A rich word over five letters with critical exponent $E^*(u_9)$, and the bound $\mathrm{RT}(C_5) \le E^*(u_9) \approx 2.0592854$

Qianghao Wu

November 26, 2026

Abstract

A finite word of length n is rich if it has $n+1$ distinct palindromic factors, counting the empty word. The repetition threshold $\mathrm{RT}(C_d)$ is the infimum of the critical exponents of the infinite rich words over d letters. Let u_d be the fixed point of $\varphi_d\colon i \mapsto 0(i+1)$ for $i < d-1$, $d-1 \mapsto 0(d-1)(d-1)$. Dvořáková and Pelantová conjectured that $\mathrm{RT}(C_d) = E^*(u_{2d-1})$, where E^* is the asymptotic critical exponent. The conjecture is known for $d = 2, 3, 4$. Dvořáková, Mol and Pelantová (2026) named a first step for $d \ge 5$: find a morphism $\pi\colon \Sigma_{2d-1}^* \to \Sigma_d^*$ such that $\pi(u_{2d-1})$ is rich and has critical exponent $E^*(u_{2d-1})$. We give such a morphism for $d = 5$. The letter images have lengths $2, 3, 1, 2, 4, 6, 11, 19, 36$. So $\mathrm{RT}(C_5) \le E^*(u_9) = 1 + 1/(3-\Lambda) \approx 2.0592854490$, where Λ is the root of $t^5 - 2t^4 - 1$ in $(2, 3)$. The proof reduces the critical exponent to bispecial factors and their return words, by a theorem of Dolce, Dvořáková and Pelantová. We describe all bispecial factors of $\pi(u_9)$ from seven exceptional bispecial factors of u_9 and sixteen infinite chains. A program compares each ratio with $1/(3-\Lambda)$ in exact arithmetic, and bounds the tail of each chain with a closed form. Two programs written separately confirm the result, and a brute-force search on long prefixes agrees with it. The lower bound $\mathrm{RT}(C_5) \ge E^*(u_9)$ remains open. This work has not yet been reviewed by experts. The computations, the proofs and the text were produced by AI models (from Anthropic and OpenAI), working under the author's direction.

Status: Computer-assisted. This work has not yet been reviewed by experts. There is no proof-assistant formalization.

1 Introduction

A finite word w of length n has at most $n+1$ distinct palindromic factors, counting the empty word. It is *rich* if it has exactly $n+1$. An infinite word is rich if all of its factors are rich. Rich words were introduced by Droubay, Justin and Pirillo [4]. For a nonempty word v and a rational $e \ge 1$, the word v^e is the prefix of length $e|v|$ of $vvv\cdots$. The *critical exponent* of an infinite word x is

$$E(x) = \sup\{e \in \mathbb{Q} : v^e \text{ is a factor of } x \text{ for a nonempty word } v\},$$

and the *asymptotic critical exponent* $E^*(x)$ is the $\limsup_{n\to\infty}$ of the same set with $|v| = n$. Always $E^*(x) \le E(x)$. With $\Sigma_d = \{0, 1, \ldots, d-1\}$ and C_d the class of infinite rich words over Σ_d, the source paper [6] defines

$$\mathrm{RT}(C_d) = \inf\{E(x) : x \text{ is an infinite rich word over } \Sigma_d\}.$$

The words u_d. Let $\varphi_d(i) = 0(i+1)$ for $i \in \{0, \dots, d-2\}$ and $\varphi_d(d-1) = 0(d-1)(d-1)$, and let u_d be the fixed point of φ_d. Dvořáková, Klouda and Pelantová [5] showed that u_d is rich and that $E^*(u_d) = 1 + 1/(3 - \Lambda_d)$, with Λ_d the spectral radius of the incidence matrix of φ_d. They also showed that the asymptotic repetition threshold of rich words is 2 for every alphabet size. Dvořáková and Pelantová [8] proved that $\mathrm{RT}(C_d) \to 2$ as $d \to \infty$, by morphic images of u_{2d+1} over $d + 2^d$ letters, and they stated the conjecture below.

Known values. Currie, Mol and Rampersad [2] proved $\mathrm{RT}(C_2) = E^*(u_3) \approx 2.70710678$. Currie, Mol and Peltomäki [1] proved $\mathrm{RT}(C_3) = E^*(u_5) \approx 2.25876324$; the morphic form of their extremal word is due to Dvořáková and Pelantová [7]. Dvořáková, Mol and Pelantová [6] proved a structure theorem for $d = 4$ and obtained $\mathrm{RT}(C_4) = E^*(u_7) \approx 2.1197196907$ [6, Corollary 1.6].

Conjecture 1 ([6, Conjecture 1.1], from [8])**.** *For every integer $d \geq 2$, we have $\mathrm{RT}(C_d) = E^*(u_{2d-1})$.*

Section 8 of [6] lists open questions. Its first item reads:

> "A first step towards proving Conjecture 1.1 for every $d \geq 5$ would be to find a projection $\pi = \pi_d : \Sigma^*_{2d-1} \to \Sigma^*_d$ such that $\pi(u_{2d-1})$ is rich and $E(\pi(u_{2d-1})) = E^*(u_{2d-1})$. This would prove the inequality $\mathrm{RT}(C_d) \leq E^*(u_{2d-1})$."

In [6] a *projection* is a morphism from the larger alphabet to the smaller one; letter images can be long (Table 1 of [6]). We do this step for $d = 5$, the first open case.

Theorem 2. *Let $\pi\colon \Sigma^*_9 \to \Sigma^*_5$ be the morphism of Table 1, and let $z = \pi(u_9)$. Then z is rich and*

$$E(z) = E^*(u_9) = 1 + \frac{1}{3 - \Lambda} \approx 2.0592854490,$$

where $\Lambda \approx 2.0559673967$ is the root of $t^5 - 2t^4 - 1$ in $(2, 3)$. Hence $\mathrm{RT}(C_5) \leq E^*(u_9)$.

i	$\pi(i)$	$\lvert\pi(i)\rvert$
0	01	2
1	011	3
2	0	1
3	02	2
4	0232	4
5	023432	6
6	02343223432	11
7	0234344343343443432	19
8	023434434334344343234344343343443432	36

Table 1: The projection π (candidate A in the programs). Each $\pi(i)$ is 0 followed by a palindrome without 0.

What is new. The searches recorded with this result (citations of [6] and [8], arXiv, GitHub, Zenodo and other sources, last checked on November 26, 2026) found no projection for $d = 5$ and no upper bound on $\mathrm{RT}(C_5)$ at $E^*(u_9)$. The public code of [6] covers $d = 3$ and $d = 4$ only. The new content is the morphism of Table 1 and the computer-assisted proof of Theorem 2. We do not prove the lower bound $\mathrm{RT}(C_5) \geq E^*(u_9)$. In the cases $d \leq 4$ that bound came from a structure theorem [2, 1, 6].

The status of the proof. The proof cites three published results (Section 2) and uses a computer for a finite list of exact checks and for an eigenvalue estimate in 60-digit floating point. Section 6 lists all assumptions.

2 Cited results

For a factor w of an infinite word x, let $\mathrm{Bext}(w) = \{(a,b) : awb \text{ is a factor of } x\}$, with a, b letters. The word w is *bispecial* if it has at least two left and two right extensions. Its *bilateral order* is $b(w) = \#\,\mathrm{Bext}(w) - \#\{a\} - \#\{b\} + 1$, and for a palindrome w, $\mathrm{Pext}(w) = \{(a,a) \in \mathrm{Bext}(w)\}$. If $j < \ell$ are consecutive occurrences of w in $x = x_0 x_1 \cdots$, then $x_j \cdots x_{\ell-1}$ is a *return word* to w.

(T1) [3, Theorem 3], quoted as Theorem 7.2 in [6]. If x is uniformly recurrent and aperiodic, then $E(x) = 1 + \sup\{|w|/|r| : w \text{ bispecial in } x,\ r \text{ a return word to } w\}$.

(T2) [8, Theorem 4]. If the language of x is closed under reversal, then x is rich if and only if every bispecial factor w satisfies $b(w) = 0$ when w is not a palindrome, and $b(w) = \#\,\mathrm{Pext}(w) - 1$ when w is a palindrome.

(T3) [5], as listed in [8, Section 2]: the language of u_d is closed under reversal, and $E^*(u_d) = 1 + 1/(3 - \Lambda_d)$, with Λ_d the root in $(2,3)$ of $t^{d-1}(t-2)^2 - 1$. For $d = 9$ we have $t^8(t-2)^2 - 1 = (t^5 - 2t^4 - 1)(t-1)(t^4 - t^3 - t^2 - t - 1)$, so $\Lambda_9 = \Lambda$.

The source [6, Proposition 7.4] (from [8, Proposition 19]) gives a shorter richness test for morphisms of class P_{ret}. Our π is of this class with marker 0. We check its hypothesis, but the proof below uses (T2) on every bispecial factor, so it does not depend on that proposition.

3 Bispecial factors of u_9 and of z

Write $u = u_9$, $\varphi = \varphi_9$, M for the incidence matrix of φ (9×9), $h_i = |\pi(i)|$, and $\vec{w}$ for the Parikh vector of a word w, so that $|\pi(w)| = h \cdot \vec{w}$. Put $P_5(t) = t^5 - 2t^4 - 1$, $Q_4(t) = t^4 - t^3 - t^2 - t - 1$ and $q = 1/(3 - \Lambda) \approx 1.0592854490$. The characteristic polynomial of M is $P_5 Q_4$, with nine distinct roots. The four roots of P_5 other than Λ have modulus at most 0.8886. The roots of Q_4 include one real root $\beta \approx 1.9276$.

Lemma 3 (blocks)**.** *The letter* 0 *occurs in* u *exactly at the first letter of each block* $\varphi(u_j)$*. So the letters* $1, \ldots, 7$ *are always preceded and followed by* 0*, and* 88 *is always preceded and followed by* 0*.*

Lemma 4 (desubstitution)**.** *The bispecial factors of* u *of length at most* 1 *are* ε*,* 0 *and* 8*. Every bispecial factor* w *of* u *with* $|w| \geq 2$ *has the form* $w = s\,\varphi(w')\,0\,t$ *with* $s, t \in \{\varepsilon, 8\}$ *and* w' *a shorter bispecial factor of* u*, its* parent*. The extensions of* w *follow from* $\mathrm{Bext}(w')$ *by an explicit rule:* $(a,b) \in \mathrm{Bext}(w')$ *gives the pair (last letter of* $\varphi(a)$*, second letter of* $\varphi(b)$*) when* $s = t = \varepsilon$*, and similar rules for* $s = 8$ *or* $t = 8$*, which admit only* $a \in \{7, 8\}$ *or* $b \in \{7, 8\}$*.*

Both lemmas follow from the fact that each $\varphi(c)$ starts with 0 and has no other 0. So the bispecial factors of u form a tree with roots ε and 8. A node is a 2×2 *node* if it has exactly two left and two right extensions, and a *multi node* otherwise.

Lemma 5 (chains)**.** *A* 2×2 *node* w' *has exactly one bispecial child, and this child is again a* 2×2 *node. Each occurrence of* w' *gives an occurrence of the child, so the return words of the child have Parikh vectors* $M\vec{r}$*, for the return words* r *of* w'*.*

The program expands the tree from the roots. The expansion stops with 7 multi nodes, of lengths $0, 1, 3, 7, 15, 31, 63$, and 16 *seeds*: the root 8, which is a 2×2 node, and the 2×2 nodes whose parent is a multi node, of lengths $1, 2, 2, 4, 4, 8, 8, 16, 16, 32, 32, 64, 64, 127, 128, 128$. By Lemmas 4 and 5, every bispecial factor of u is a multi node or lies on the infinite chain $w_0, w_1, w_2, \dots$ of one seed w_0.

The images $\pi(i) = 0x_i$, with x_i a palindrome without 0, put a 0 at the start of each block. So a factor v of z without 0 lies inside some x_c, and a factor v with a 0 has a unique form $v = x\,\pi(w)\,0\,y$, with x a suffix of some x_a, y a prefix of some x_b and w a factor of u. If v is bispecial, then w is bispecial in u. For a 2×2 node w with Bext letters a, a' on the left and b, b' on the right, exactly one bispecial factor of z has this w:

$$v(w) = \mathrm{lcs}(\pi(a), \pi(a'))\,\pi(w)\,\mathrm{lcp}(\pi(b)0, \pi(b')0),$$

and the return words of $v(w)$ have lengths $h\cdot\vec{r}$ for the return words r of w. So the bispecial factors of z are: the empty word, 15 nonempty factors without 0, 51 factors from multi nodes, and the factors $v(w_n)$ on the 16 chains.

Lemma 6 (exact comparison)**.** *For integers $a, b \geq 1$: $a/b < q$ if and only if $P_5(3 - b/a) < 0$.*

Proof. $a/b < q$ is equivalent to $3 - b/a < \Lambda$. For $s \leq 2$ we have $P_5(s) = s^4(s-2) - 1 < 0$, and P_5 is increasing on $(8/5, \infty)$ with the single root Λ there. □

Return words of the seeds. The program builds the prefix $\varphi^{21}(0)$ of u, of length 2 310 194. The 17 two-letter factors of u all occur in $\varphi^9(0)$, so this prefix contains every factor of u of length at most 4153. For each seed w, let W be the length of the longest window of the prefix that contains no occurrence of w. If $W + 2 \leq 4153$, then every factor of u of length $W + 1$ contains w, so every return word r satisfies $|r| + |w| - 2 \leq W$, and rw occurs in the prefix. So the gaps between consecutive occurrences in the prefix give all return words. Each of the 16 seeds passes this test and has exactly 9 return words. This certificate replaces the cited fact that every factor of u_9 has 9 return words.

Chains. Along each chain, the type of the node (its Bext and the letters s, t) is periodic with period 8 from $n = 0$, for all 16 seeds. For a seed, a return word r, and $n \geq 0$, put $a_n = |v(w_n)|$ and $b_n = h \cdot M^n\vec{r}$. We need $a_n/b_n < q$ for all n.

Lemma 7 (tail)**.** *Write $n = 8N + i$ with $0 \leq i < 8$. Then $D_n := q\,b_n - a_n = h \cdot M^n X + c_i$, where X is a vector with entries in $\mathbb{Q}(\Lambda)$ and the c_i are rational numbers, all computed from the seed and r. Expanding in eigenvectors, $h \cdot M^n X = \sum_\mu \kappa_\mu \mu^n$. The coefficients κ_μ vanish for the four roots of Q_4, because h is orthogonal to their right eigenvectors. Hence*

$$D_n \;\geq\; \kappa_\Lambda \Lambda^n + c_i - \varepsilon\rho^n, \qquad \varepsilon = \sum_{\mu\neq\Lambda,\ P_5(\mu)=0} |\kappa_\mu|, \quad \rho \leq 0.8886.$$

The orthogonality $h \cdot v_\mu = 0$ for the roots of Q_4 is the length condition of [8, Proposition 17]. With $v_t = (t^7(t-2), \dots, t(t-2), t-2, 1)$ it becomes four linear equations: $h_5 = h_0 + h_4$, $h_6 = h_1 - h_4 + 2h_5$, $h_7 = h_2 - h_4 + 2h_6$, $h_8 = h_3 - h_4 + 2h_7$. The program checks it exactly, as a remainder modulo Q_4.

4 Proof of Theorem 2

The word z is uniformly recurrent, as a morphic image of the fixed point of a primitive morphism. It is aperiodic: the cut of z at its letters 0 recovers u, and the letter frequencies of u are irrational because P_5 is irreducible. The language of z is closed under reversal, because the reversal of $\pi(w)0$ is $\pi(\overline{w})0$ and the language of u is closed under reversal (T3).

Richness. The program checks the condition of (T2) on the empty word, on the 15 bispecial factors without 0, on the 51 factors from multi nodes, and on the chain factors $v(w_n)$ for $n \le 60$. Along a chain the type and the palindrome property are periodic from $n = 0$, so $n \le 60$ covers every case. By (T2), z is rich.

Upper bound. For the $51 + 15$ finite cases, the program finds the shortest return word in a prefix of z that provably contains it, and checks $|v|/|r| < q$ by Lemma 6. For the chains, it checks $a_n/b_n < q$ exactly for $n \le 60$. For $n > 60$ it uses Lemma 7. It decides exactly, in $\mathbb{Q}[t]/(P_5)$, whether $\kappa_\Lambda = 0$. In 142 of the 144 pairs (seed, return word), $\kappa_\Lambda > 0$, with smallest value 2.14, so $D_n \to \infty$. In the other 2 pairs, $\kappa_\Lambda = 0$ and $\min_i c_i = 2/5$, exactly. The largest ε in these two pairs is 0.5454, and $\varepsilon\rho^n < 2/5$ for $n \ge 3$. In all cases $D_n > 0$ for $n > 60$. So every ratio is below q, and $E(z) \le 1 + q$ by (T1).

Lower bound. For the pair on the chain of the seed $w_0 = 8$ with $\kappa_\Lambda = 0$, D_n stays bounded while $b_n \to \infty$. So $a_n/b_n \to q$ and $E(z) \ge 1 + q$ by (T1).

So $E(z) = 1 + q = E^*(u_9)$ by (T3). The word z is an infinite rich word over Σ_5 (it uses all five letters), so $\mathrm{RT}(C_5) \le E(z)$. □

The two critical chains are those of the seed $w_0 = 8$ and of the seed of length 127. The largest ratio that the program meets is $a_{60}/b_{60} \approx 1.0592854490$ on the chain of the seed of length 127, which is below q by Lemma 6.

5 How the result was checked

The folder that accompanies this paper contains four programs. Its `checks/README.md` gives the commands.

1. `checks/verify.py` does the proof of Section 4 for a given morphism. It uses exact rational and polynomial arithmetic (SymPy), and mpmath with 60 digits only for κ_μ ($\mu \ne \Lambda$), ε, ρ and the sign of a nonzero κ_Λ. It also compares the tree with the bispecial factors of u of length at most 40 found directly (36 factors). It ends with `RESULT: ALL CHECKS PASSED`.

2. `checks/crosscheck.py` uses only φ, π and prefixes of z. It finds all bispecial factors of z of length at most 60 by brute force (64 factors) and their shortest return words, and they equal the list that the theory predicts. It also confirms all 149 predicted bispecial factors of length at most 3000.

3. `checks/independent_check.py` was written in the verification stage from the definitions and the proof plan, before its author read `verify.py`. It finds by brute force the 75 bispecial factors of u of length at most 300 and the 146 bispecial factors of z of length at most 2000, checks the condition of (T2) on all of them, and checks $|v|/|r| < q$ for the 125 of them with a certified shortest return word (largest ratio 18/17). It checks the chains exactly for $n \le 1000$, with its own closed form and tail bound. At $n = 1000$ the term $\varepsilon\rho^n$ is below 10^{-50}.

4. `checks/prefix_check.c` uses only the definitions. On the prefix of z of length 9 727 357 (the image of 3 000 000 letters of u), an eertree finds 9 727 357 distinct nonempty palindromes,

so the prefix is rich. The largest exponent of a repetition in this prefix, over all periods, is $3684955/1789434 \approx 2.0592852265 < E^*(u_9)$.

Figure 1 shows the output of the last program: the largest exponent in each band of periods rises towards $E^*(u_9)$ from below.

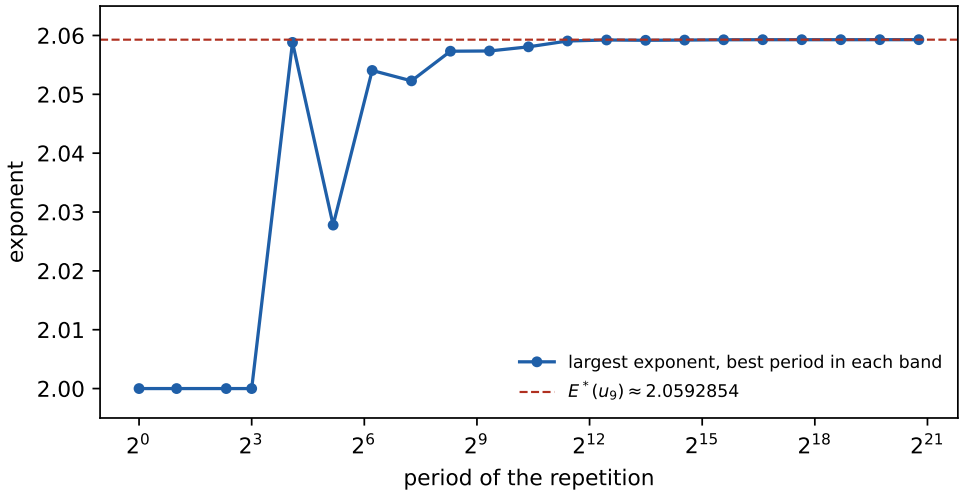


Figure 1: Largest exponent of a repetition in the first 9 727 357 letters of z, for periods in $[2^k, 2^{k+1})$, $k = 0, \ldots, 20$. The dashed line is $E^*(u_9)$.

Other projections. Two more morphisms pass the same programs. Candidate B differs from Table 1 only in $\pi(6) =$ `02343443432`. Candidate C is $0 \mapsto$ `01`, $1 \mapsto$ `011`, $2 \mapsto$ `0`, $3 \mapsto$ `02`, $4 \mapsto$ `022322`, $5 \mapsto$ `02234322`, $6 \mapsto$ `0223434434322`, $7 \mapsto$ `022343443433434434322`, $8 \mapsto$ `0223434434334344343234344343343443432 2`. We did not check whether the languages of the three words differ up to a permutation of the letters. So we do not claim three distinct words.

How π was found. A depth-first search over morphisms of class P_{ret} with marker 0 and palindromic x_i imposed the four linear equations of Section 3, and pruned by richness and by the largest exponent on prefixes of $\pi(u_9)$. Three candidates survived on a prefix of 1200 letters of u_9. The search is not part of the proof.

6 Limits

- The proof assumes (T1), (T2) and the two facts (T3) from the literature. We did not re-prove them, and no program checks them.
- The tail estimate of Lemma 7 uses 60-digit floating point for κ_μ ($\mu \neq \Lambda$), ε and ρ, not interval arithmetic. The decision $\kappa_\Lambda = 0$ is exact. The margins are large: every nonzero κ_Λ is at least 2.14; where $\kappa_\Lambda = 0$, $\min c_i = 2/5$ against $\varepsilon\rho^{60} < 2 \cdot 10^{-3}$, and the exact check of `independent_check.py` reaches $n = 1000$, where $\varepsilon\rho^n < 10^{-50}$.
- The programs are not formally verified, and there is no Lean proof.

- We do not prove $\mathrm{RT}(C_5) = E^*(u_9)$. The lower bound needs a structure theorem as in [6, Section 6].

References

[1] J. D. Currie, L. Mol and J. Peltomäki, *The repetition threshold for ternary rich words*, Electron. J. Combin. **32**(2) (2025), #P2.55, doi:10.37236/13499, arXiv:2409.12068.

[2] J. D. Currie, L. Mol and N. Rampersad, *The repetition threshold for binary rich words*, Discrete Math. Theor. Comput. Sci. **22**(1) (2020), article 6, doi:10.23638/DMTCS-22-1-6.

[3] F. Dolce, L. Dvořáková and E. Pelantová, *On balanced sequences and their critical exponent*, Theoret. Comput. Sci. **939** (2023), 18–47, doi:10.1016/j.tcs.2022.10.014.

[4] X. Droubay, J. Justin and G. Pirillo, *Episturmian words and some constructions of de Luca and Rauzy*, Theoret. Comput. Sci. **255** (2001), 539–553, doi:10.1016/S0304-3975(99)00320-5.

[5] L. Dvořáková, K. Klouda and E. Pelantová, *The asymptotic repetition threshold of sequences rich in palindromes*, European J. Combin. **126** (2025), 104124, doi:10.1016/j.ejc.2025.104124.

[6] L. Dvořáková, L. Mol and E. Pelantová, *An automated approach to structure theorems for minimally repetitive rich words*, arXiv:2611.13676v1, November 10, 2026.

[7] L. Dvořáková and E. Pelantová, *Symmetries of rich sequences with minimum critical exponent*, in Combinatorics on Words (WORDS 2025), Lecture Notes in Comput. Sci. **15729**, Springer, 2025, 143–154, doi:10.1007/978-3-031-97548-6_11, arXiv:2501.15135.

[8] L. Dvořáková and E. Pelantová, *The limit of repetition thresholds of rich sequences*, Electron. J. Combin. **33**(2) (2026), #P2.15, doi:10.37236/14600, arXiv:2509.06104.