A CLOSED FORM FOR THE REDUCED ABELIAN COMPLEXITY OF THE THUE–MORSE WORD, AND THE FIVE OPEN PROBLEMS OF CAMPBELL, CURRIE AND RAMPERSAD

KOREA SUPERINTELLIGENCE LABS

Abstract. Campbell, Currie and Rampersad recently introduced the *reduced* factor and abelian complexity functions of an infinite word: length-n factors are counted after every maximal block of equal letters has been collapsed to a single letter, up to equality (respectively, up to permutation) of the collapsed words. They evaluate both functions for the paperfolding word and the reduced factor complexity $\rho_{\mathbf{t}}^{\mathrm{red}}$ of the Thue–Morse word $\mathbf{t}$, and close with five printed questions about the reduced abelian complexity $\rho_{\mathbf{t}}^{\mathrm{ab,red}}$: an apparent halving identity, an apparent dichotomy $|\rho_{\mathbf{t}}^{\mathrm{ab,red}}(4n+2) - \rho_{\mathbf{t}}^{\mathrm{ab,red}}(4n)| \in \{0,1\}$ governed by $\mathbf{t}_{n+1} = \mathbf{t}_{3n+1}$, the sign of that difference, a recursion for $\rho_{\mathbf{t}}^{\mathrm{ab,red}}(4n)$, and non-$k$-automaticity.

We give a closed form. Writing m_n and M_n for the least and greatest number of alternations in a length-n factor of $\mathbf{t}$ and $\Lambda_n = M_n - m_n + 1 = \rho_{\mathbf{t}}^{\mathrm{red}}(n)/2$,

$$2\,\rho_{\mathbf{t}}^{\mathrm{ab,red}}(n) \;=\; 3\Lambda_n + \varepsilon_n, \qquad \varepsilon_n = \begin{cases} 0 & \Lambda_n \text{ even}, \\ (-1)^{m_n} & \Lambda_n \text{ odd}. \end{cases}$$

From it we prove the halving identity and the dichotomy; we evaluate the sign of the difference as $(-1)^{m_{n+1}}$, with m mod 2 given by an explicit $O(\log n)$ rule, equivalently by a four-state base-4 automaton, and we report two exhaustive searches that find no closed form of simpler shape; we give the recursion for $\rho_{\mathbf{t}}^{\mathrm{ab,red}}(4n)$ and hence a full recursion for $\rho_{\mathbf{t}}^{\mathrm{ab,red}}$ on the state $(\Lambda_n, m_n \bmod 2)$, which the value $\rho_{\mathbf{t}}^{\mathrm{ab,red}}(n)$ alone does not carry. The fifth question is settled as printed: the sequence is unbounded, hence has infinite range, hence is not k-automatic for any k. Outside the formal development we record that the strengthening a reader reaches for is false: $\rho_{\mathbf{t}}^{\mathrm{ab,red}}$ *is* 2-regular, its 2-kernel spanning an 11-dimensional space over $\mathbb{Q}$.

One caveat is uniform and is stated with every result below. All theorems are about the closed form $\widehat{\rho}(n) = (3\Lambda_n + \varepsilon_n)/2$ built from the (m_n, M_n) recursion of Campbell–Currie–Rampersad, and all are machine-checked in Lean 4 for that function. The identification $\widehat{\rho}(n) = \rho_{\mathbf{t}}^{\mathrm{ab,red}}(n)$ is checked by direct computation for $1 \le n \le 64$ and, in general, rests on that recursion together with the two structural facts established inside their proof of the recursion for $\rho_{\mathbf{t}}^{\mathrm{red}}$ — the run counts of the length-n factors fill the whole interval $[m_n + 1, M_n + 1]$, and $\mathbf{t}$ is closed under complement — which we cite and do not reprove.

1. Introduction

The complexity functions. Let $\mathbf{x}$ be an infinite word over a finite alphabet. Its *factor complexity* $\rho_{\mathbf{x}}(n)$ counts the distinct factors of length n, and its *abelian complexity* $\rho_{\mathbf{x}}^{\mathrm{ab}}(n)$ counts them up to rearrangement of letters. Campbell, Currie and Rampersad [3] introduce a reduction operation on finite words and, with it, reduced versions of both functions. We quote their definitions verbatim ([3], §1):

> For a nonempty and finite word w, let w be written as $w = c_1 \cdots c_1 c_2 \cdots c_2 \cdots c_n \cdots c_n$ for some positive integer n, where $c_1, c_2, \ldots, c_n$ are characters such that $c_i \neq c_{i+1}$ for all possible indices i. A maximal block of identical characters, $c_i \cdots c_i$ is called a *run*. We then define the *reduction* $\mathrm{red}(w)$ of w so that $\mathrm{red}(w) = c_1 c_2 \cdots c_n$.

Two finite words are *reduced-equivalent*, $w \sim_{\mathrm{red}} v$, when $\mathrm{red}(w) = \mathrm{red}(v)$; the *reduced factor complexity* $\rho_{\mathbf{x}}^{\mathrm{red}}(n)$ counts the length-n factors up to $\sim_{\mathrm{red}}$, and the *reduced abelian complexity* $\rho_{\mathbf{x}}^{\mathrm{ab,red}}$ is defined so that it

> maps $n \geq 0$ to the number of distinct factors of $\mathbf{x}$ of length n, where two factors v and w are considered to be equivalent if $\mathrm{red}(v)$ is of the same length as $\mathrm{red}(w)$ and $\mathrm{red}(v)$ can be obtained by rearranging the characters of $\mathrm{red}(w)$ (possibly by an identity permutation).

Following [3] we disregard throughout the trivial case $n = 0$. Over the binary alphabet $\mathrm{red}(w)$ alternates, so, as [3] observes, the $\sim_{\mathrm{red}}$ class of a binary word w is determined by the pair

$$\big(\text{first letter of } w,\ \text{number of runs of } w\big). \tag{1}$$

The indexing convention of [3] is fixed explicitly in its introduction and is load-bearing for one of the questions below. Verbatim: *"For an infinite sequence denoted as an infinite word, let the index of the initial term be* 1*. For* $n \in \mathbb{N}$*, set* $\mathbf{t}_n$ *to be equal to the number of* 1*'s, modulo* 2*, in the base-*2 *expansion of* $n - 1$*."* So the Thue–Morse word of [3] is 1-indexed, $\mathbf{t} = \mathbf{t}_1\mathbf{t}_2\mathbf{t}_3\cdots = 0110100110010110\cdots$. We keep that convention when quoting [3] and, when it is more convenient, also use the 0-indexed word

$$\mathfrak{t}(k) \;=\; s_2(k) \bmod 2, \qquad \text{so that} \qquad \mathbf{t}_k = \mathfrak{t}(k-1), \tag{2}$$

where s_2 is the binary digit sum.

What is known, and the five printed questions. The results of [3] are numbered consecutively across that paper. Its Theorem 1 evaluates the reduced factor complexity of $\mathbf{t}$:

$$\rho^{\mathrm{red}}_{\mathbf{t}}(n) = \begin{cases} \rho^{\mathrm{red}}_{\mathbf{t}}\big(\frac{n+1}{2}\big) & n \text{ odd}, \\ \rho^{\mathrm{red}}_{\mathbf{t}}(m+1) + 2 & n = 4m \text{ or } n = 4m+2, \end{cases} \tag{3}$$

for every positive integer n, whence $(\rho^{\mathrm{red}}_{\mathbf{t}}(n))_{n\geq 1}$ is 2-regular ([3], immediately after its Theorem 1; for k-regular sequences see [2], §16). Its Theorems 2 and 3 evaluate $\rho^{\mathrm{red}}_{\mathbf{f}}$ and $\rho^{\mathrm{ab,red}}_{\mathbf{f}}$ for the paperfolding word $\mathbf{f}$, both eventually periodic. For the Thue–Morse word the reduced *abelian* complexity is left open; [3] prints, as its equation (10),

$$\big(\rho^{\mathrm{ab,red}}_{\mathbf{t}}(n) : n \in \mathbb{N}\big) = (2,3,3,4,3,5,4,5,3,4,5,6,4,6,5,4,3,5,4,\ldots), \tag{4}$$

calls the problem of a recursion for it "much more challenging, relative to the above Theorems", and devotes its Conclusion to it. That Conclusion is one paragraph, which we quote in full ([3], §3; the display is its equation (11)):

> Although it appears that $\rho^{\mathrm{ab,red}}_{\mathbf{t}}(2n+1) = \rho^{\mathrm{ab,red}}_{\mathbf{t}}(n+1)$ for nonnegative integers n, the problem of determining a full recursion for $\rho^{\mathrm{ab,red}}_{\mathbf{t}}(n)$ seems to be challenging. It appears that
>
> $$\big|\rho^{\mathrm{ab,red}}_{\mathbf{t}}(4n+2) - \rho^{\mathrm{ab,red}}_{\mathbf{t}}(4n)\big| = \begin{cases} 0 & \text{if } \mathbf{t}_{n+1} = \mathbf{t}_{3n+1}, \\ 1 & \text{otherwise}, \end{cases} \tag{5}$$
>
> but it is unclear how the sign of $\rho^{\mathrm{ab,red}}_{\mathbf{t}}(4n+2) - \rho^{\mathrm{ab,red}}_{\mathbf{t}}(4n)$ could be evaluated in an explicit way for the nonzero case, and we leave this as an open problem. Also, it is unclear as to how a suitable recursion could be determined for $\rho^{\mathrm{ab,red}}_{\mathbf{t}}(4n)$, and we leave this as an open problem. We also leave it as an open problem to prove (5). It also appears that the integer sequence in (4) is not k-automatic, and we leave it as an open problem to prove this.

We label the five questions (C1)–(C5) in the order in which they are printed: (C1) the halving identity, together with the remark that a full recursion "seems to be challenging"; (C2) the dichotomy (5); (C3) the sign in the nonzero case; (C4) a recursion for $\rho^{\mathrm{ab,red}}_{\mathbf{t}}(4n)$; (C5) non-$k$-automaticity. The attributions differ and we record them exactly: (C3) and (C4) each carry "we leave this as an open problem" and (C2) and (C5) each carry "we leave it as an open problem", while (C1) is printed as an unproved observation — the sentence "We also leave it as an open problem to prove (5)" refers to the display, that is to (C2), not to the halving identity.

The display (5) is printed with no stated range for n, and we read it, and prove it, for $n \geq 1$: at $n = 0$ it would involve the value at length 0, which [3] disregards — verbatim ([3], §2.1), *"For convenience, we typically disregard the trivial case whereby a complexity function may or may not count the unique word of length* 0*, i.e., the empty or null word."*

Results. Let m_n and M_n denote the least and the greatest number of *alternations* (occurrences of 01 or 10) in a factor of $\mathbf{t}$ of length n, and set $\Lambda_n = M_n - m_n + 1$. The two structural facts that [3] establishes about these quantities inside the proof of (3) give $\rho_{\mathbf{t}}^{\mathrm{red}}(n) = 2\Lambda_n$. Our starting point is that they also *evaluate* the reduced abelian complexity, once one adds the parity observation that [3] makes for the paperfolding word but not for $\mathbf{t}$: two $\sim_{\mathrm{red}}$ classes with the same run count merge under abelian equivalence exactly when that run count is even. The result is Theorem 3.2:

$$\rho_{\mathbf{t}}^{\mathrm{ab,red}}(n) \;=\; \Lambda_n + \#\{\, r \in [m_n + 1, M_n + 1] : r \text{ odd} \,\}, \qquad \text{i.e.} \qquad 2\,\rho_{\mathbf{t}}^{\mathrm{ab,red}}(n) \;=\; 3\Lambda_n + \varepsilon_n \tag{6}$$

with $\varepsilon_n = 0$ when Λ_n is even and $\varepsilon_n = (-1)^{m_n}$ when Λ_n is odd. Since $\Lambda_n = \rho_{\mathbf{t}}^{\mathrm{red}}(n)/2$ obeys the recursion (3) of [3], everything then reduces to arithmetic on Λ_n and on the parity of m_n. Section by section:

- Theorem 4.1 is the arithmetic fact that drives the rest: Λ_{n+1} is odd if and only if $\mathfrak{t}(n) = \mathfrak{t}(3n)$, i.e. if and only if $\mathbf{t}_{n+1} = \mathbf{t}_{3n+1}$ in the notation of [3]. Equivalently, $\rho_{\mathbf{t}}^{\mathrm{red}}(n+1) \equiv 2 \pmod 4$ exactly on the criterion of (5) — so that criterion is a statement about the 2-regular sequence $\rho_{\mathbf{t}}^{\mathrm{red}}$ that [3] had already evaluated. Its proof reduces to the Thue–Morse identity $\mathfrak{t}(j) \oplus \mathfrak{t}(j+1) = \mathfrak{t}(3j+2) \oplus \mathfrak{t}(3j+3)$, which says that j and $3j+2$ have the same number of trailing 1s in binary.
- Theorem 5.1 is (C1).
- Theorem 6.1 evaluates the difference of (C2)–(C3) with its sign: it is 0 when $\mathbf{t}_{n+1} = \mathbf{t}_{3n+1}$ and $(-1)^{m_{n+1}}$ otherwise. Theorem 6.2 is (C2) as printed, and Theorem 6.3 makes the sign explicit: $m_{4p+j} \equiv m_{p+1} + [\, j \in \{0, 3\}\,] \pmod 2$, an $O(\log n)$ rule, equivalently a four-state automaton reading the base-4 expansion of n (Remark 6.4). Remark 6.5 reports two exhaustive searches which find no closed form of simpler shape.
- Theorem 7.1 is (C4), and with Theorem 5.1 it is a full recursion for $\rho_{\mathbf{t}}^{\mathrm{ab,red}}$ — carried on the pair $(\Lambda_n, m_n \bmod 2)$, which is necessary: the value $\rho_{\mathbf{t}}^{\mathrm{ab,red}}(n)$ alone is not a state (Remark 7.2).
- Proposition 8.1 settles (C5) as literally printed. It is a one-line consequence of (6) and (3) and we claim no more for it. Remark 8.2, outside the formal development, records that the natural strengthening is false: $\rho_{\mathbf{t}}^{\mathrm{ab,red}}$ is 2-regular.

The standing caveat. Everything below is proved unconditionally about an explicitly defined function $\widehat{\rho}$ (Definition 2.1) built from the recursion that [3] proves for (m_n, M_n). Reading the results as statements about $\rho_{\mathbf{t}}^{\mathrm{ab,red}}$ uses two results of [3] that we cite and do not reprove; we call them (B1) and (B2) and collect them in Section 2. The identification $\widehat{\rho}(n) = \rho_{\mathbf{t}}^{\mathrm{ab,red}}(n)$ is in addition verified by direct computation from the definition for every $1 \leq n \leq 64$ (Proposition 9.3). Every theorem head below says which of the two readings is meant. This is the only gap in the formal development, and closing it means formalising the Thue–Morse morphism and the parsing of a long factor, which we have not done.

2. Preliminaries

Throughout, $\mathbf{t}$ is the Thue–Morse word with the 1-indexing of [3] and $\mathfrak{t}$ is its 0-indexed form (2), so $\mathfrak{t}(2k) = \mathfrak{t}(k)$ and $\mathfrak{t}(2k+1) = 1 - \mathfrak{t}(k)$. We write $\oplus$ for addition modulo 2. A factor of $\mathbf{t}$ of length n with α alternations has $\alpha + 1$ runs; m_n and M_n are the least and greatest values of α over the length-n factors, and

$$\Lambda_n \;=\; M_n - m_n + 1.$$

The first values are $m_1 = M_1 = 0$, $m_2 = 0$, $M_2 = 1$, $m_3 = 1$, $M_3 = 2$.

The two results of [3] that we use without reproving them are the following.

(B1) ([3], Lemma 2.) For $n \geq 2$,

$$m_{2n} = 2n-1-M_{n+1}, \qquad M_{2n} = 2n-1-m_n, \qquad m_{2n+1} = 2n-M_{n+1}, \qquad M_{2n+1} = 2n-m_{n+1}.$$

(B2) ([3], inside the proof of its Theorem 1, verbatim.) *"Now let w be a factor of length n of $\mathbf{t}$ with i runs. Then $i \in [m_n+1, M_n+1]$. Furthermore, for every i in this interval, there is such a w in $\mathbf{t}$ (sliding w to the left or right in $\mathbf{t}$ can only increase or decrease the number of runs by at most 1). Since $\mathbf{t}$ is closed under complement, we see that $\rho_{\mathbf{t}}^{\mathrm{red}}(n) = 2(M_n+1-m_n-1+1) = 2(M_n-m_n+1)$."*

So (B2) says that the set of pairs (1) realised by the length-n factors of $\mathbf{t}$ is exactly $\{0,1\} \times [m_n+1, M_n+1]$, and in particular $\Lambda_n = \rho_{\mathbf{t}}^{\mathrm{red}}(n)/2$. Iterating (B1) gives the mod-4 form ([3], Lemma 3): for $n \geq 1$,

$$m_{4n} = 2n-1+m_{n+1}, \quad M_{4n} = 2n+M_{n+1}, \quad m_{4n+2} = 2n+m_{n+1}, \quad M_{4n+2} = 2n+1+M_{n+1}, \tag{7}$$

and, in the same way, $m_{4n+1} = 2n + m_{n+1}$ and $m_{4n+3} = 2n+1+m_{n+1}$. Two consequences used repeatedly: by (7),

$$\Lambda_{4n} = \Lambda_{4n+2} = \Lambda_{n+1} + 1 \qquad (n \geq 1), \tag{8}$$

so the two indices $4n$ and $4n+2$ carry the *same* Λ, while $m_{4n} = 2n-1+m_{n+1}$ and $m_{4n+2} = 2n+m_{n+1}$ have *opposite* parities. That single asymmetry is the whole content of (C2) and (C3). By (B1) — at $n = 1$, where (B1) does not apply, directly from the initial values —

$$\Lambda_{2n+1} = \Lambda_{n+1} \quad \text{and} \quad m_{2n+1} \equiv M_{n+1} \pmod 2 \qquad (n \geq 1). \tag{9}$$

Definition 2.1. Let $\widehat{m}_n$, $\widehat{M}_n$ be defined for $n \geq 1$ by the recursion of (B1) together with the initial values $\widehat{m}_1 = \widehat{M}_1 = 0$, $\widehat{m}_2 = 0$, $\widehat{M}_2 = 1$, $\widehat{m}_3 = 1$, $\widehat{M}_3 = 2$, and set $\widehat{\Lambda}_n = \widehat{M}_n - \widehat{m}_n + 1$,

$$\varepsilon_n = \begin{cases} 0 & \widehat{\Lambda}_n \text{ even}, \\ (-1)^{\widehat{m}_n} & \widehat{\Lambda}_n \text{ odd}, \end{cases} \qquad \widehat{\rho}(n) \;=\; \frac{3\widehat{\Lambda}_n + \varepsilon_n}{2}.$$

By (B1), together with the five initial values, which are read off $\mathbf{t}$, $\widehat{m}_n = m_n$ and $\widehat{M}_n = M_n$ for every $n \geq 1$, and we drop the hats whenever the reading through (B1) is meant. All formal statements are about $\widehat{m}$, $\widehat{M}$, $\widehat{\Lambda}$, $\widehat{\rho}$, which are defined by the recursion and are therefore unconditional; the agreement of $\widehat{m}, \widehat{M}$ with the true extremal alternation counts of $\mathbf{t}$ was also verified directly for $1 \leq n \leq 64$ (Proposition 9.2).

Note that $\widehat{\rho}$ is an integer: $3\widehat{\Lambda}_n + \varepsilon_n$ is even in both branches. Note also that $\widehat{\rho}(0) = 2$ whereas $\rho_{\mathbf{t}}^{\mathrm{ab,red}}(0) = 1$, the empty word being the unique factor of length 0; the identification below is asserted for $n \geq 1$ only, consistently with the convention of [3] that the trivial case is disregarded.

3. The evaluation

The counting step is elementary and is stated here in the form in which it is used. It combines (1) with the merge rule that [3] states in the first paragraph of the proof of its Theorem 3, verbatim: *"Suppose two binary words w and w' begin with different letters but have the same number of runs. If w and w' have an even number of runs then* $\mathrm{red}(w)$ *and* $\mathrm{red}(w')$ *are abelian equivalent; whereas, if w and w' have an odd number of runs then* $\mathrm{red}(w)$ *and* $\mathrm{red}(w)$ *are not abelian equivalent."* (Indeed a reduced binary word with r runs carries $\lceil r/2 \rceil$ copies of its first letter and $\lfloor r/2 \rfloor$ of the other, and these two multisets agree exactly when r is even.) In [3] the rule is applied only to the paperfolding word, case by case through six lemmas.

Lemma 3.1. *Let $a, L \geq 0$ be integers and let W be a family of nonempty binary words whose set of pairs (1) is exactly $\{0,1\} \times [a, a+L)$. Then the number of classes of W under reduced abelian equivalence is*

$$L + \#\{\, r \in [a, a+L) : r \text{ odd}\,\} \;=\; L + \Big\lfloor \frac{L + (a \bmod 2)}{2} \Big\rfloor,$$

and twice that number equals $3L + \delta$, where $\delta = 0$ if L is even, $\delta = 1$ if L is odd and a is odd, and $\delta = -1$ if L is odd and a is even.

Proof. By (1) the reduced abelian class of a word in W is determined by its run count r together with, when r is odd, its first letter; for even r the two words with run count r and different first letters are identified. So the classes are in bijection with $[a, a+L)$ together with one extra element for each odd r in that interval, giving the first display. The count of odd integers in $[a, a+L)$ is $\lfloor (L + (a \bmod 2))/2 \rfloor$ by induction on L, and the four parity cases of (a, L) give the second display. □

Theorem 3.2. *For every $n \geq 1$,*

$$\widehat{\rho}(n) \;=\; \Lambda_n + \#\{\, r \in [m_n + 1,\, M_n + 1] : r \text{ odd}\,\}, \qquad \textit{equivalently} \qquad 2\,\widehat{\rho}(n) = 3\Lambda_n + \varepsilon_n.$$

Granting (B1) and (B2), $\widehat{\rho}(n) = \rho_{\mathbf{t}}^{\mathrm{ab,red}}(n)$, so that

$$2\,\rho_{\mathbf{t}}^{\mathrm{ab,red}}(n) \;=\; 3\Lambda_n + \varepsilon_n, \qquad \varepsilon_n = \begin{cases} 0 & \Lambda_n \text{ even}, \\ (-1)^{m_n} & \Lambda_n \text{ odd}, \end{cases} \qquad \Lambda_n = \tfrac{1}{2}\rho_{\mathbf{t}}^{\mathrm{red}}(n).$$

Proof. The interval $[m_n + 1, M_n + 1]$ has Λ_n elements, so Lemma 3.1 with $a = m_n + 1$ and $L = \Lambda_n$ gives both displays after checking that its δ agrees with ε_n: δ is nonzero exactly when Λ_n is odd, and then $\delta = 1$ exactly when $a = m_n + 1$ is odd, that is when m_n is even. For the second sentence, (B2) says that the family of length-n factors of $\mathbf{t}$ satisfies the hypothesis of Lemma 3.1 with those a and L, and (B1) identifies $\widehat{m}$, $\widehat{M}$ with m, M. □

Both ingredients of Theorem 3.2 are in [3]: the merge rule quoted above, and the description (B2) of the run counts of $\mathbf{t}$. What is new is their combination, and what it yields: as the next four sections show, (6) settles four of the five printed questions.

4. The parity criterion

By (6), $\rho_{\mathbf{t}}^{\mathrm{ab,red}}$ is determined by Λ_n together with the parity of m_n, and the correction term ε_n is switched on and off by the parity of Λ_n. The following theorem identifies that parity with the criterion printed in (5), and is the reason (C2) is true.

Theorem 4.1. *For every $n \geq 0$: Λ_{n+1} is odd if and only if $\mathfrak{t}(n) = \mathfrak{t}(3n)$; equivalently, in the 1-indexed notation of* [3]*, if and only if $\mathbf{t}_{n+1} = \mathbf{t}_{3n+1}$. Granting (B1) and (B2), the same criterion reads*

$$\rho_{\mathbf{t}}^{\mathrm{red}}(n+1) \equiv 2 \pmod 4 \qquad \Longleftrightarrow \qquad \mathbf{t}_{n+1} = \mathbf{t}_{3n+1}.$$

Proof. Write $F(j) = \Lambda_{j+1} \bmod 2$. By (9) and (8),

$$F(2j) = F(j) \quad (j \geq 1), \qquad F(4j+1) = 1 \oplus F(j) \quad (j \geq 1), \qquad F(4j+3) = 1 \oplus F(j+1) \quad (j \geq 0),$$

with $F(0) = 1$ and $F(1) = 0$. Put $g(j) = \mathfrak{t}(j) \oplus \mathfrak{t}(3j)$. From $\mathfrak{t}(2k) = \mathfrak{t}(k)$ and $\mathfrak{t}(2k+1) = 1 \oplus \mathfrak{t}(k)$,

$$g(2j) = g(j), \qquad g(4j+1) = 1 \oplus g(j), \qquad g(4j+3) = 1 \oplus \big(\mathfrak{t}(j) \oplus \mathfrak{t}(3j+2)\big).$$

Strong induction on j now gives $F(j) = 1 \oplus g(j)$ for all j, the case $4j+3$ requiring

$$\mathfrak{t}(j) \oplus \mathfrak{t}(j+1) \;=\; \mathfrak{t}(3j+2) \oplus \mathfrak{t}(3j+3) \qquad (j \geq 0), \tag{10}$$

which converts $\mathfrak{t}(j) \oplus \mathfrak{t}(3j+2)$ into $\mathfrak{t}(j+1) \oplus \mathfrak{t}(3j+3) = g(j+1)$. For (10), note that $\mathfrak{t}(j) \oplus \mathfrak{t}(j+1) \equiv 1 + \nu(j) \pmod 2$ where $\nu(j)$ is the number of trailing 1s in the binary expansion of j, since passing

from j to $j+1$ clears those $\nu(j)$ ones and sets one zero. It therefore suffices that $\nu(3j+2)=\nu(j)$, and indeed if $j = a\cdot 2^{k+1}+(2^k-1)$ with $k=\nu(j)$ then

$$3j+2 \;=\; 3a\cdot 2^{k+1}+2^{k+1}+(2^k-1),$$

whose bit k is 0 and whose bits $0,\dots,k-1$ are 1, so $\nu(3j+2)=k$ as well. Finally $F(n)=1$ says Λ_{n+1} is odd, and $g(n)=0$ says $\mathfrak{t}(n)=\mathfrak{t}(3n)$, which by (2) is $\mathbf{t}_{n+1}=\mathbf{t}_{3n+1}$. Granting (B1) and (B2), $\Lambda_{n+1}=\rho_{\mathbf{t}}^{\mathrm{red}}(n+1)/2$, so Λ_{n+1} odd is $\rho_{\mathbf{t}}^{\mathrm{red}}(n+1)\equiv 2 \pmod 4$. □

The criterion of (5) is thus a statement about $\rho_{\mathbf{t}}^{\mathrm{red}}$ alone — the sequence whose recursion (3) is the main theorem of [3], and which [3] notes is 2-regular. The identity (10) was verified independently for every $j<200{,}000$ before it was proved.

5. (C1): the halving identity

Theorem 5.1. *For every integer $n\ge 0$, $\widehat{\rho}(2n+1)=\widehat{\rho}(n+1)$. Granting (B1) and (B2), this is (C1):*

$$\rho_{\mathbf{t}}^{\mathrm{ab,red}}(2n+1)=\rho_{\mathbf{t}}^{\mathrm{ab,red}}(n+1)\qquad(n\ge 0).$$

Proof. The case $n=0$ is trivial, so let $n\ge 1$. By (9), $\Lambda_{2n+1}=\Lambda_{n+1}$, so by Theorem 3.2 it suffices that $\varepsilon_{2n+1}=\varepsilon_{n+1}$. If Λ_{n+1} is even both terms vanish. If $\Lambda_{n+1}=M_{n+1}-m_{n+1}+1$ is odd then $M_{n+1}\equiv m_{n+1} \pmod 2$, and (9) gives $m_{2n+1}\equiv M_{n+1}\equiv m_{n+1}$, so the two correction terms are equal. □

The proof shows why the identity is not a formal consequence of $\Lambda_{2n+1}=\Lambda_{n+1}$ alone: the first-letter bookkeeping carried by ε survives the halving precisely because an odd Λ forces M and m to have the same parity.

6. (C2) and (C3): the dichotomy and its sign

Theorem 6.1. *For every $n\ge 1$,*

$$\widehat{\rho}(4n+2)-\widehat{\rho}(4n)=\begin{cases}0 & \text{if } \mathfrak{t}(n)=\mathfrak{t}(3n),\\ (-1)^{m_{n+1}} & \text{otherwise.}\end{cases}$$

Granting (B1) and (B2), this answers (C3): with the criterion of (5) written in the notation of [3],

$$\rho_{\mathbf{t}}^{\mathrm{ab,red}}(4n+2)-\rho_{\mathbf{t}}^{\mathrm{ab,red}}(4n)=\begin{cases}0 & \text{if } \mathbf{t}_{n+1}=\mathbf{t}_{3n+1},\\ (-1)^{m_{n+1}} & \text{otherwise,}\end{cases}\qquad(n\ge 1),$$

where m_{n+1} is the least number of alternations in a length-$(n+1)$ factor of $\mathbf{t}$.

Proof. By (8) the two indices carry the same Λ, namely $\Lambda_{n+1}+1$, so by Theorem 3.2 the difference is $(\varepsilon_{4n+2}-\varepsilon_{4n})/2$. If Λ_{n+1} is odd, both Λ_{4n} and Λ_{4n+2} are even and both correction terms vanish, giving 0; by Theorem 4.1 this is exactly the case $\mathfrak{t}(n)=\mathfrak{t}(3n)$. If Λ_{n+1} is even, both are odd, and by (7) $m_{4n}=2n-1+m_{n+1}$ and $m_{4n+2}=2n+m_{n+1}$ have opposite parities, so $\varepsilon_{4n+2}=-\varepsilon_{4n}=(-1)^{m_{n+1}}$ and the difference is $(-1)^{m_{n+1}}$. □

Theorem 6.2. *For every $n\ge 1$,*

$$\left|\widehat{\rho}(4n+2)-\widehat{\rho}(4n)\right|=\begin{cases}0 & \text{if } \mathbf{t}_{n+1}=\mathbf{t}_{3n+1},\\ 1 & \text{otherwise.}\end{cases}$$

Granting (B1) and (B2), this is (C2), the displayed conjecture (5) of [3], in its own 1-indexed notation.

Proof. Immediate from Theorem 6.1, since $|(-1)^{m_{n+1}}|=1$, and from $\mathbf{t}_{n+1}=\mathfrak{t}(n)$, $\mathbf{t}_{3n+1}=\mathfrak{t}(3n)$. □

The 1-indexing is load-bearing: read 0-indexed, the criterion of (5) already fails at $n = 1$ (Proposition 9.1(i)).

It remains to make $(-1)^{m_{n+1}}$ explicit, which is what (C3) asks for.

Theorem 6.3. *For every $p \geq 1$ and every $j \in \{0, 1, 2, 3\}$,*

$$m_{4p+j} \;\equiv\; m_{p+1} + [\, j \in \{0, 3\}\,] \pmod 2.$$

Granting (B1), this computes $m_n \bmod 2$, and hence the sign in Theorem 6.1, in $O(\log n)$ arithmetic operations.

Proof. Immediate from (7) and the two companion identities $m_{4p+1} = 2p + m_{p+1}$, $m_{4p+3} = 2p + 1 + m_{p+1}$ recorded in Section 2: the four right-hand sides are $2p - 1 + m_{p+1}$, $2p + m_{p+1}$, $2p + m_{p+1}$, $2p + 1 + m_{p+1}$. □

Remark 6.4. Theorem 6.3 is a four-state automaton reading base-4 digits. Put $\mu(p) = m_{p+1} \bmod 2$ and track the pair $(\mu(p), \mu(p+1))$. The theorem gives, for $p \geq 1$ and $(a, b) = (\mu(p), \mu(p+1))$,

$$\begin{aligned} (\mu(4p), \mu(4p+1)) &= (a, a), & (\mu(4p+1), \mu(4p+2)) &= (a, \bar{a}), \\ (\mu(4p+2), \mu(4p+3)) &= (\bar{a}, \bar{b}), & (\mu(4p+3), \mu(4p+4)) &= (\bar{b}, b), \end{aligned}$$

where $\bar{\cdot}$ is complementation. Reading the base-4 expansion of p most significant digit first from the initial state $(\mu(0), \mu(1)) = (0, 0)$ and outputting the first coordinate therefore computes $\mu(p)$; the four transitions out of the initial state hold as well, since $m_1 = m_2 = 0$, $m_3 = m_4 = 1$ and $m_5 = 2$. The automaton was checked against a direct computation of μ for every $p < 131{,}070$. Its transitions commute with complementing both coordinates, so the two-state quotient recording whether $\mu(p) = \mu(p+1)$ is itself automatic: from the "equal" state the digits 0 and 2 stay and 1 and 3 leave, and from the "differ" state only the digit 0 returns.

Remark 6.5. Two exhaustive searches, run outside the formal development with an independent program, say that no closed form of simpler shape is being missed within their ranges. First, no function of the form $\sigma(n) = \mathfrak{t}(a_1 n + b_1) - \mathfrak{t}(a_2 n + b_2)$ with $1 \leq a_i \leq 12$ and $0 \leq b_i \leq 12$ agrees with the sign sequence of Theorem 6.1: all 24,180 ordered pairs of parameters were tested against 30,000 terms and all were rejected. Second, no exclusive-or of one or two terms $\mathfrak{t}(am + b)$ with $1 \leq a \leq 8$, $0 \leq b \leq 8$, with or without an added constant, equals $m_{n+1} \bmod 2$: $144 + 5{,}112$ candidates, all rejected against 20,000 terms. The sign sequence itself begins $1, -1, 0, 1, 0, 0, 0, -1, 0, 0, 1, 0, -1, \ldots$ and was not found in the OEIS [7] (searched September 2026).

7. (C4): the recursion

Theorem 7.1. *For every $n \geq 1$,*

$$2\,\widehat{\rho}(4n) = 3\Lambda_{n+1} + 3 + \begin{cases} 0 & \Lambda_{n+1} \text{ odd}, \\ -(-1)^{m_{n+1}} & \Lambda_{n+1} \text{ even}, \end{cases}$$

$$2\,\widehat{\rho}(4n+2) = 3\Lambda_{n+1} + 3 + \begin{cases} 0 & \Lambda_{n+1} \text{ odd}, \\ (-1)^{m_{n+1}} & \Lambda_{n+1} \text{ even}. \end{cases}$$

Granting (B1) and (B2), the first display answers (C4), and together with Theorem 5.1 the two displays are a full recursion for $\rho_{\mathbf{t}}^{\mathrm{ab,red}}$, carried on the state $(\Lambda_n, m_n \bmod 2)$: Λ obeys (3), $m \bmod 2$ obeys Theorem 6.3, and $\rho_{\mathbf{t}}^{\mathrm{ab,red}}$ is the function $(3\Lambda_n + \varepsilon_n)/2$ of the pair.

Proof. By (8), $\Lambda_{4n} = \Lambda_{4n+2} = \Lambda_{n+1} + 1$, so $3\Lambda_{4n} = 3\Lambda_{n+1} + 3$ and it remains to evaluate the two correction terms. If Λ_{n+1} is odd then Λ_{4n} and Λ_{4n+2} are even and both corrections vanish. If Λ_{n+1} is even then both are odd, and by (7) $m_{4n} \equiv m_{n+1} + 1$ and $m_{4n+2} \equiv m_{n+1} \pmod 2$, giving $\varepsilon_{4n} = -(-1)^{m_{n+1}}$ and $\varepsilon_{4n+2} = (-1)^{m_{n+1}}$. Now apply Theorem 3.2. □

Remark 7.2. The pair is needed; the value $\rho_{\mathbf{t}}^{\mathrm{ab,red}}(n)$ alone is not a state for this recursion. From the printed values (4), $\rho_{\mathbf{t}}^{\mathrm{ab,red}}(2) = \rho_{\mathbf{t}}^{\mathrm{ab,red}}(3) = 3$, while the recursion of Theorem 7.1 sends the two indices to $\rho_{\mathbf{t}}^{\mathrm{ab,red}}(4) = 4$ and $\rho_{\mathbf{t}}^{\mathrm{ab,red}}(8) = 5$. The two inputs differ only in the parity of m: $\Lambda_2 = \Lambda_3 = 2$ but $m_2 = 0$ and $m_3 = 1$. This is presumably why [3] finds a full recursion "challenging": it is challenging for $\rho_{\mathbf{t}}^{\mathrm{ab,red}}$ alone.

8. (C5): unboundedness, and what it does not give

Proposition 8.1. *The sequence $(\widehat{\rho}(n))_{n\geq 1}$ is unbounded, hence has infinite range. Granting (B1) and (B2), $(\rho_{\mathbf{t}}^{\mathrm{ab,red}}(n) : n \in \mathbb{N})$ takes infinitely many values and is therefore not k-automatic for any k, which is (C5) as literally printed.*

Proof. By (8), $\Lambda_{4m} = \Lambda_{m+1} + 1$ for $m \geq 1$, so the indices $n_0 = 2$ and $n_{i+1} = 4(n_i - 1)$, that is $2, 4, 12, 44, 172, \ldots$, satisfy $\Lambda_{n_i} = i + 2$. Since $|\varepsilon_n| \leq 1$, Theorem 3.2 gives $2\widehat{\rho}(n_i) \geq 3(i + 2) - 1$, which is unbounded. A sequence with unbounded values has infinite range, whereas a k-automatic sequence is by definition the output of a deterministic finite automaton with output, whose output alphabet is finite, and so has finite range ([2], Ch. 5, where k-automatic sequences are defined). □

We claim no more for Proposition 8.1 than that it settles the printed question: once (6) is available it is one line from the recursion (3) of [3]. The interest of (C5) lies rather in what it suggests and what is in fact true.

Remark 8.2. The natural strengthening of (C5) — that $\rho_{\mathbf{t}}^{\mathrm{ab,red}}$ is not *k-regular*, which is the form in which the two positive results of [3] are stated — is **false**. The following argument is complete but is *not* part of the formal development below: it grants (B1) and (B2), as does every reading of $\widehat{\rho}$ as $\rho_{\mathbf{t}}^{\mathrm{ab,red}}$ above, and k-regular sequences [2] are in any case not available in the library we use. The numerical rank reported at the end is exact arithmetic, not a proof.

(Λ_n) is 2-regular, being $\rho_{\mathbf{t}}^{\mathrm{red}}/2$ ([3], after its Theorem 1). Next, $n \mapsto \Lambda_n \bmod 2$ is 2-automatic: by Theorem 4.1 it is $1 \oplus g(n)$ with $g(m) = \mathfrak{t}(m) \oplus \mathfrak{t}(3m)$, and writing $h(m) = \mathfrak{t}(m) \oplus \mathfrak{t}(3m + 1)$ and $q(m) = \mathfrak{t}(m) \oplus \mathfrak{t}(3m + 2)$, the relations $\mathfrak{t}(2k) = \mathfrak{t}(k)$, $\mathfrak{t}(2k + 1) = 1 \oplus \mathfrak{t}(k)$ give

$$\begin{aligned} g(2m) &= g(m), & h(2m) &= 1 \oplus g(m), & q(2m) &= h(m), \\ g(2m + 1) &= h(m), & h(2m + 1) &= 1 \oplus q(m), & q(2m + 1) &= q(m), \end{aligned}$$

a six-state system read least significant digit first, with $g(0) = 0$ and $h(0) = q(0) = 1$. Third, $n \mapsto m_n \bmod 2$ is 2-automatic by Remark 6.4, base-4 digits being pairs of base-2 digits. Hence ε, a function of those two parities, is 2-automatic, so its 2-kernel is a finite set; and every member of the 2-kernel of $2\rho_{\mathbf{t}}^{\mathrm{ab,red}} = 3\Lambda + \varepsilon$ lies in the span of the 2-kernel of Λ plus the span of the 2-kernel of ε, which is finite-dimensional. So $\rho_{\mathbf{t}}^{\mathrm{ab,red}}$ is 2-regular. The dimension of the $\mathbb{Q}$-span of its 2-kernel was measured to be 11 (and 7 for Λ): by exact Gaussian elimination over $\mathbb{Q}$ on vectors of length 400 with 127 kernel rows, and over $\mathbb{F}_p$ for $p = 2^{31} - 1$ and $p = 10^9 + 7$ on vectors of length 1500 with 511 kernel rows, stable from the fourth level on. So the printed cell (C5) has no hard reading: literally it is true and immediate, and strengthened to k-regularity it is false.

9. Controls, and the computation from the definition

A closed form checked only where it was designed to fit is worth little, and an off-by-one in an indexing convention would be invisible in most of the statements above. The following were checked; each refutes one specific way in which the development could have been satisfied vacuously or by a misreading.

Proposition 9.1. (i) The convention is load-bearing. *Reading the criterion of* (5) *with the* 0*-indexed word makes it* false *at $n = 1$: $\mathfrak{t}(2) = \mathfrak{t}(4)$, which would predict a difference of* 0*, whereas $|\widehat{\rho}(6) - \widehat{\rho}(4)| = 1$. The same computation performed directly from the definition of $\rho_{\mathbf{t}}^{\mathrm{ab,red}}$, rather than from $\widehat{\rho}$, gives the same refutation.*

(ii) The hypothesis is not vacuous. *With the* 1*-indexing of* [3] *the criterion at the same place reads* $\mathbf{t}_2 \neq \mathbf{t}_4$*, predicting* 1*, which is correct.*
(iii) Not an identity. *It is false that* $\widehat{\rho}(4n+2) = \widehat{\rho}(4n)$ *for all* $n \geq 1$*; the case* $n = 1$ *fails.*
(iv) Not always nonzero. $\widehat{\rho}(14) = \widehat{\rho}(12)$*, so the dichotomy of Theorem 6.2 genuinely has two cases.*
(v) The windows used below are not decoration. *Counting the reduced abelian classes of the length-*23 *factors over only the first* 40 *starting positions of* $\mathbf{t}$ *gives a strictly smaller number than counting over the first* 900*.*

The last item is about the following computation, which is how the closed form was tested against the actual word. For a window bound P, let $\rho_{\mathbf{t}}^{\mathrm{ab,red}}(n \mid P)$ be the number of reduced abelian classes among the length-n factors occurring at the positions $i < P$ of $\mathbf{t}$. This is unconditionally a lower bound for $\rho_{\mathbf{t}}^{\mathrm{ab,red}}(n)$, and equals it as soon as the positions $i < P$ carry every length-n factor of $\mathbf{t}$. Rather than appeal to a bound on the recurrence function of $\mathbf{t}$ we checked that directly: for every $n \leq 64$ the positions $i < 200$ already realise $\rho_{\mathbf{t}}(n)$ distinct factors of length n, the whole ordinary factor complexity of $\mathbf{t}$ — classical, and the sequence [3] prints in its §2.1 and identifies with OEIS `A005942` [7]. So nothing is missed at $P = 200$, nor at the larger windows $P = 400$, 480, 900, 1400 used below. That count was made outside the formal development; inside it, Proposition 9.2(ii) exhibits the saturation of the reduced abelian count itself.

Proposition 9.2. (i) *For* $1 \leq n \leq 19$*,* $\rho_{\mathbf{t}}^{\mathrm{ab,red}}(n \mid 400)$ *equals the* n*-th term printed in* (4)*; for* $1 \leq n \leq 23$*, the corresponding count* $\rho_{\mathbf{t}}^{\mathrm{red}}(n \mid 480)$ *for* $\sim_{\mathrm{red}}$ *equals the* n*-th term of the sequence* $(2, 4, 4, 6, 4, 6, 6, 6, 4, 6, 6, 8, 6, 8, 6, 6, 4, 6, 6, 8, 6, 8, 8, \ldots)$ *printed for* $\rho_{\mathbf{t}}^{\mathrm{red}}$ *in* [3]*, §2.1.*
(ii) *For* $1 \leq n \leq 40$*,* $\rho_{\mathbf{t}}^{\mathrm{ab,red}}(n \mid 200) = \rho_{\mathbf{t}}^{\mathrm{ab,red}}(n \mid 900)$*: the count has saturated well before the windows used.*
(iii) *For* $1 \leq n \leq 64$*, the recursively defined* $\widehat{m}_n$ *and* $\widehat{M}_n$ *of Definition 2.1 are the true minimum and maximum numbers of alternations over the length-*n *factors occurring at the positions* $i < 1400$*.*

Proposition 9.3. *For every* $1 \leq n \leq 64$*,* $\rho_{\mathbf{t}}^{\mathrm{ab,red}}(n \mid 1400) = \widehat{\rho}(n)$*. Since every factor of* $\mathbf{t}$ *of length* $n \leq 64$ *occurs at a position* $i < 200$*, hence at one* $i < 1400$*, the identification* $\widehat{\rho}(n) = \rho_{\mathbf{t}}^{\mathrm{ab,red}}(n)$ *used in the theorem heads above therefore holds for* $1 \leq n \leq 64$ *without appeal to (B1) or (B2).*

Beyond $n = 64$ the identification is the one described in Section 1: it rests on (B1) and (B2), which are results of [3]. Making it unconditional means formalising the Thue–Morse morphism $\mu : 0 \mapsto 01$, $1 \mapsto 10$, the four-way parsing of a long factor as $\mu(u)$, $^{-}\mu(u)^{-}$, $\mu(u)^{-}$ or $^{-}\mu(u)$ used in the proof of (B1), and the alternation count of $\mu(w)$; we have not done this. In addition to the checks above, the statements of Theorems 4.1, 5.1, 6.1 and 6.2 were tested numerically against a direct computation of $\rho_{\mathbf{t}}^{\mathrm{ab,red}}$ from the definition — for every $n < 150{,}000$ (halving), $n < 200{,}000$ (parity criterion) and $n < 100{,}000$ (dichotomy and sign) — with no violation, before any of them was proved.

10. Formal verification

Every lemma, theorem and proposition stated above is formalised and machine-checked in Lean 4 [6] against the Mathlib library; the exceptions, each labelled as such in the text, are Remark 8.2 (the 2-regularity of $\rho_{\mathbf{t}}^{\mathrm{ab,red}}$ and the measured rank 11), the two exhaustive searches of Remark 6.5, the parts of Remark 6.4 that go beyond the recursion it restates (the transitions out of the initial state and the check for $p < 131{,}070$), the values quoted in Remark 7.2, which are terms printed in [3] and are reproduced by Proposition 9.2(i), the numerical tests recorded at the end of Section 9, and the two results (B1), (B2) of [3], which are cited and not reproved. The development is three files — the definitions; the theory, which is Sections 3–8 together with the controls of Proposition 9.1(i)–(iv); and the computations from the definition, which are Propositions 9.2, 9.3 and 9.1(v). There is no incomplete proof and no added axiom. The axiom set of every headline statement was printed.

The fifty theorems of the theory file, which include all of Theorems 3.2, 4.1, 5.1, 6.1, 6.2, 6.3, 7.1 and Proposition 8.1, depend only on propositional extensionality, quotient soundness and choice — no compiled evaluation — which is the formal counterpart of their being proofs for all n rather than checks over a range. The eight statements of the computational file, namely Propositions 9.2 and 9.3, the item (v) and the from-the-definition half of item (i) of Proposition 9.1, and a check of Theorems 6.1 and 6.2 straight from the definition of $\rho_{\mathbf{t}}^{\mathrm{ab,red}}$ for $1 \le n \le 15$, each additionally depend on exactly one compiled-evaluation axiom, one per named computation. In those computations the Thue–Morse word, the run counts and the two class maps are built from the definitions quoted in Section 1, independently of the closed form, so that Proposition 9.3 compares two separately implemented objects. A full check of the development takes well under one CPU-hour, with a peak memory use of about seven gigabytes in the compiled evaluations. The repository is private; repository reference to be added at submission.

Acknowledgement of the source

Every question answered here is printed in [3], and both ingredients of the evaluation (6) — the merge rule for reduced binary words of equal run count, and the description (B2) of the run counts of $\mathbf{t}$ — are that paper's. What is added is their combination, the arithmetic criterion of Theorem 4.1, the parity recursion of Theorem 6.3, and the consequences drawn in Sections 5–8.

References

[1] G. Allouche, J.-P. Allouche and J. Shallit, *Kolam indiens, dessins sur le sable aux îles Vanuatu, courbe de Sierpinski et morphismes de monoïde*, Annales de l'Institut Fourier **56** (2006), 2115–2130; DOI 10.5802/aif.2235 (publisher record and Crossref, 3 September 2026). Cited by [3] for the non-automaticity of the run-length sequence of $\mathbf{t}$, the fixed point of $1 \to 121$, $2 \to 12221$; it is the nearest published relative of (C5). Not used in any proof above.

[2] J.-P. Allouche and J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge University Press, 2003. Used for two standard facts that are not reproved here: a k-automatic sequence has finite range (its Ch. 5, where such sequences are defined as outputs of finite automata), and the theory of k-regular sequences (its Ch. 16, the chapter [3] cites for the same purpose).

[3] J. M. Campbell, J. Currie and N. Rampersad, *Reduced complexities for sequences over finite alphabets*, arXiv: 2509.16034v1 (submitted 19 September 2025), 18 pages, math.CO with a cs.FL cross-list, MSC 11B85. All quotations above are from the arXiv source of v1, retrieved on 3 September 2026; the abstract page was checked the same day, and v1 is the only version and carries no journal reference or publisher DOI. The theorem environments of [3] are numbered consecutively across the paper, without a section counter, so that its Theorem 1 is (3), its Lemma 2 is (B1), its Lemma 3 is (7), and its Theorem 3 evaluates $\rho_{\mathbf{f}}^{\mathrm{ab,red}}$; the number of every result and equation cited above was read off a compilation of the v1 source, and is pinned to v1, the only version.

[4] J. Chen and Z.-X. Wen, *On the abelian complexity of generalized Thue–Morse sequences*, Theoretical Computer Science **780** (2019), 66–73; DOI 10.1016/j.tcs.2019.02.014 (Crossref, 3 September 2026). Cited by [3] as motivation; a different object from the one studied here, since no reduction is applied to the factors.

[5] F. Greinecker, *On the 2-abelian complexity of the Thue–Morse word*, Theoretical Computer Science **593** (2015), 88–105; DOI 10.1016/j.tcs.2015.05.047; arXiv:1404.3906 (Crossref, 3 September 2026; the arXiv record carries no journal reference). Cited by [3] as motivation; again a different object, and its main result — that the 2-abelian complexity of $\mathbf{t}$ is 2-regular — is the analogue for that object of Remark 8.2 here.

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

[7] OEIS Foundation Inc., *The On-Line Encyclopedia of Integer Sequences*, https://oeis.org. The factor complexity $\rho_{\mathbf{t}}$ is entry A005942, as [3] notes; neither $\rho_{\mathbf{t}}^{\mathrm{red}}$ nor $\rho_{\mathbf{t}}^{\mathrm{ab,red}}$ nor the sequences (m_n), (M_n), (Λ_n) was found there (searched 3 September 2026).

[8] G. Richomme, K. Saari and L. Q. Zamboni, *Abelian complexity of minimal subshifts*, Journal of the London Mathematical Society (2) **83** (2011), 79–95; DOI 10.1112/jlms/jdq063; arXiv:0911.2914 (volume, issue, pages and year from Crossref and zbMATH 1211.68300, 3 September 2026; the arXiv journal-reference field gives only the online year, and the preprint's title reads *…in Minimal Subshifts*). This is the reference for the ordinary abelian complexity of $\mathbf{t}$: its Theorem 3.1 (numbering of arXiv:0911.2914v1) gives $\rho_{\mathbf{t}}^{\mathrm{ab}}(n) = 2$ for n odd and 3 for $n \neq 0$ even, so that complexity is bounded — in contrast with the reduced abelian complexity studied here, which is unbounded by Proposition 8.1. Not used in any proof above.