2$ is sharp: relaxing "square-free" to "overlap-free" — that is, moving the exponent from 2 to 2^+ — removes the obstruction, and nothing less does. Neither half is difficult; the content is in the assembled two-sided statement, and we record it as such.

Theorem 1.4 (Theorems 6.1, 6.2 and 6.5)**.** $\mathrm{RT}(2,3) \geq 5/2$*;* $\mathrm{RT}(2,p) \geq 7/3$ *for every* $p \in \{5,6,7,9,10,11,12\}$*; and* $\mathrm{RT}(3,4) \geq 11/6$*.*

A bound on the infimum (1) is a statement about a continuum of exponents, and could not be settled by finitely many searches were it not for the elementary monotonicity of Lemma 2.2: an α^+-power-free word is β-power-free, in the non-strict sense, for every $\beta > \alpha$. One exhaustive search at the *non-strict* threshold β therefore rules out the whole half-line of exponents $\alpha < \beta$ at once. Each of the four bounds above is one instance of this lever, and $\mathrm{RT}(2,3) \geq 5/2$ costs a single search of 325 words.

What is not proved. The matching upper bounds are open. Except along the Thue–Morse column — the powers of two for $k = 2$, and $p = 2$ for $k = 3$ — we have no infinite word at any of the conjectured values, and producing one is not a matter of more computation: it needs a morphism criterion for α^+-power-freeness in the style of Crochemore's characterisation of square-free morphisms [2] and its k-power-free descendants [15], so that finitely many finite checks certify an infinite word. Section 7 says what the searches see from above at each open cell, and marks that evidence clearly as computation outside the formal development.

Organisation. Section 2 sets up the predicates, the prefix-closure that makes the searches sound, and the exponent lever. Section 3 proves Thue's theorem and the positive half of the main theorem. Section 4 proves the main theorem and gives the table. Section 5 states $\mathrm{RT}(k,p)$ two-sidedly and settles the powers of two and $\mathrm{RT}(3,2)$. Section 6 proves the four lower bounds. Section 7 records the open cells and the numerical evidence at them, and Section 8 describes the verification.

2. Preliminaries

Throughout, $\alpha = A/B$ is a rational number greater than 1 with A, B positive integers, and $\mathrm{st} \in \{+,-\}$ records whether the threshold is strict (α^+-power-freeness,

$|u| > \alpha q$ forbidden) or non-strict (α-power-freeness, $|u| \geq \alpha q$ forbidden). Both flavours are needed: the infimum (1) is taken over the strict predicate, while the classical results this paper is measured against — Thue's, Harju's and Dejean's — are naturally stated with the non-strict one.

It is convenient to state repetition-freeness so that closure under prefixes is immediate.

Lemma 2.1. *For a finite word w the following are equivalent:* (i) *no factor of w has a period $q \geq 1$ with $|u| > \alpha q$ (resp. $\geq$);* (ii) *no prefix of w has a suffix with such a period. Consequently repetition-freeness at a fixed threshold is inherited by prefixes, and, since $\langle v \rangle_p$ is a prefix of $\langle w \rangle_p$ whenever v is a prefix of w, so is the conjunction of the two conditions defining* (1).

Formulation (ii) is the definition used in the formal development; (i) is the usual factor form. The next lemma is what makes a single search speak about an interval of exponents.

Lemma 2.2 (monotonicity in the exponent)**.** *Let $\alpha = A/B$ and $\beta = A'/B'$ be positive rationals and w any word.*

1. *If $\alpha \leq \beta$ and w is α^+-power-free, then w is β^+-power-free.*
2. *If $\alpha < \beta$ and w is α^+-power-free, then w is β-power-free in the non-strict sense.*

Clause (1) says that the set of exponents appearing in (1) is upward closed, so that the infimum there is the infimum of an interval and a lower bound for it is exactly a statement that small exponents are unattainable. Clause (2) with $(A', B') = (2, 1)$ reads: for every $\alpha < 2$, an α^+-power-free word is square-free — which is how Harju's theorem will cover a half-line of exponents in Proposition 5.4.

Definition 2.3. Let $p \geq 1$. A finite word v is (α, st, p)-*admissible* when v and $\langle v \rangle_p$ are both repetition-free at the threshold α with strictness st; an infinite word w is admissible when w and $\langle w \rangle_p$ are.

Lemma 2.4 (the finite/infinite bridge)**.** *Let $p \geq 1$ and let w be an infinite word. For every n there is an L with $n \leq Lp < n+p$ such that the decimation of the length-n prefix of w is exactly the length-L prefix of $\langle w \rangle_p$. Consequently w is admissible if and only if every prefix of w is admissible.*

The searches of Sections 4–6 enumerate admissible words by length. Two observations make them cheap and make them mean something.

Proposition 2.5. *Fix $\alpha = A/B > 1$, a strictness, an alphabet size k and a modulus $p \geq 1$.*

1. (One test per period.) *For a fixed period $q \geq 1$ the set of lengths L at which a factor of period q is forbidden is an up-set with least element*

$$\mathrm{minLen}(q) \;=\; \left\lfloor \tfrac{Aq}{B} \right\rfloor + 1 \quad (\textit{strict}), \qquad \mathrm{minLen}(q) \;=\; \left\lceil \tfrac{Aq}{B} \right\rceil \quad (\textit{non-strict}),$$

   and $\mathrm{minLen}(q) > q$. Hence deciding whether appending one letter to an admissible word creates a forbidden repetition ending at the new position is, for each q, a single comparison of two blocks of the fixed length $\mathrm{minLen}(q) - q$.
2. (Exhaustion.) *If there is no admissible word of some length N over Σ_k, then there is no infinite admissible word over Σ_k.*

Part (2) is the criterion that Delépine, Ochem and Rosenfeld state as "If the search terminates, then no such infinite word exists" [5], here at an arbitrary alphabet size and an arbitrary rational threshold; it is immediate from Lemma 2.1 and Lemma 2.4. Part (1) is the design point that keeps the general exponent as cheap as the square case: at exponent 2 one tests for a square at the new position by walking half-lengths, but at a general rational threshold the exponent of a factor grows with its length at fixed period, so there is a single shortest forbidden length per period and no scan over lengths is required.

3. The Thue–Morse word

Let $\mathbf{t}\colon \mathbb{N} \to \{0,1\}$ be the Thue–Morse word, $\mathbf{t}_n$ being the parity of the number of 1s in the binary expansion of n, so that

$$\mathbf{t}_{2n} = \mathbf{t}_n, \qquad \mathbf{t}_{2n+1} = 1 - \mathbf{t}_n. \tag{2}$$

Its first eight letters are 0 1 1 0 1 0 0 1.

Theorem 3.1 (Thue [16])**.** *The Thue–Morse word is overlap-free: there are no $i \geq 0$ and $q \geq 1$ with $\mathbf{t}_{i+j} = \mathbf{t}_{i+j+q}$ for all $0 \leq j \leq q$.*

Proof sketch. The result is classical; we sketch the argument used in the formal development because it differs from the textbook one and uses no computation. Everything is derived from (2) by descent on q, with no appeal to the morphism $0 \mapsto 01$, $1 \mapsto 10$ or to a synchronisation lemma.

First, $\mathbf{t}$ has no three consecutive equal letters: of the two pairs $(n, n+1)$, $(n+1, n+2)$ one is of the form $(2m, 2m+1)$, and $\mathbf{t}_{2m} = \mathbf{t}_{2m+1}$ reads $\mathbf{t}_m = 1 - \mathbf{t}_m$ by (2). This disposes of $q = 1$. If $q = 2r$ is even, then reading the hypothesis at the positions of a single parity class turns it into the same hypothesis for the period r at a smaller starting index, and the induction descends. If $q = 2r+1$ is odd with $r \geq 1$, the constraints at positions $2m$ and $2m+1$ land, by (2), on *opposite* sides of $\mathbf{t}_m$: one gives $\mathbf{t}_m = 1 - \mathbf{t}_{m+r}$ and the other $1 - \mathbf{t}_m = \mathbf{t}_{m+r+1}$, whence $\mathbf{t}_{m+r} = \mathbf{t}_{m+r+1}$. Applying this at two consecutive values of m produces three consecutive equal letters, which the first step forbids. The two parities of the starting position require different pairings, $(2m, 2m+1)$ in one case and $(2m+1, 2m+2)$ in the other. □

Proposition 3.2. *For every $j \geq 0$ we have $\langle \mathbf{t} \rangle_{2^j} = \mathbf{t}$. Consequently $\mathbf{t}$ is an infinite overlap-free word whose decimation modulo 2^j is overlap-free, for every $j \geq 0$; and since a binary word is a legal word over any larger alphabet, the same holds when $\mathbf{t}$ is read over Σ_k for any $k \geq 2$.*

Proof sketch. Iterating $\mathbf{t}_{2n} = \mathbf{t}_n$ gives $\mathbf{t}_{n2^j} = \mathbf{t}_n$ for all n and j, which is the asserted equality of infinite words; the rest is Theorem 3.1. □

Proposition 3.2 is the positive half of the main theorem, and it is uniform in j: no bound on j enters. The positive direction is also, for the single word $\mathbf{t}$, the easy half of [1, Theorem 5.3].

4. The power-of-two theorem

Theorem 4.1. *Let $1 \leq p \leq 40$. There exists an infinite binary overlap-free word w with $\langle w \rangle_p$ overlap-free if and only if $p \in \{1, 2, 4, 8, 16, 32\}$, i.e. if and only if p is a power of two.*

TABLE 1. For each non-power of two $p \leq 40$: the maximum length $M(p)$ of a binary overlap-free word whose decimation modulo p is overlap-free, and the number $N(p)$ of such words of all lengths — that is, the size of the whole search tree.

p	$M(p)$	$N(p)$	p	$M(p)$	$N(p)$
3	18	233	23	320	91 577
5	80	4 125	24	168	32 895
6	42	1 291	25	500	107 285
7	77	5 043	26	604	235 639
9	72	4 017	27	297	73 965
10	160	20 365	28	336	121 347
11	217	29 133	29	396	148 311
12	84	6 535	30	270	79 233
13	302	49 403	31	372	112 879
14	154	25 007	33	297	108 075
15	132	15 983	34	408	139 415
17	204	28 797	35	494	234 567
18	144	20 283	36	288	102 631
19	383	95 503	37	592	232 009
20	320	102 301	38	766	477 383
21	378	56 295	39	546	190 021
22	434	144 195	40	640	503 899

Proof sketch. The "if" direction is Proposition 3.2 and holds for every power of two, in range or not. The "only if" direction is 34 separate assertions, one for each $p \leq 40$ that is not a power of two, each of the form "no infinite binary overlap-free word has overlap-free decimation modulo p". By Proposition 2.5(2), each follows from the existence of one length N_p at which the enumeration of admissible words is empty, and Theorem 4.2 supplies such an N_p, namely $M(p) + 1$. Each of the 34 enumerations is a finite exhaustive computation, performed inside the proof assistant; Section 8 describes them and their independent reproduction. No other case analysis enters: the 40 values of p are split into the six powers of two and the 34 exhausted moduli. □

Theorem 4.2. *Let $p \leq 40$ be one of the 34 integers that are not powers of two and let $M(p)$ be the value listed in Table 1. Then*

(1) *no binary word of length $M(p)+1$ is overlap-free with overlap-free decimation modulo p; and*
(2) *some binary word of length $M(p)$ is overlap-free with overlap-free decimation modulo p.*

Thus $M(p)$ is exactly the length of the longest such word.

Proof sketch. Both halves rest on exhaustive computation, and each is stated and checked separately for each of the 34 moduli. For (1) the enumeration of admissible words is carried out level by level, from length 0 up to length $M(p) + 1$, and the level at $M(p) + 1$ is found to be empty; the enumeration is legitimate because admissibility is closed under prefixes (Lemma 2.1), so a word of length N can survive only if all of its prefixes do. The number of words visited in the whole

enumeration is the value $N(p)$ of Table 1. For (2) an explicit word of length $M(p)$ is exhibited and its admissibility verified by a single evaluation; the witnesses were produced by an independent depth-first implementation of the same search and transcribed mechanically. The two halves together also give, for every $n \le M(p)$, an admissible word of length n, by taking a prefix. □

Remark 4.3 (the table is not a doubling law). Of the 15 moduli p for which both p and $2p$ appear in Table 1, twelve satisfy $M(2p) = 2M(p)$, namely $p \in \{5, 6, 7, 9, 10, 11, 12, 13, 17, 18, 19, 20\}$. Three do not: $M(3) = 18$ but $M(6) = 42 \neq 36$; $M(14) = 154$ but $M(28) = 336 \neq 308$; and $M(15) = 132$ but $M(30) = 270 \neq 264$. So the table is not explained by any simple doubling rule, and in particular it is not a repackaging of the power-of-two dichotomy.

Proposition 4.4. *The Thue–Morse word itself is not overlap-free modulo* 3*; that is,* $\langle \mathbf{t} \rangle_3$ *contains an overlap.*

Proof sketch. Immediate from the case $p = 3$ of Theorem 4.1, applied to the overlap-free word $\mathbf{t}$. We record it because it shows that the power-of-two hypothesis in Proposition 3.2 is doing work rather than being satisfied vacuously. □

Remark 4.5. Theorem 4.2 produces 34 numbers that are, to our knowledge, new; they are not in the OEIS, neither the sequence of maxima nor the sequence of tree sizes, and no statement about the single word $\mathbf{t}$ can produce a finite maximum. The nearest published maximum for a related binary condition is the value 10 of [7], for the stronger condition of avoiding overlaps in all progressions of odd difference, at every starting position, at once.

5. The threshold $\mathrm{RT}(k, p)$

Say that a rational α is *attainable* at (k, p) when some infinite word over Σ_k is α^+-power-free with α^+-power-free decimation modulo p; by Lemma 2.2(1) the attainable exponents form an up-set, and $\mathrm{RT}(k, p)$ of (1) is the infimum of that up-set. We write $\mathrm{RT}(k, p) = c$ for the two-sided statement that c is attainable and that no $\alpha < c$ is; this is the form an infimum takes when one wants to say, in addition, whether it is attained.

Lemma 5.1 (the exponent lever)**.** *Let* β *be a rational and suppose that for some* N *there is no word of length* N *over* Σ_k *that is* β*-power-free (non-strict) with* β*-power-free decimation modulo* p*. Then no* $\alpha < \beta$ *is attainable at* (k, p)*; that is,* $\mathrm{RT}(k, p) \ge \beta$*.*

Proof sketch. If $\alpha < \beta$ were attainable, witnessed by w, then by Lemma 2.2(2) both w and $\langle w \rangle_p$ would be β-power-free in the non-strict sense, and Proposition 2.5(2) applied at the non-strict threshold β would contradict the hypothesis. □

The lever is the reason the lower bounds of Section 6 are affordable: a single terminating search at one non-strict threshold settles a whole half-line of exponents.

Proposition 5.2. $\mathrm{RT}(2, p) \ge 2$ *for every modulus* $p \ge 1$*.*

Proof sketch. No binary word of length 4 is square-free — an exhaustive check over a tree of 7 words, performed by the proof assistant's kernel without any compiled evaluation. Hence for $\alpha < 2$ the first condition alone already fails, by Lemma 2.2(2), uniformly in p and with no search per modulus. □

Proposition 5.3. $\mathrm{RT}(2, 2^j) = 2 = \mathrm{RT}(2)$ *for every* $j \geq 0$*: at a power-of-two modulus the decimation constraint costs nothing over Dejean's threshold.*

Proof sketch. Attainability of 2 is Proposition 3.2; unattainability below 2 is Proposition 5.2. Both halves are one step from published results — the positive direction of [1] and Thue's theorem — and we state the composition only because it is the contrast that makes Theorem 4.1 a statement about a threshold rather than a table: the powers of two are exactly the moduli at which the cost is zero, and Table 1 measures a cost that is not zero everywhere else. □

Proposition 5.4. $\mathrm{RT}(3, 2) = 2$*, with the infimum attained just above* 2 *and not at* 2*. Explicitly:*

(1) *some infinite ternary word is overlap-free with overlap-free decimation modulo* 2*;*
(2) *no infinite ternary word is square-free with square-free decimation modulo* 2*; indeed no ternary word of length* 15 *is, the longest such word having length* 14*, for example* 01201021012010*;*
(3) *hence no* $\alpha < 2$ *is attainable at* $(3, 2)$*.*

Proof sketch. For (1), a binary word is a legal ternary word, so $\mathbf{t}$ read over Σ_3 works by Proposition 3.2 with $j = 1$. Part (2) is Harju's theorem [6] at $p = 2$; we reprove it here by exhaustion, the whole search tree consisting of 178 words and dying at length 15. Part (3) is Lemma 5.1 applied to (2) with $\beta = 2$. □

Each half of Proposition 5.4 is a single step from a published theorem; what is new is only the two-sided reading, which answers the question "how far must the exponent in Harju's theorem be relaxed?" with: to 2^+, and no further, and 2 itself does not suffice.

Remark 5.5 (indexing). Harju indexes positions from 1 and reads the progression $w(p)\, w(2p)\, w(3p) \cdots$, whereas $\langle w \rangle_p$ reads positions $0, p, 2p, \ldots$. For the existence question the two conventions agree: deleting the first $p-1$ letters carries one to the other, and a suffix of a repetition-free word is repetition-free. They do not agree on the extremal finite length, which at $p = 2$ is 14 in the indexing used here and 15 in Harju's — a shift of one caused by the convention alone. Only the first of the two numbers is part of the formal development.

Proposition 5.6 (Dejean's floor)**.** *No ternary word of length* 39 *is* $7/4$*-power-free in the non-strict sense; no* 4*-ary word of length* 122 *is* $7/5$*-power-free; no* 5*-ary word of length* 7 *is* $5/4$*-power-free. Consequently* $\mathrm{RT}(3, p) \geq 7/4$ *and* $\mathrm{RT}(4, p) \geq 7/5$ *for every modulus* $p \geq 1$*.*

Proof sketch. The three non-existence statements are exhaustive searches at $p = 1$, the ternary and 4-ary searches visiting 3 196 and 236 345 words respectively; the longest words they leave have lengths 38, 121 and 6, of which the first and the last are the classical extremal lengths at the Dejean exponents $7/4$ for $k = 3$ and $5/4$ for $k = 5$ [4]. Only the three exhaustions belong to the formal development; the attaining words do not. The consequence for arbitrary p needs no search: a word witnessing attainability at (k, p) is in particular a witness at $(k, 1)$, and Lemma 2.2(2) converts a strict threshold below $7/4$ (resp. $7/5$) into the non-strict threshold that the search refutes. □

Proposition 5.6 is included for two reasons. It is the floor $\mathrm{RT}(k,p) \geq \mathrm{RT}(k)$ made explicit at $k = 3, 4$; and it is a control on the machinery, since the searches land on the classical extremal lengths 38 and 6 before any new cell is asserted.

6. Lower bounds at four cells

Theorem 6.1. $\mathrm{RT}(2,3) \geq 5/2$. *Explicitly, no infinite binary word w has both w and $\langle w \rangle_3$ α^+-power-free for any $\alpha < 5/2$. The bound comes from one search: no binary word of length* 19 *is* $5/2$*-power-free (non-strict) with* $5/2$*-power-free decimation modulo* 3*, and the longest such word has length* 18*, for example* 001001101001100100.

Proof sketch. The search at the non-strict threshold $5/2$ and modulus 3 visits 325 words in total and its level at length 19 is empty; the witness of length 18 is checked by evaluation. Both facts are finite computations. Lemma 5.1 with $\beta = 5/2$ turns the first into the assertion for every $\alpha < 5/2$ at once. □

Theorem 6.2. $\mathrm{RT}(2,p) \geq 7/3$ *for each $p \in \{5,6,7,9,10,11,12\}$. Explicitly, for each such p there is no binary word of the length N_p tabulated below that is* $7/3$*-power-free (non-strict) with* $7/3$*-power-free decimation modulo p; hence every such word has length at most $N_p - 1$, and no $\alpha < 7/3$ is attainable at $(2,p)$.*

p	5	6	7	9	10	11	12
N_p	94	44	103	82	188	236	88
$N_p - 1$	93	43	102	81	187	235	87

Proof sketch. Seven exhaustive searches, one per modulus, all at the single non-strict threshold $7/3$; the largest of them visits at most 96 427 words. Lemma 5.1 with $\beta = 7/3$ converts each into the corresponding statement about all $\alpha < 7/3$. □

Remark 6.3. Theorem 6.2 asserts only the exhaustion, so the tabulated $N_p - 1$ are proved to be upper bounds for the length of an admissible word. Our searches outside the proof assistant find admissible words of exactly those lengths, so we expect each $N_p - 1$ to be the exact maximum, as in Theorems 4.2 and 6.1; but unlike there, the attaining witnesses at these seven cells are not part of the formal development, and the numbers $N_p - 1$ are asserted here as upper bounds only.

Remark 6.4. Karhumäki and Shallit [8] showed that $7/3$ is exactly the dividing line between polynomial and exponential growth for binary repetition-free languages: there are only polynomially many binary words of length n avoiding $7/3$-powers, and exponentially many avoiding $(7/3)^+$-powers. The searches of Theorem 6.2 run at the non-strict threshold, on the polynomial side of that line. The data say that the decimation constraint at these moduli becomes satisfiable at, or very near, the transition; see Section 7 for what the searches show just below and just above it.

Theorem 6.5. $\mathrm{RT}(3,4) \geq 11/6$. *Explicitly, no ternary word of length* 105 *is* $11/6$*-power-free (non-strict) with* $11/6$*-power-free decimation modulo* 4*; the longest such word has length* 104*; and hence no $\alpha < 11/6$ is attainable at $(3,4)$.*

Proof sketch. One exhaustive search at the non-strict threshold $11/6$ and modulus 4, visiting 16 570 words and dying at length 105, together with an explicit admissible word of length 104, checked by evaluation. Lemma 5.1 with $\beta = 11/6$ gives the statement about all $\alpha < 11/6$. □

Theorem 6.5 places the cell $(3, 4)$ strictly above Dejean's ternary threshold $7/4$ of Proposition 5.6: at $p = 4$ the decimation constraint has a positive cost, at least $11/6 - 7/4 = 1/12$.

7. The open cells, and what the searches see

Every upper bound in this paper, other than along the Thue–Morse column of Propositions 3.2, 5.3 and 5.4, is missing, and the reason is structural rather than computational. An upper bound $\mathrm{RT}(k, p) \le \alpha$ requires an *infinite* word, and the standard way to produce one is to apply a morphism to a known infinite word and certify the image by finitely many finite tests — Crochemore's criterion [2] does exactly this for square-free morphisms, and the companion paper [9] uses it to settle several cells of the two-modulus problem, among them the pair $(3, 9)$. What is required here is the analogue for α^+-power-freeness at a general rational α, in the style of the test-set criteria for k-power-free morphisms [15]. We do not have it, and without it no amount of search produces an infinite word. The gap is exactly at the exponent, not at the method: the square-free criterion exists and does its work one family over; it is the criterion at a general rational exponent that has no counterpart.

The following measurements were made with search programs outside the proof assistant and are *not* part of the formal development; we record them because they say what value each open cell should have. In each case the search was capped at words of 20 000 letters, and "reaches the cap" means that the enumeration produced a word of that length without exhausting; that behaviour is what distinguishes an open positive cell from a negative one in practice.

- $(2, 3)$. At the strict threshold $5/2$ the search reaches the cap, after 67 446 words, so the expected value is $\mathrm{RT}(2, 3) = 5/2$ exactly. Below it, every exponent from 2 up to $52/21$ gives the same maximum length 18 as Theorem 6.1; the search tree, however, is not the same, having 233 words at 2^+ and 325 from $27/11^+$ onwards. The obstruction at this cell therefore localises at a single period, and $5/2$ is exactly the exponent at which factors *ababa* of period 2 become legal.
- $(2, p)$ *for* $p \in \{5, 6, 7, 9, 10, 11, 12\}$. At the strict threshold $7/3$ every one of the seven searches reaches the cap, and just below, at $9/4$, every one of them exhausts. So the expected value is $\mathrm{RT}(2, p) = 7/3$ at all seven moduli.
- *Beyond the seven moduli.* The same non-strict exhaustion at $7/3$ terminates for *every* non-power of two $p \le 24$, with longest admissible words of lengths 305, 205, 155, 238, 162, 417, 375, 405, 471, 377, 175 at $p = 13, 14, 15, 17, 18, 19, 20, 21, 22, 23, 24$. Only the seven cells of Theorem 6.2 were carried into the formal development.
- *The ternary column.* At the strict threshold $7/4$ — Dejean's own value — the search reaches the cap at $p = 3, 5, 6, 7, 8, 9, 12$, so at those moduli the decimation appears to be free, $\mathrm{RT}(3, p) = 7/4 = \mathrm{RT}(3)$; the lower bound $\mathrm{RT}(3, p) \ge 7/4$ is proved for every p in Proposition 5.6. The exceptions found are $p = 2$, settled in Proposition 5.4, and $p = 4$, bounded from below in Theorem 6.5 and reaching the cap at $11/6$, so that $\mathrm{RT}(3, 4) = 11/6$ is the expected value. Note that $p = 8$ is a power of two and is *not* exceptional: for the ternary alphabet the exceptional moduli found are 2 and 4 specifically,

and not the powers of two as a class — the exact opposite of the binary picture of Theorem 4.1.

- *Cube-free binary words.* The analogue of Theorem 4.1 one exponent higher is uneventful: at the threshold 3 every modulus $p \le 12$, and every pair of moduli tested, reaches the cap within a few thousand words. The phenomenon isolated in this paper is a threshold phenomenon and does not survive to exponent 3.

Two further problems are worth naming. First, the negative half of Theorem 4.1 for all p, not only $p \le 40$: the natural route is the factorisation $w = u\,\mu(w')$ of infinite binary overlap-free words [14] combined with a parity recursion of the kind used in [1]; the recursion carries a complementation, which is where the difficulty lies. Second, the exact value of $\mathrm{RT}(k, p)$ as a function of p: Table 1 shows that the finite data are not governed by any obvious arithmetic of p (Remark 4.3), so even a conjectural formula would be of interest.

8. Verification

Every statement asserted above as a lemma, proposition or theorem has been formalised and machine-checked in Lean 4 against Mathlib [10, 11], and no proof in the development is incomplete. The definitions of Section 2 are the formal ones — repetition-freeness at a rational threshold with a strictness flag, the decimation, admissibility for finite and infinite words, and the level-by-level enumeration — and Lemma 2.4 and Proposition 2.5 are theorems there, so the passage from a terminating search to a statement about infinite words is itself checked rather than assumed. Theorem 3.1, Proposition 3.2, Proposition 5.2, Proposition 5.3, Lemma 5.1 and Proposition 2.5 are checked by the kernel alone, with no compiled evaluation: in particular $\mathrm{RT}(2, 2^j) = 2$ and the overlap-freeness of the Thue–Morse word depend on no computation. The exhaustive searches — the 34 cells of Theorem 4.2 and hence Theorem 4.1, the searches of Propositions 5.4 and 5.6, and those of Theorems 6.1, 6.2 and 6.5 — and the checks of the explicit witnesses are discharged by compiled evaluation inside the proof assistant, so those statements additionally rest on the assumption that the compiler's evaluation of a closed boolean term agrees with its logical value. In total the development evaluates 4 069 126 admissible words, of which 3 599 640 are the cells of Table 1. The development also contains the controls one wants against a vacuous formalisation: that $\langle w\rangle_3$ really reads positions $0, 3, 6, \ldots$; that $\langle w\rangle_1 = w$; that 000 and 01010 are *not* overlap-free, with the offending period named in each case; and that 0110 is overlap-free but not square-free, so that the two predicates are distinct and neither is vacuous. Independently of the proof assistant, a depth-first search written in C reproduces every value of Table 1, both the maxima and the tree sizes; before any new cell was believed it was also checked against published numbers, reproducing the classical extremal lengths at the Dejean exponents for alphabet sizes $2, 3, 4, 5$ and the length-14, 178-word cell of Proposition 5.4, which is the case $p = 2$ of Harju's theorem. Repository reference to be added at submission.

References

[1] S. Brown, N. Rampersad, J. Shallit, T. Vasiga, *Squares and overlaps in the Thue–Morse sequence and some variants*, RAIRO Theor. Inform. Appl. **40** (2006), no. 3, 473–484, doi:10.1051/ita:2006030.

[2] M. Crochemore, *Sharp characterizations of squarefree morphisms*, Theoret. Comput. Sci. **18** (1982), no. 2, 221–226, doi:10.1016/0304-3975(82)90023-8.

[3] J. Currie, N. Rampersad, *A proof of Dejean's conjecture*, Math. Comp. **80** (2011), no. 274, 1063–1070, doi:10.1090/S0025-5718-2010-02407-X.

[4] F. Dejean, *Sur un théorème de Thue*, J. Combin. Theory Ser. A **13** (1972) 90–99, doi:10.1016/0097-3165(72)90011-8.

[5] T. Delépine, P. Ochem, M. Rosenfeld, *Pairs of square-free arithmetic progressions in infinite words*, arXiv:2605.29853v1, 28 May 2026.

[6] T. Harju, *On square-free arithmetic progressions in infinite words*, Theoret. Comput. Sci. **770** (2019) 95–100, doi:10.1016/j.tcs.2018.09.032.

[7] J.-Y. Kao, N. Rampersad, J. Shallit, M. Silva, *Words avoiding repetitions in arithmetic progressions*, Theoret. Comput. Sci. **391** (2008), no. 1–2, 126–137, doi:10.1016/j.tcs.2007.10.039; arXiv:math/0608607. The binary maximum 10 quoted above is the optimality remark of arXiv version v2.

[8] J. Karhumäki, J. Shallit, *Polynomial versus exponential growth in repetition-free binary words*, J. Combin. Theory Ser. A **105** (2004), no. 2, 335–347, doi:10.1016/j.jcta.2003.12.004.

[9] Korea Superintelligence Labs, *Square-free words with square-free arithmetic subsequences: a correction at* $(3, 9)$*, and the first data on shifted triples*, companion manuscript, 2026.

[10] L. de Moura, S. Ullrich, *The Lean 4 theorem prover and programming language*, in: Automated Deduction — CADE 28, Lecture Notes in Computer Science **12699**, Springer, 2021, 625–635, doi:10.1007/978-3-030-79876-5_37.

[11] The mathlib Community, *The Lean mathematical library*, in: Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020), ACM, 2020, 367–381, doi:10.1145/3372885.3373824.

[12] J.-J. Pansiot, *A propos d'une conjecture de F. Dejean sur les répétitions dans les mots*, Discrete Appl. Math. **7** (1984) 297–311, doi:10.1016/0166-218X(84)90006-4.

[13] M. Rao, *Last cases of Dejean's conjecture*, Theoret. Comput. Sci. **412** (2011), no. 27, 3010–3018, doi:10.1016/j.tcs.2010.06.020.

[14] A. Restivo, S. Salemi, *Overlap-free words on two symbols*, in: M. Nivat, D. Perrin (eds.), Automata on Infinite Words, Lecture Notes in Computer Science **192**, Springer, 1985, 198–206, doi:10.1007/3-540-15641-0_35.

[15] G. Richomme, F. Wlazinski, *Some results on k-power-free morphisms*, Theoret. Comput. Sci. **273** (2002), no. 1–2, 119–142, doi:10.1016/S0304-3975(00)00437-0.

[16] A. Thue, *Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen*, Norske Vid. Selsk. Skr. I Mat.-Nat. Kl. Christiania, no. 1 (1912) 1–67; JFM 44.0462.01.