On a counting conjecture for binary rich words

Max Glass

November 6, 2026

Abstract

Let $R_2(n)$ denote the number of binary rich words of length n. We prove that

$$\limsup_{n\to\infty} \frac{\log R_2(n)}{\sqrt{n}\log n} = \infty.$$

Consequently, $R_2(n)$ is not $O(n^{C\sqrt{n}})$ for any fixed $C > 0$, disproving the conjectured upper bound $O((n/g(n))^{\sqrt{n}})$ with $g(n) \to \infty$. The proof combines independent permutation families through an ordered concatenation construction and a binary gap encoding.

hexagon:2611.00167v1

1 The result

Let $R_2(n)$ count rich words of length n over $\{0, 1\}$. Guo, Shallit, and Shur [2, Conjecture 1] conjectured the bound

$$R_2(n) = O\left(\left(\frac{n}{g(n)}\right)^{\sqrt{n}}\right) \qquad \text{for some positive } g(n) \longrightarrow \infty. \tag{1}$$

This formulation also appears in [4, Section 1]. Using run-length constructions, Guo, Shallit, and Shur [3] obtained a lower bound exponential in $\sqrt{n}$.

Theorem 1. *The binary rich-word counting function satisfies*

$$\limsup_{n\to\infty} \frac{\log R_2(n)}{\sqrt{n}\log n} = \infty.$$

Corollary 2. *For every $C > 0$, $R_2(n) \neq O(n^{C\sqrt{n}})$. In particular,* (1) *is false.*

Proof. An eventual bound $R_2(n) \leq Kn^{C\sqrt{n}}$ would give a limsup at most C. Since $g(n) \geq 1$ eventually, (1) would imply such a bound with $C = 1$. □

The construction starts with $r!$ rich blocks of length $O(r^2)$. To multiply the choices, we place successive blocks on disjoint ordered pairs of letters and then encode the resulting word by binary gaps. The encoding charges $O(jr_j^2)$ for block j, suggesting parameters r_j proportional to $1/j$. We implement this choice dyadically: 2^t blocks of parameter $2s\,2^{k-t}$ at level t. Each level has length cost $O(s^2 4^k)$ and contributes at least $s2^k \log s$ to the logarithm of the count. Summing over $k + 1$ levels gives a lower bound of order $\sqrt{k+1}$ for the normalized logarithm, with $s \to \infty$ first. All logarithms are natural.

2 Ordered concatenation

We use the standard facts that a word w has at most $|w| + 1$ distinct palindromic factors and that every factor of a rich word is rich [1, Proposition 2.1 and Corollary 2.5]. Write $\mathcal{P}(w)$ for the set of palindromic factors of w, including ε, and $D(w) = |w| + 1 - |\mathcal{P}(w)|$ for its palindromic defect. For a set S of letters, let $|w|_S$ count the positions of w carrying letters of S, and put

$$I_S(w) = \{p \in \mathcal{P}(w) : apb \text{ is a factor of } w \text{ for some } a, b \in S\},$$
$$\sigma_S(w) = \max\{|w|_S - 1, 0\} - |I_S(w)|.$$

The two endpoint letters need not coincide. We abbreviate $I_{\{b\}}$ to I_b. If w is rich and b occurs in w, then

$$|I_b(w)| = |w|_b - 1. \tag{2}$$

Each position carrying b first introduces exactly one palindrome ending in b; these are b and the words bpb with $p \in I_b(w)$ [1, proof of Proposition 2.7].

If all letters are marked and $|w| \geq 2$, then $I_{\mathrm{alph}(w)}(w)$ is the set of palindromes in w with its first and last letters deleted. Thus $\sigma_{\mathrm{alph}(w)}(w) = 0$ for rich w, also when $|w| \leq 1$.

For a word $d = a_1 \cdots a_m$ of nonnegative integers, define

$$G(d) = 0\,1^{a_1}0\,1^{a_2}0 \cdots 1^{a_m}0, \qquad G(\varepsilon) = 0.$$

For nonempty d, with $M = \max_i a_i$ and $S_h = \{a : a \geq h\}$, one has

$$D(G(d)) = D(d) + \sum_{h=1}^{M} \sigma_{S_h}(d). \tag{3}$$

To see this, the nonempty zero-ended palindromes are exactly $G(p)$, $p \in \mathcal{P}(d)$. The one-ended palindromes are $1, \ldots, 1^M$ and $1^h G(p) 1^h$ with $p \in I_{S_h}(d)$. Parsing at zeros makes these classes disjoint and exhaustive. Hence

$$|\mathcal{P}(G(d))| = 1 + |\mathcal{P}(d)| + M + \sum_{h=1}^{M} |I_{S_h}(d)|.$$

Now use $|G(d)| = 1 + m + \sum_i a_i$ and $\sum_{h=1}^{M} |d|_{S_h} = \sum_i a_i$.

Lemma 3. *Let $q \geq 1$ and let $z_1, \ldots, z_q$ be nonempty binary words ending in 1, with $z_j 1$ rich for every j. Relabel $0, 1$ in z_j as $2j-2, 2j-1$, respectively, obtaining B_j. Then $G(B_1 \cdots B_q)$ is rich.*

Proof. Put $d = B_1 \cdots B_q$. Each block is rich, and their alphabets are pairwise disjoint. No nonempty palindrome crosses a block boundary, so d is rich. We show that every term in (3) vanishes.

At an even threshold $h = 2j-2$, the marked positions form the entire rich suffix $B_j \cdots B_q$, so the marked deficit is zero. At an odd threshold $h = 2j-1$, put $U = B_j$, $b = 2j-1$, and $V = B_{j+1} \cdots B_q$. Only b is marked in U, and all letters are marked in V. If V is empty, use (2). Otherwise, writing V^- for V with its last letter deleted,

$$I_{\{b\} \cup \mathrm{alph}(V)}(UV) = I_b(Ub) \cup \mathcal{P}(V^-). \tag{4}$$

For an interior contained in U, the first letter of V can replace the auxiliary final b in Ub. For an interior contained in V, the last letter of U supplies a left endpoint. No nonempty palindrome crosses the disjoint-alphabet boundary. These observations give (4). Its two sets intersect only in ε, which belongs to both since Ub ends in bb. As Ub and V^- are rich, the union has $|U|_b + |V| - 1$ elements. The marked deficit is therefore zero. Earlier blocks contain no marked positions and do not affect these interiors. Equation (3) completes the proof. □

3 A product lower bound

Lemma 4. *For each $r \geq 1$ there are $r!$ distinct binary words z ending in 1 such that $z1$ is rich and*

$$|z| = L_r := \frac{5r^2 + 9r + 4}{2}, \qquad |z|_1 = (r+1)^2. \tag{5}$$

Proof. For a permutation $p = (p_1, \ldots, p_r)$ of $1, \ldots, r$, put $H = r + 1$ and

$$W_p = 0\,1^H \prod_{i=1}^{r} (0^i 1^{p_i} 0^i 1^H)\, 0^r.$$

The following nonempty palindromic factors occur in W_p:

$$\begin{array}{lll} 0^t, \quad 0^t 1^H 0^t & 1 \le t \le r, & \\ 0^t 1^{p_i} 0^t & 1 \le i \le r, & 1 \le t \le i, \\ 0^t 1^H 0^i 1^{p_i} 0^i 1^H 0^t & 1 \le i \le r, & 1 \le t \le \max\{1, i-1\}, \\ 1^t & 1 \le t \le H, & \\ 1^t 0^i 1^t & 1 \le i \le r, & 1 \le t \le p_i, \\ 1^t 0^i 1^{p_i} 0^i 1^t & 1 \le i \le r, & 1 \le t \le H. \end{array}$$

In the third row, the zero-run before the left 1^H has length 1 when $i = 1$ and $i - 1$ otherwise; the zero-run on the right is at least as long. This gives the stated range of t. The run patterns, $p_i < H$, and the distinct values p_i distinguish all the listed words. Their total number is $L_r = |W_p|$, so W_p is rich. Its longest zero-run has length r; appending one zero adds the new palindrome 0^{r+1}. Thus $W_p 0$ is rich. The permutation is recoverable from the successive short one-runs, and $|W_p|_0 = (r+1)^2$. Taking z to be the bitwise complement of W_p gives the assertion. □

Proposition 5. *For $q \ge 1$ and positive integers $r_1, \dots, r_q$, define*

$$N = 1 + \sum_{j=1}^{q} \big((2j-1)L_{r_j} + (r_j+1)^2\big).$$

Then $R_2(N) \ge \prod_{j=1}^{q} r_j!$.

Proof. Choose the words in Lemma 4 independently and apply Lemma 3. A zero of block j contributes $2j-1$ letters to the gap encoding, and a one contributes $2j$, giving the stated common length. Parsing at zeros recovers the integer word; the value $\lfloor a/2 \rfloor$ identifies the block containing each integer a. Thus every permutation is recoverable, and all outputs are distinct. □

Proof of Theorem 1. Fix $k \ge 0$. For integers $s \ge 2$, take $q = 2^{k+1} - 1$ and

$$r_j = 2s\, 2^{k-t} \qquad (2^t \le j < 2^{t+1}, \quad 0 \le t \le k),$$

and denote the length in Proposition 5 by $N_{s,k}$. Since $(2u)! \ge s^u$ for $u \ge s$, that proposition gives

$$\log R_2(N_{s,k}) \ge (k+1)s2^k \log s. \tag{6}$$

Also $L_r \le 9r^2$ and $(r+1)^2 \le 4r^2$ for $r \ge 1$. The 2^t blocks at level t have total length contribution at most

$$2^t(9 \cdot 2^{t+2} + 4)(2s\, 2^{k-t})^2 \le 160 s^2 4^k.$$

The first block also gives $N_{s,k} \ge s$. Thus

$$s \le N_{s,k} \le 200(k+1)s^2 4^k + 1. \tag{7}$$

Letting $s \to \infty$ at fixed k in (6)–(7) gives

$$\liminf_{s\to\infty} \frac{\log R_2(N_{s,k})}{\sqrt{N_{s,k}} \log N_{s,k}} \ge \frac{\sqrt{k+1}}{2\sqrt{200}}.$$

Since $N_{s,k} \to \infty$ and k is arbitrary, the theorem follows. □

Formal verification. Theorem 1 and the negation of (1) have also been formalized in Lean 4 using Mathlib. The formal proof uses the product construction and explicit integer bounds for the scaled profiles above. The formalization and paper source are available at `https://github.com/glasss13/rich-word-counting`.

Acknowledgment. GPT-6 Astra (OpenAI) was used for the main proof construction and drafting of this paper.

References

[1] A. Glen, J. Justin, S. Widmer, and L. Q. Zamboni, Palindromic richness, *European Journal of Combinatorics* **30**(2) (2009), 510–531. doi:10.1016/j.ejc.2008.04.006.

[2] C. Guo, J. Shallit, and A. M. Shur, On the combinatorics of palindromes and antipalindromes, arXiv preprint arXiv:1503.09112v1 (2015).

[3] C. Guo, J. Shallit, and A. M. Shur, Palindromic rich words and run-length encodings, *Information Processing Letters* **116**(12) (2016), 735–738. doi:10.1016/j.ipl.2016.07.001.

[4] J. Rukavicka, Subexponential upper bound on the number of rich words, arXiv preprint arXiv:2511.11069v3 (2026).