

The minimum number of distinct abelian squares in a binary circular word

Deep Bhattacharjee^{a,1,*}

^a*Formerly, Electro-Gravitational Space Propulsion Laboratory (EGSPL), Bhubaneswar, 751030, Odisha, India*

Abstract

We prove that every binary circular word of length n contains at least $\lfloor (n-1)/2 \rfloor$ distinct abelian squares, and we determine all words attaining this bound. This settles a conjecture of Simpson: its odd case holds, and in its even case the list of extremal words misses $(ab)^{n/2}$ for $n \equiv 2 \pmod{4}$, $n \geq 6$.

Keywords: abelian square, circular word, Parikh vector, combinatorics on words

2020 MSC: Primary 68R15, Secondary 05A05, 68V15

1. Introduction

Words are over the alphabet $\{a, b\}$. For a letter ℓ we write $\bar{\ell}$ for the other letter and $|u|_{\ell}$ for the number of occurrences of ℓ in a word u . An *abelian square* is a word uv with u nonempty and $|u|_a = |v|_a$, $|u|_b = |v|_b$; then $|u| = |v|$. A circular word of length n is a word $w = w_0 \cdots w_{n-1}$ read cyclically; its *factors* are the words $w_i w_{i+1} \cdots w_{i+L-1}$, indices modulo n , with $0 \leq i < n$ and $1 \leq L \leq n$. As in [1], factors longer than n are not counted. Let $D(w)$ be the number of distinct factors of w that are abelian squares. Conjugate words (cyclic shifts) and complementary words (obtained by exchanging a and b) have the same value of D . Simpson [1] surveyed extremal problems on abelian squares in linear and circular words;

*Corresponding author.

Email addresses: itsdeep@live.com (Deep Bhattacharjee),
d.bhattacharjee@erl-forschung.de (Deep Bhattacharjee)

¹ORCID: <https://orcid.org/0000-0003-0466-750X>

for the minimum number of distinct abelian squares in a circular word he stated the following, based on computer experiments.

Conjecture 0.6 of [1]. The minimum number of distinct abelian squares in a circular word of length n is:

- (a) $(n - 1)/2$ if n is odd and this bound is attained only by a^n , $a^{n-1}b$ and their complements and conjugates.
- (b) $(n - 2)/2$ if n is even and this bound is attained only by $a^k b^{n-k}$ and its complement and their conjugates, where $k \in \{1, 3, 5, \dots, n - 1\}$.

For non-equivalent abelian squares in circular words the minimum was found by Fraenkel, Simpson and Paterson [2]. For linear binary words the minimum number of distinct abelian squares is conjectured by Fici and Saarela to be $\lfloor n/4 \rfloor$, and [3] proves this in special cases; maximal numbers of abelian squares were studied in [4, 5], and [6] surveys the area. We prove:

Theorem 1.1. *Let $n \geq 1$. Every binary circular word w of length n satisfies $D(w) \geq \lfloor (n - 1)/2 \rfloor$. Equality holds if and only if w is a conjugate of one of the following words:*

- (a) n odd: a^n , b^n , $a^{n-1}b$, $b^{n-1}a$;
- (b) n even: $a^k b^{n-k}$ with k odd, and $(ab)^{n/2}$ if $n \equiv 2 \pmod{4}$.

The complement $b^k a^{n-k}$ of $a^k b^{n-k}$ is a conjugate of $a^{n-k} b^k$, and $n - k$ is odd when n is even and k is odd; the complement of $(ab)^{n/2}$ is a conjugate of $(ab)^{n/2}$. Hence the lists in (a) and (b) are closed under complement and conjugation.

Corollary 1.2. *Part (a) of Conjecture 0.6 of [1] holds. In part (b) the minimum $(n - 2)/2$ is correct, and the list of extremal words is complete if and only if $n \not\equiv 2 \pmod{4}$ or $n = 2$. For $n \equiv 2 \pmod{4}$, $n \geq 6$, the words $(ab)^{n/2}$ and $(ba)^{n/2}$ attain the minimum and are not conjugates of $a^k b^{n-k}$ or $b^k a^{n-k}$ for any k .*

Proof. For $n \geq 4$ the word $(ab)^{n/2}$ has $n/2 \geq 2$ maximal blocks of a , while every conjugate of $a^k b^{n-k}$ or $b^k a^{n-k}$ has at most one maximal block of a when read cyclically. For $n = 2$, $(ab)^1 = a^1 b^1$. The rest is Theorem 1.1. $\square$

For example, the abelian squares among the factors of the circular word $ababab$ are $abab$ and $baba$, so $D(ababab) = 2 = (6 - 2)/2$. Counting conjugates,

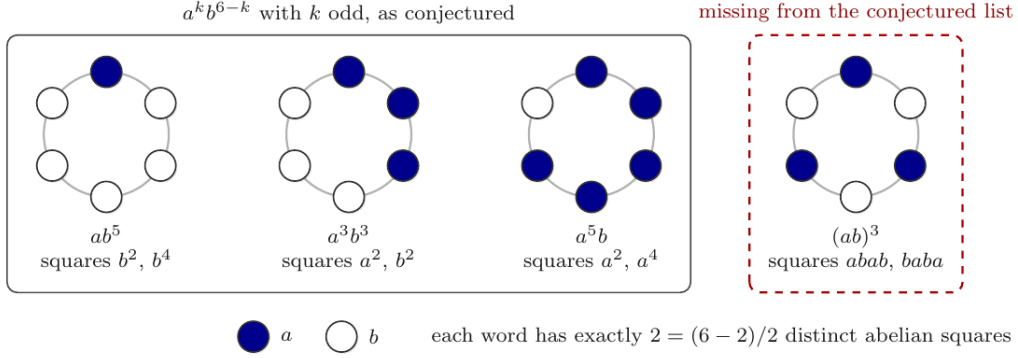


Figure 1: The extremal circular words of length 6, read clockwise from the top, up to complement. The three words on the left are those of Conjecture 0.6(b); the alternating word on the right also has exactly two distinct abelian squares.

the number of words of length n with $D(w) = \lfloor (n-1)/2 \rfloor$ is $2n+2$ for odd $n \geq 3$, and $n^2/2$ for even n , plus 2 when $n \equiv 2 \pmod{4}$ and $n \geq 6$. Figure 1 shows the extremal words of length 6 up to conjugation and complement.

Section 2 proves the lower bound by a double count of the abelian squares of half-length m and $n-m$. Section 3 describes the words in which this count is tight for one letter, and Section 4 completes the proof. Section 5 describes the Lean 4 formalization.

2. The lower bound

Fix a circular word w of length $n \geq 1$ and set $x_i = w_{i \bmod n}$ for $i \in \mathbb{Z}$. For a letter ℓ let $t_\ell = |w|_\ell$, so $t_a + t_b = n$. For $i \in \mathbb{Z}$ and $L \geq 0$ let $c_\ell(i, L)$ be the number of $s \in [i, i+L)$ with $x_s = \ell$; a window of n consecutive positions contains t_ℓ occurrences of ℓ , so $c_\ell(i, L) < t_\ell$ implies $L < n$. For $m \geq 1$ let $W(i, m) = x_i x_{i+1} \cdots x_{i+2m-1}$; we call $W(i, m)$ a *square* if $c_a(i, m) = c_a(i+m, m)$, equivalently $c_b(i, m) = c_b(i+m, m)$.

A factor of w that is an abelian square has even length $2m \leq n$, and it equals $W(i, m)$ for some i for which $W(i, m)$ is a square. Hence

$$D(w) = \sum_{m=1}^{\lfloor n/2 \rfloor} |S_m|, \quad S_m = \{W(i, m) : i \in \mathbb{Z}, W(i, m) \text{ is a square}\}.$$

Definition 2.1. For $1 \leq m \leq n-1$ let Q_m be the set of pairs (ℓ, k) such that some square $W(i, m)$ satisfies $x_{i+m-1} = \ell$ and $c_\ell(i, m) = k$. For a letter ℓ and $k \geq 0$, the *column* $C_\ell(k)$ is the set of $m \in [1, n-1]$ with $(\ell, k) \in Q_m$.

So $(\ell, k) \in Q_m$ records a square of half-length m whose left half ends with ℓ and contains k letters ℓ . The sets S_m , Q_m and $C_\ell(k)$ do not change when w is replaced by a conjugate.

Lemma 2.2. $|Q_m| \leq |S_m|$ for $1 \leq m \leq n/2$.

Proof. The pair (ℓ, k) attached to a square $W(i, m)$ is determined by the word $W(i, m)$: ℓ is its m -th letter and k the number of letters ℓ among its first m letters. So Q_m is the image of S_m under a map. $\square$

Lemma 2.3. Let $1 \leq m \leq n-1$. If $(\ell, k) \in Q_m$, then $(\ell, t_\ell - k) \in Q_{n-m}$. Hence $|Q_m| = |Q_{n-m}|$, and $1 \leq k \leq t_\ell - 1$ for every $(\ell, k) \in Q_m$.

Proof. Let $W(i, m)$ be a square with $x_{i+m-1} = \ell$ and $c_\ell(i, m) = c_\ell(i+m, m) = k$. Consider $W(i+2m, n-m)$. Its left half $[i+2m, i+m+n)$ and the right half $[i+m, i+2m)$ of $W(i, m)$ together form the window $[i+m, i+m+n)$, so the left half contains $t_\ell - k$ letters ℓ . Its right half $[i+m+n, i+2n)$ is the translate by n of $[i+m, i+n)$, which together with $[i, i+m)$ forms a window of length n ; so it also contains $t_\ell - k$ letters ℓ . The last letter of the left half is $x_{i+m+n-1} = x_{i+m-1} = \ell$. Thus $(\ell, t_\ell - k) \in Q_{n-m}$. The map $(\ell, k) \mapsto (\ell, t_\ell - k)$ is injective and is its own inverse, so $|Q_m| = |Q_{n-m}|$. Finally $k \geq 1$ because the left half ends with ℓ , and $t_\ell - k \geq 1$ by the same argument applied to $W(i+2m, n-m)$. $\square$

If $t_\ell \geq 1$, let $(p_j)_{j \in \mathbb{Z}}$ be the increasing enumeration of $\{s \in \mathbb{Z} : x_s = \ell\}$; then $p_{j+t_\ell} = p_j + n$. For $j \in \mathbb{Z}$ and $1 \leq k < t_\ell$ let $I_j(k)$ be the interval of integers m with

$$\max(p_{j+k} - p_{j+1} + 1, p_{j+2k} - p_{j+k}) \leq m \leq \min(p_{j+k} - p_j, p_{j+2k+1} - p_{j+k} - 1). \quad (1)$$

Lemma 2.4. Let $1 \leq k < t_\ell$ and $m \in I_j(k)$. Then $W(p_{j+k} + 1 - m, m)$ is a square whose left half ends with ℓ and contains k letters ℓ . In particular $m \in C_\ell(k)$.

Proof. Let $i = p_{j+k} + 1 - m$. By (1), $p_j < i \leq p_{j+1}$, so the left half $[i, p_{j+k}]$ contains exactly the occurrences $p_{j+1}, \dots, p_{j+k}$ and ends with $x_{p_{j+k}} = \ell$. Also

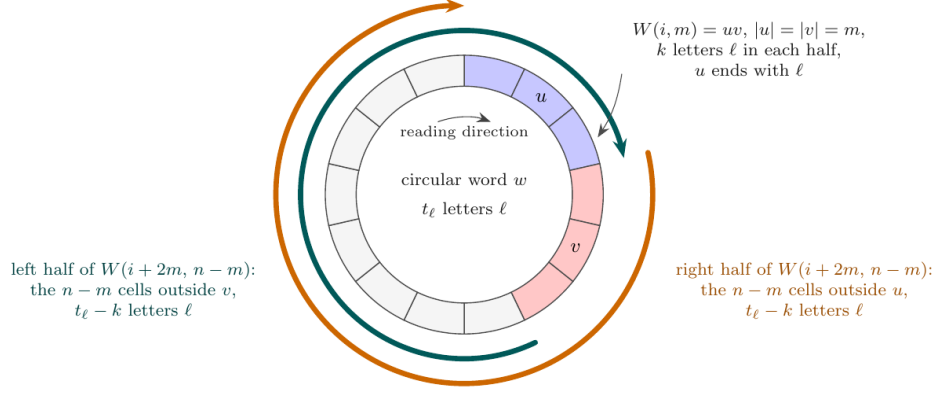


Figure 2: The proof of Lemma 2.3. The two halves of the square $W(i+2m, n-m)$ are the complements of v and of u in the circle, so each contains $t_\ell - k$ letters ℓ , and the left one ends at the last letter of u .

$p_{j+2k} \leq p_{j+k} + m < p_{j+2k+1}$, so the right half $[p_{j+k} + 1, p_{j+k} + m]$ contains exactly $p_{j+k+1}, \dots, p_{j+2k}$. Both halves contain k letters ℓ . Since $m \geq 1$ and $k < t_\ell$, we have $1 \leq m < n$. $\square$

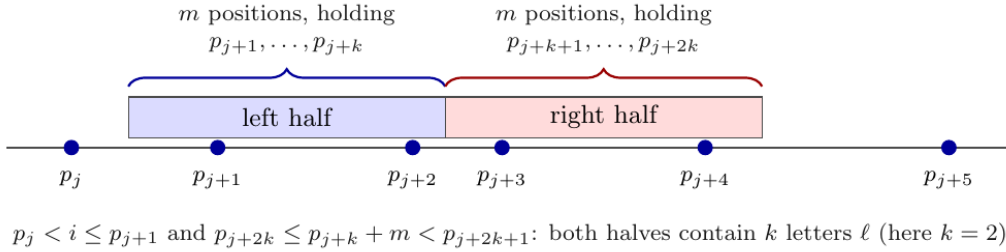


Figure 3: The square of Lemma 2.4 for $k = 2$. The occurrences of ℓ are the dots; the left half ends at p_{j+k} and the right half stops before p_{j+2k+1} .

Lemma 2.5. *For every letter ℓ and $1 \leq k \leq t_\ell - 1$, the column $C_\ell(k)$ is nonempty.*

Proof. Let $d_j = p_{j+k} - p_j$ and $e_j = d_{j+k} - d_j$. Both are periodic in j with period t_ℓ , so $\sum_{j=r}^{r+t_\ell-1} e_j = 0$ for every r . Hence some $e_{j_0} \leq 0$. If $e_j \leq 0$ implied $e_{j+1} < 0$ for every j , then $e_j < 0$ for all $j > j_0$, and $\sum_{j=j_0+1}^{j_0+t_\ell} e_j < 0$.

So there is j with $e_j \leq 0 \leq e_{j+1}$, that is $d_{j+k} \leq d_j$ and $d_{j+1} \leq d_{j+k+1}$. We show $I_j(k) \neq \emptyset$ by comparing each lower bound in (1) with each upper bound. First, $p_{j+k} - p_{j+1} + 1 \leq p_{j+k} - p_j$ and $p_{j+2k} - p_{j+k} \leq p_{j+2k+1} - p_{j+k} - 1$. Second, $p_{j+2k} - p_{j+k} = d_{j+k} \leq d_j = p_{j+k} - p_j$. Third, with $g = p_{j+k+1} - p_{j+k} \geq 1$,

$$\begin{aligned} p_{j+k} - p_{j+1} + 1 &= d_{j+1} - g + 1 \leq d_{j+1} \leq d_{j+k+1} \\ &\leq d_{j+k+1} + g - 1 = p_{j+2k+1} - p_{j+k} - 1. \end{aligned}$$

By Lemma 2.4, $C_\ell(k) \supseteq I_j(k) \neq \emptyset$. $\square$

Theorem 2.6. $D(w) \geq \lfloor (n-1)/2 \rfloor$.

Proof. Let $N = \sum_{m=1}^{n-1} |Q_m|$ and $F = \sum_{m=1}^{\lfloor n/2 \rfloor} |Q_m|$. Counting the pairs $(m, (\ell, k))$ with $(\ell, k) \in Q_m$ by columns,

$$N = \sum_{\ell} \sum_{k \geq 0} |C_\ell(k)|, \quad \sum_{k \geq 0} |C_\ell(k)| \geq t_\ell - 1 \quad \text{if } t_\ell \geq 1, \quad (2)$$

by Lemmas 2.3 and 2.5. If both letters occur, $N \geq t_a + t_b - 2 = n - 2$; otherwise $N \geq n - 1$. By Lemma 2.3, $|Q_m| = |Q_{n-m}|$, so

$$N + \varepsilon |Q_{n/2}| = 2F, \quad \varepsilon = 1 \text{ if } n \text{ is even, } \varepsilon = 0 \text{ if } n \text{ is odd.} \quad (3)$$

Hence $2F \geq n - 2$ and $F \geq \lceil (n-2)/2 \rceil = \lfloor (n-1)/2 \rfloor$. By Lemma 2.2, $D(w) \geq F$. $\square$

We call a letter ℓ with $t_\ell \geq 1$ *tight* if $|C_\ell(k)| \leq 1$ for every k . By Lemmas 2.3 and 2.5, ℓ is tight if and only if $\sum_k |C_\ell(k)| = t_\ell - 1$. We call w *rigid* if $|S_m| = |Q_m|$ for $1 \leq m \leq n/2$: two squares $W(i, m)$, $W(i', m)$ with $2m \leq n$, $x_{i+m-1} = x_{i'+m-1} = \ell$ and $c_\ell(i, m) = c_\ell(i', m)$ are equal words.

Proposition 2.7. *If $D(w) = \lfloor (n-1)/2 \rfloor$, then w is rigid and $N + \varepsilon |Q_{n/2}| = 2 \lfloor (n-1)/2 \rfloor$.*

Proof. In the proof of Theorem 2.6, $\lfloor (n-1)/2 \rfloor \leq F \leq D(w)$, so $F = D(w)$, and $|Q_m| \leq |S_m|$ for each m forces $|Q_m| = |S_m|$. Then (3) gives the second statement. $\square$

3. Rigid words with a tight letter

Theorem 3.1. *Let w be rigid and let ℓ be a tight letter, $t = t_\ell$. Then w is a conjugate of $\ell^t \bar{\ell}^{n-t}$ with t odd or $t = n$, or $n = 2t$ and w is a conjugate of $(\ell \bar{\ell})^t$.*

Throughout this section w and ℓ are as in Theorem 3.1, (p_j) enumerates the occurrences of ℓ , and $g_j = p_{j+1} - p_j \geq 1$. Since ℓ is tight, Lemma 2.4 gives:

for $1 \leq k < t$, the integers in $\bigcup_j I_j(k)$ form a set with at most one element. (4)

Proof of Theorem 3.1. If $t = n$, then $w = \ell^n$. Assume $t < n$, so $g_j \geq 2$ for some j .

Case A: $g_j \geq 2$ for all j . If $t = 1$, then w is a conjugate of $\ell \bar{\ell}^{n-1}$. Let $t \geq 2$. By (1), $I_j(1)$ is the set of m with $g_{j+1} \leq m \leq \min(g_j, g_{j+1} + g_{j+2} - 1)$. If $g_j > g_{j+1}$, then $I_j(1) \supseteq \{g_{j+1}, g_{j+1} + 1\}$ since $g_{j+2} \geq 2$, contradicting (4). So $g_j \leq g_{j+1}$ for all j , and as g is periodic, $g_j = \delta$ is constant and $n = t\delta$. If $\delta = 2$, then $n = 2t$ and w is a conjugate of $(\ell \bar{\ell})^t$. Let $\delta \geq 3$, so $p_j = p_0 + j\delta$. The words $W(p_1, \delta)$ and $W(p_1 + 2, \delta)$ have halves $\ell \bar{\ell}^{\delta-1}$, $\ell \bar{\ell}^{\delta-1}$ and $\bar{\ell}^{\delta-2} \ell \bar{\ell}$, $\bar{\ell}^{\delta-2} \ell \bar{\ell}$. Both are squares whose left halves end with $\bar{\ell}$ and contain $\delta - 1$ letters $\bar{\ell}$, and $2\delta \leq n$. By rigidity they are equal, but their first letters are ℓ and $\bar{\ell}$.

Case B: $g_j = 1$ for some j . A run of length $L \geq 1$ consists of occurrences $p_{J+1}, \dots, p_{J+L}$ with $g_J \geq 2$, $g_{J+1} = \dots = g_{J+L-1} = 1$ and $g_{J+L} \geq 2$; its gaps are $\lambda = g_J$ and $\rho = g_{J+L}$. Since g has period t and some $g_j \geq 2$, every run has length $L \leq t$. The two occurrences p_j, p_{j+1} with $g_j = 1$ lie in a run of length at least 2.

(B1) *Every run has odd length.* Suppose a run has length $2i$, $i \geq 1$. In (1) for $I_J(i)$ the lower bounds are i and i , and the upper bounds are $\lambda + i - 1 \geq i + 1$ and $i + \rho - 1 \geq i + 1$. So $I_J(i) \supseteq \{i, i + 1\}$ with $i < 2i \leq t$, contradicting (4).

For a run $p_{J+1}, \dots, p_{J+2h+1}$ with $h \geq 1$ and gaps λ, ρ , (1) gives, with $\lambda' = g_{J-1}$ and $\rho' = g_{J+2h+2}$,

$$I_{J-1}(h+1) = [\lambda + h, \min(\lambda + h + \lambda' - 1, h + \rho)], \quad (5)$$

$$I_J(h+1) = [\rho + h, \min(\lambda + h, h + \rho + \rho' - 1)], \quad (6)$$

where we used $\lambda, \rho \geq 2$. Here $h + 1 < 2h + 1 \leq t$.

(B2) *All runs of length at least 3 have the same length.* Let a run $p_{J+1}, \dots, p_{J+2h+1}$ have $h \geq 1$, and suppose another run $p_{J'+1}, \dots, p_{J'+L'}$ has $L' \geq 2h+2$. In (1) for $I_{J'}(h+1)$ the lower bounds are $h+1$ and $h+1$, and the upper bounds are at least $h+1$; so $h+1 \in I_{J'}(h+1)$. By (5) and (6), $\lambda+h \in I_{J-1}(h+1)$ if $\lambda \leq \rho$ and $\rho+h \in I_J(h+1)$ if $\rho \leq \lambda$. Both numbers exceed $h+1$, contradicting (4).

By (B1) and (B2) there is $h \geq 1$ such that every run has length 1 or $2h+1$, and some run has length $2h+1$. Consider such a run $R = (p_{J+1}, \dots, p_{J+2h+1})$ with gaps λ, ρ .

(B3) *$\rho < \lambda$ is impossible.* Let R' be the run following R , starting at p_{J+2h+2} . If R' has length 1, then $\rho' \geq 2$, and by (6) $I_J(h+1) \supseteq \{\rho+h, \rho+h+1\}$. If R' has length $2h+1$, let $\sigma \geq 2$ be its right gap. Then $t \neq 2h+1$: otherwise R' is the translate of R by n and $\rho = g_{J+t} = g_J = \lambda$. In (1) for $I_J(2h+1)$ the lower bounds are $2h+1$ and $2h+\rho$, and the upper bounds are $2h+\lambda \geq 2h+\rho+1$ and $2h+\rho+\sigma-1 \geq 2h+\rho+1$; so $I_J(2h+1) \supseteq \{2h+\rho, 2h+\rho+1\}$ with $2h+1 < t$. Both cases contradict (4).

(B4) *If $\lambda < \rho$, the occurrence p_J is not a run of length 1.* Otherwise $\lambda' \geq 2$, and by (5) $I_{J-1}(h+1) \supseteq \{\lambda+h, \lambda+h+1\}$, contradicting (4).

(B5) *If $\lambda = \rho$, then $t = 2h+1$ and w is a conjugate of $\ell^{2h+1}\bar{\ell}^{\lambda-1}$.* By (5) and (6), $m = \lambda+h$ lies in $I_{J-1}(h+1)$ and in $I_J(h+1)$. By Lemma 2.4 the squares $W(p_J, m)$ and $W(p_J+1, m)$ have left halves ending with ℓ and containing $h+1$ letters ℓ . Their first letters are $x_{p_J} = \ell$ and $x_{p_J+1} = \bar{\ell}$, since $p_J < p_J+1 < p_{J+1}$. By rigidity, $2m > n$. Then $p_{J+t} = p_J+n < p_J+2\lambda+2h = p_{J+2h+2}$, so $t \leq 2h+1$, and $t \geq 2h+1$ as R has $2h+1$ occurrences. So $t = 2h+1$, $n = p_{J+2h+2} - p_{J+1} = 2h+\lambda$, and $x_{p_{J+1}} \cdots x_{p_{J+1}+n-1} = \ell^{2h+1}\bar{\ell}^{\lambda-1}$.

Now choose the run R of length $2h+1$ with the least left gap λ . By (B3), $\rho \geq \lambda$. If $\rho = \lambda$, (B5) applies. If $\rho > \lambda$, then by (B4) the run R^- preceding R has length $2h+1$. Its right gap is λ , and its left gap is at least λ by the choice of R ; by (B3) the two are equal, and (B5) applies to R^- . In all cases w is a conjugate of $\ell^t\bar{\ell}^{n-t}$ with $t = 2h+1$ odd. $\square$

4. Proof of Theorem 1.1

Lemma 4.1. *Let ℓ be a letter.*

- (i) *Let $0 \leq \alpha \leq n$ and $\beta = n - \alpha$, and suppose each of α, β is 0, odd, or at least $n-1$. Then every abelian square factor of the circular word $\ell^\alpha\bar{\ell}^\beta$ is ℓ^{2m} or $\bar{\ell}^{2m}$, and $D(\ell^\alpha\bar{\ell}^\beta) = \lfloor \alpha/2 \rfloor + \lfloor \beta/2 \rfloor$.*

(ii) For even n , $D((\ell\bar{\ell})^{n/2}) = 2\lfloor n/4 \rfloor$.

Proof. (i) Let $s = uv$ be an abelian square factor, $|u| = |v| = m$, $2m \leq n$, containing both letters; then $1 \leq \alpha, \beta \leq n-1$, and read cyclically the word consists of the two blocks ℓ^α and $\bar{\ell}^\beta$. If s crosses exactly one block boundary, $s = c^r \bar{c}^{2m-r}$ with $c \in \{\ell, \bar{\ell}\}$ and $0 < r < 2m$, and $|u|_c = \min(r, m) > \max(r-m, 0) = |v|_c$. Otherwise s contains a whole block in its interior, say $s = \ell^r \bar{\ell}^\beta \ell^{r'}$ with $r, r' \geq 1$. Then $\beta \leq 2m-2 \leq n-2$, so β is odd, whereas $\beta = |s|_{\bar{\ell}} = 2|u|_{\bar{\ell}}$. The case $s = \bar{\ell}^r \ell^\alpha \bar{\ell}^{r'}$ is symmetric. The factor ℓ^{2m} occurs if and only if $2m \leq \alpha$, and $\bar{\ell}^{2m}$ if and only if $2m \leq \beta$.

(ii) Every factor of length $2m$ alternates. If m is odd, its halves start with different letters $c \neq \bar{c}$, and $|u|_c = (m+1)/2 \neq (m-1)/2 = |v|_c$. If m is even, both halves contain $m/2$ letters of each kind, and there are exactly two factors of length $2m \leq n$. So $D = 2|\{m : m \text{ even}, 2 \leq m \leq n/2\}| = 2\lfloor n/4 \rfloor$. $\square$

Lemma 4.2. Let $\alpha \geq 3$, $h \geq 1$, $n = \alpha + 2h$ and $w = \ell^\alpha \bar{\ell}^{2h}$. Then $C_{\bar{\ell}}(h) \supseteq \{h, h+1, n-h\}$, three distinct numbers.

Proof. $W(\alpha, h) = \bar{\ell}^{2h}$ is a square whose left half $\bar{\ell}^h$ ends with $\bar{\ell}$, so $h \in C_{\bar{\ell}}(h)$. $W(\alpha-1, h+1) = \ell \bar{\ell}^{2h} \ell$ is a square with halves $\ell \bar{\ell}^h$ and $\bar{\ell}^h \ell$, and $2h+2 \leq n$; so $h+1 \in C_{\bar{\ell}}(h)$. By Lemma 2.3 with $t_{\bar{\ell}} = 2h$, $n-h \in C_{\bar{\ell}}(2h-h)$. Finally $n-h = \alpha + h \geq h+3$. $\square$

Proof of Theorem 1.1. The lower bound is Theorem 2.6. If w is a conjugate of a word in the list, Lemma 4.1 gives $D(w) = \lfloor (n-1)/2 \rfloor$: for a^n (n odd) $\lfloor n/2 \rfloor = (n-1)/2$; for $a^{n-1}b$ (n odd) $(n-1)/2 + 0$; for $a^k b^{n-k}$ (n even, k odd) $(k-1)/2 + (n-k-1)/2 = (n-2)/2$; for $(ab)^{n/2}$ ($n \equiv 2 \pmod{4}$) $2(n-2)/4$. Complements and conjugates have the same D .

Conversely let $D(w) = \lfloor (n-1)/2 \rfloor$. By Proposition 2.7, w is rigid and $N + \varepsilon|Q_{n/2}| = 2\lfloor (n-1)/2 \rfloor$.

n even. Then $N + |Q_{n/2}| = n-2$. If only one letter occurs, $N \geq n-1$; so both occur, and by (2) $N \geq n-2$. Hence $|Q_{n/2}| = 0$ and $\sum_k |C_\ell(k)| = t_\ell - 1$ for both letters, so both letters are tight. Theorem 3.1 for $\ell = a$ shows that w is a conjugate of $a^k b^{n-k}$ with k odd ($k = n$ is excluded since b occurs), or $n = 2t_a$ and w is a conjugate of $(ab)^{n/2}$. In the second case, if $n \equiv 0 \pmod{4}$, both halves of $W(0, n/2)$ alternate and have even length $n/2$, so $W(0, n/2)$ is a square and $Q_{n/2} \neq \emptyset$, a contradiction. So $n \equiv 2 \pmod{4}$.

n odd. Then $N = n-1$. If only one letter occurs, w is a^n or b^n . Otherwise (2) gives $\sum_k |C_a(k)| + \sum_k |C_b(k)| = (t_a - 1) + (t_b - 1) + 1$ with $\sum_k |C_\ell(k)| \geq$

$t_\ell - 1$ for both letters. So one letter ℓ is tight, and $\sum_k |C_{\bar{\ell}}(k)| \leq t_{\bar{\ell}}$. By Theorem 3.1, w is a conjugate of $\ell^t \bar{\ell}^{n-t}$ with $t = t_\ell$ odd, since $t < n$ and n is odd. Then $n - t = 2h$ with $h \geq 1$. If $t = 1$, then w is a conjugate of $\bar{\ell}^{n-1} \ell$. If $t \geq 3$, Lemma 4.2 gives $|C_{\bar{\ell}}(h)| \geq 3$ with $1 \leq h < 2h = t_{\bar{\ell}}$, and with Lemma 2.5, $\sum_k |C_{\bar{\ell}}(k)| \geq (t_{\bar{\ell}} - 1) + 2 > t_{\bar{\ell}}$, a contradiction. $\square$

5. Formal verification

Theorem 1.1 and Corollary 1.2 are formalized in Lean 4 [7] with Mathlib [8], version 4.34.1, and the formal proofs are checked by the Lean kernel. Words are lists over a two-letter alphabet, and the definitions follow Section 1 literally: an abelian square is a word uv with u nonempty and the same number of each letter in u and v ; the factors of a circular word w of length n are the prefixes of length L , $1 \leq L \leq n$, of its n rotations; and $D(w)$ is the number of distinct factors that are abelian squares. The formal statements are the following. For every $n \geq 1$ the least value of D on the words of length n is $\lfloor (n-1)/2 \rfloor$. A word of odd length, and a word of even length, attains this value exactly when it is a conjugate of one of the words in (a), respectively (b), of Theorem 1.1. Part (a) of Conjecture 0.6 holds, and part (b), stated literally, is false. Finally $D(ababab) = 2$, and $ababab$ is not a conjugate of $a^k b^{6-k}$ or of its complement for any k . The proofs depend only on the axioms of Lean's logic, namely propositional extensionality, choice and the soundness of quotients; they contain no unproved step and add no axiom.

Independent computations confirm Theorem 1.1 on every binary circular word of length at most 22 and on one word of each necklace class up to length 28, and they confirm each lemma of Sections 2 and 3 on every word of length at most 16.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author used Claude (Anthropic) in order to assist with the proofs, the Lean formalization, the computations and the preparation of the manuscript. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Data availability

Deep Bhattacharjee computed the Lean formalization and the Julia, C, C++, shell and Python computations with the help of AI assistance (Claude, Anthropic). The Lean formalization of the proofs in this paper and the programs are in release v2.1.0 of the repository <https://github.com/creelie/literate-octo-lamp>, <https://github.com/creelie/literate-octo-lamp/releases/tag/v2.1.0>.

Competing interests

None.

Funding

None.

References

- [1] J. Simpson, Solved and unsolved problems about abelian squares, preprint (2018), arXiv:1802.04481.
- [2] A.S. Fraenkel, J. Simpson, M. Paterson, On weak circular squares in binary words, in: Combinatorial Pattern Matching (CPM 1997), Lecture Notes in Computer Science, Springer, 1997, pp. 76–82, doi:10.1007/3-540-63220-4_51.
- [3] S.Z. Fazekas, A. Mammoliti, R. Mercas, J. Simpson, Binary words containing few abelian squares, preprint (2026), arXiv:2604.23188.
- [4] G. Fici, F. Mignosi, Words with the maximum number of abelian squares, in: Combinatorics on Words (WORDS 2015), Lecture Notes in Computer Science, Springer, 2015, pp. 122–134, doi:10.1007/978-3-319-23660-5_11.
- [5] G. Fici, F. Mignosi, J. Shallit, Abelian-square-rich words, Theoret. Comput. Sci. 684 (2017) 29–42, doi:10.1016/j.tcs.2017.02.012.
- [6] G. Fici, S. Puzynina, Abelian combinatorics on words: a survey, Comput. Sci. Rev. 47 (2023) 100532, doi:10.1016/j.cosrev.2022.100532.

- [7] L. de Moura, 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.
- [8] 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, pp. 367–381, doi:10.1145/3372885.3373824.