Improved Bounds on the Maximum Number of Distinct Squares in Circular Words

Panagiotis Charalampopoulos
King's College London, UK

Manal Mohamed
King's College London, UK

Jakub Radoszewski
University of Warsaw, Poland

Wojciech Rytter
University of Warsaw, Poland

Tomasz Waleń
University of Warsaw, Poland

Wiktor Zuba
University of Warsaw, Poland

Abstract

We investigate the asymptotic growth of function $\mathsf{CS}(n)$, which maps n to the maximum number of distinct squares in a circular word of length n (that is, the maximum number of distinct squares of length at most n in a word ww of length $2n$). We improve upon the lower bound of $1.25n$ established by Amit and Gawrychowski [SPIRE 2017] and the straightforward upper bound of $2n$, which follows from the recent result of Brlek and Li [*Comb. Theory*, 2025] stating that there are fewer than n squares in standard (i.e., non-circular) words of length n. (Previously, Amit and Gawrychowski gave an upper bound of $3\frac{2}{15}n$ using a weaker upper bound on squares in standard words.) Specifically, we show that $\mathsf{CS}(n) \le \lceil 1.8n \rceil$ and that, for infinitely many n, $\mathsf{CS}(n) \ge 1.5n - \mathcal{O}(\sqrt{n})$.

For the lower bound, we exploit the combinatorial structure of Fibonacci words to construct a family of square-rich circular words. For the upper bound, we exploit density properties of the starting positions of long squares, adapting an approach of Amit and Gawrychowski.

2012 ACM Subject Classification Mathematics of computing → Combinatorics on words

Keywords and phrases circular words, squares, repetitions

Digital Object Identifier 10.4230/LIPIcs.CPM.2026.6

Funding *Jakub Radoszewski*: Supported by the Polish National Science Center, grant no. 2022/46/E/ST6/00463.

1 Introduction

The study of square and symmetric fragments of words is central in combinatorics and algorithms on words. A square is a word of the form uu for a word u. Here, we focus on square fragments of circular words. That is, for a standard word w, we study the square fragments that occur in any rotation of w or, equivalently, the square fragments of ww that are of length at most $|w|$.

The earliest result concerning squares in words is the construction of square-free ternary words of all lengths by Thue [19]. The analogue of this problem for circular words is more intricate. Currie [4] showed that square-free circular words over the ternary alphabet exist for all lengths except for 5, 7, 9, 10, 14, and 17 using a computer-aided proof. Shur [16] later gave a computer-free proof that also implied an exponential (in the length) lower bound on the number of such words of any fixed length.

© Panagiotis Charalampopoulos, Manal Mohamed, Jakub Radoszewski, Wojciech Rytter, Tomasz Waleń, and Wiktor Zuba;
licensed under Creative Commons License CC-BY 4.0
37th Annual Symposium on Combinatorial Pattern Matching (CPM 2026).
Editors: Philip Bille and Nicola Prezza; Article No. 6; pp. 6:1–6:12
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Let $\mathsf{SQ}(n)$ be the maximum number of distinct squares in a (standard) word of length n. Fraenkel and Simpson [7] proved that $\mathsf{SQ}(n) < 2n$ and conjectured that $\mathsf{SQ}(n) < n$. The resolution of this conjecture took more than two decades. Milestone results include the upper bound of $2n - \log n$ due to Ilie [9], the upper bound of $\frac{11}{6}n$ due to Deza, Franek, and Thierry [5], and the upper bound of $1.5n$ due to Thierry [18]. Finally, Brlek and Li [3] confirmed the Fraenkel-Simpson conjecture using a graph-theoretic approach, establishing that $\mathsf{SQ}(n) < n$.

Let $\mathsf{CS}(n)$ denote the maximum number of distinct squares in a circular word of length n. The growth of $\mathsf{CS}(n)$ was first studied by Amit and Gawrychowski [1] who showed an upper bound of $3\frac{2}{15}n$ and a lower bound of $1.25n$ (for infinitely many n). Note that this work was published when the best known upper bound for standard squares was $\mathsf{SQ}(n) \le \frac{11}{6}n$. Further, observe that the number of squares of length at most n in a word ww is at most $\mathsf{SQ}(2|w|)$. The result of Brlek and Li [3] thus implies that $\mathsf{CS}(n) \le \mathsf{SQ}(2n) < 2n$.

Our results. Here, we improve upon both the upper bound of $2n$ that follows from [3] and the lower bound of $1.25n$ from [1] for $\mathsf{CS}(n)$. Specifically, we show that $\mathsf{CS}(n) \le \lceil 1.8n \rceil$ for all n and that $\mathsf{CS}(n) \ge 1.5n - \mathcal{O}(\sqrt{n})$ for infinitely many n.

Our lower and upper bounds are presented in Section 3 and Section 4, respectively. Then, Section 5 discusses the limitations of our approach.

Discussion. As we remark at the end of Section 4, the proof of the $3\frac{2}{15}n$ upper bound of Amit and Gawrychowski could be modified to yield an upper bound of $\lceil \frac{13}{7}n \rceil$ by incorporating the upper bound of n for standard squares. However, the resulting argument remains technically involved. We establish a stronger upper bound of $\lceil 1.8n \rceil$ and also a (still stronger) upper bound of $\lceil \frac{11}{6}n \rceil$ which is considerably simpler.

2 Preliminaries

A word $u = u[0] \cdots u[n-1]$ is a sequence of length $|u| = n$ of letters from some alphabet. For any two integers $i, j \in [0\,..\,|u|)$, $u[i\,..\,j]$ and $u[i\,..\,j+1)$ denote the fragment $u[i] \cdots u[j]$ if $i \le j$ and the empty word otherwise. A fragment $u[0\,..\,j]$ is called a *prefix* of u. A prefix of u is called a *proper prefix* of u if its length smaller than n. A fragment $u[i\,..\,n-1]$ is called a *suffix* of u.

We say that integer $p \in [1\,..\,|u|]$ is a *period* of a word u if $u[i] = u[i+p]$ for all $i \in [0\,..\,|u|-p)$. We denote the smallest period of u by $\mathrm{per}(u)$. We say that u is *periodic* if $\mathrm{per}(u) \le \frac{1}{2}|u|$; otherwise u is called *non-periodic*.

▶ **Lemma 1** (Periodicity Lemma [6])**.** *If a word u has periods p and q such that $p + q \le |u| + \gcd(p, q)$, then $\gcd(p, q)$ is also a period of u.*

A word u is called primitive if the equality $u = v^k$ for a positive integer k implies that $k = 1$. Primitive words satisfy the following synchronization property that follows from the periodicity lemma: a primitive word u occurs in u^2 only as a prefix and as a suffix.

For a word u, for any integer $i \in [0\,..\,|u|)$, we call $u[i\,..\,|u|)u[0\,..\,i)$ a *rotation* of u.

For a word w, we denote by $\mathbf{S}_n(w)$ the set of squares of length at most n that occur w. Thus, the set of squares of a circular word w equals $\mathbf{S}_n(w^2)$, where $n = |w|$.

Fibonacci words are defined by the recurrence

$$F_0 = b, \quad F_1 = a, \quad F_t = F_{t-1}F_{t-2} \text{ for } t \ge 2.$$

The number of distinct squares in the (standard) word F_n is equal to $2|F_{n-2}|-2$; see [8]. As a warm-up, we show an exact bound on the number of squares in circular Fibonacci words.

▶ **Fact 2.** *If $n \ge 2$, $|\mathbf{S}_{|F_n|}(F_nF_n)| = |F_n| - 2$.*

Proof. Every square in F_n is a square of a rotation of some shorter Fibonacci word; see [10, Theorem 2.3]. The same holds for F_n^2 since F_n^2 is a prefix of the infinite Fibonacci word for $n \ge 3$, and for $n \le 2$ it can be easily verified.

For $n = 2$, we have $\mathbf{S}_{|F_n|}(F_nF_n) = \emptyset$ as expected. Assume $n \ge 3$. The set $\mathbf{S}_{|F_n|}(F_n^2)$ does not contain $F_0^2 = bb$ and F_{n-1}^2 (because $2|F_{n-1}| > |F_n|$). Let us show by induction that for all $k \in [1\,..\,n-2]$, each rotation of F_k^2 occurs in F_n^2. Correctness in the base cases of $n \in \{3, 4\}$ can be easily checked. Assume that $n \ge 5$ and the property holds for $n-1$. By the hypothesis, F_{n-1}^2 contains all rotations of F_k^2 for $k \in [1\,..\,n-3]$. Since $n \ge 4$, F_{n-1}^2 is a prefix of F_n^2, so all these squares are also in $\mathbf{S}_{|F_n|}(F_n^2)$. Moreover, F_n has a suffix F_{n-2} and a prefix F_{n-2}^2 (since $n-2 \ge 3$). Hence, F_n^2 has a fragment F_{n-2}^3, so each rotation of F_{n-2}^2 occurs in F_n^2.

In total, we obtain $|F_1| + |F_2| + \cdots + |F_{n-2}| = |F_n| - 2$ squares. All these squares are different because each Fibonacci word is primitive. ◀

▶ **Example 3.** We have

$$F_2 = a\,b,\ F_3 = ab\,a,\ F_4 = aba\,ab,\ F_5 = abaab\,aba,\ F_6 = abaababa\,abaab.$$

For F_6, we have

$$\mathbf{S}_{13}(F_6) = \{a^2, (ab)^2, (ba)^2, (aba)^2, (aab)^2, (baa)^2, (abaab)^2, (baaba)^2\},$$

and

$$\mathbf{S}_{13}(F_6^2) = \mathbf{S}_n(F_6) \cup \{(aabab)^2, (ababa)^2, (babaa)^2\}.$$

In particular, F_6 has $8 = 2\,|F_3| - 2$ distinct squares, while the circular F_6 contains $11 = |F_6| - 2$ distinct squares in total.

3 Lower Bound

We use Fibonacci words as building blocks. For any word v of length at least two, let $\mathbf{cut}(v) := v[0\,..\,|v|-2)$ be the word obtained from v by deleting its two last letters. Let $\mathbf{ex}(v)$ denote the word v with the last two letter exchanged, that is, $\mathbf{ex}(v) := \mathbf{cut}(v)v[|v|-1]v[|v|-2]$.

▶ **Example 4.** $\mathbf{cut}(abcde) = abc$ and $\mathbf{ex}(abcd) = abdc$.

The following (folklore) property of Fibonacci words follows by a straightforward inductive argument.

▶ **Observation 5.** *For $t \ge 2$, we have $\mathbf{cut}(F_t) = \mathbf{cut}(F_{t-2}F_{t-1})$ and $\mathbf{ex}(F_t) = F_{t-2}F_{t-1}$.*

The next observation follows directly from the synchronization property of primitive words; see Figure 1.

▶ **Observation 6** (Squares in periodic fragments)**.**

(a) *If u' is a proper prefix of a primitive word u, then uuu' contains $|u'| + 1$ distinct squares of length $|uu|$.*

(b) *u^k contains at least $\lfloor \frac{k-1}{2} \rfloor |u|$ words that are rotations of words in $\{u^{2j} \mid 1 \le j \le \frac{1}{2}k\}$ and are pairwise distinct.*

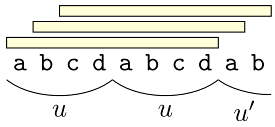

Figure 1 Illustration of Observation 6(a). For the primitive word $u = \texttt{abcd}$ and its proper prefix $u' = \texttt{ab}$, the word u^2u' contains $|u'| + 1$ distinct squares of length $|u^2|$: $(\texttt{abcd})^2$, $(\texttt{bcda})^2$, $(\texttt{cdab})^2$.

Before describing our construction, we first identify the primitive roots of the squares under consideration. These roots correspond to rotations of squares of words of the following types.

▶ **Observation 7** (Primitive words). *For any $i > 0$ and $t \geq 1$, the words F_{t-1}, $F_t^i F_{t-1}$, and $F_t^i F_{t-1} F_t$ are primitive.*

Proof. For any $t \geq 1$, words F_{t-1} and $F_t^k F_{t-1}$ are standard Sturmian words for $k \geq 1$. Since every standard Sturmian word is primitive (cf. [15]), it follows that all words of the form F_{t-1} and $F_t^i F_{t-1}$ are primitive. Moreover, since any rotation of a primitive word is primitive, the word $F_t^i F_{t-1} F_t$ is primitive as its rotation $F_t^{i+1} F_{t-1}$ is primitive. ◀

The construction. We define an infinite family of words. For any $k, t \in \mathbb{Z}_+$, let

$$\mathcal{W}_{k,t} := F_t^k F_{t-1} F_t^{k+1} F_{t-1} F_t^{k+2} F_{t-1}.$$

Note that $\mathcal{W}_{k,t}$ is parametrized by both k and t (eventually, we will set $k := |F_t|$), and let $n := |\mathcal{W}_{k,t}|$. For convenience, we express bounds on the lengths of fragments and the number of square fragments in terms of

$$\alpha := |F_t^k| = k|F_t|.$$

Note that we have

$$|\mathcal{W}_{k,t}| \;=\; (3k+3)|F_t| + 3|F_{t-1}| \;=\; 3\alpha \,+\, 3\,|F_{t+1}|. \tag{1}$$

Squares in $\mathcal{W}_{k,t}$ and $\mathcal{W}^2_{k,t}$. We show that the words $\mathcal{W}_{k,t}$ are almost cubes (except the last two letters); see Figure 2 for an illustration.

▶ **Lemma 8.** *For all $k, t \in \mathbb{Z}_+$, we have that $\mathbf{cut}(\mathcal{W}_{k,t}) = \mathbf{cut}((F_t^k F_{t-1} F_t)^3)$. Further, $\mathcal{W}_{k,t}$ contains at least α distinct squares which are rotations of $(F_t^{k+1} F_{t-1})^2$.*

Proof. We have

$$\mathcal{W}_{k,t} \;=\; F_t^k F_{t-1} F_t^{k+1} F_{t-1} F_t^{k+2} F_{t-1} \;=\; (F_t^k F_{t-1} F_t)(F_t^k F_{t-1} F_t) F_t^k F_t F_{t-1}.$$

By Observation 5,

$$\mathbf{cut}(\mathcal{W}_{k,t}) \;=\; (F_t^k F_{t-1} F_t)(F_t^k F_{t-1} F_t)(F_t^k\;\mathbf{cut}(F_{t-1} F_t)) \;=\; \mathbf{cut}((F_t^k F_{t-1} F_t)^3).$$

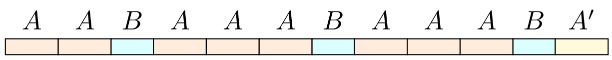

Figure 2 The structure of $\mathcal{W}_{k,t}$ for $k = 2$ and $t \geq 2$: $\mathcal{W}_{k,t} = A^2\,B\,A^3\,B\,A^4\,B$, where $A = F_t$ and $B = F_{t-1}$. We have $\mathcal{W}_{k,t} = (AABA)^2\,AABA'$, where $A' = \mathbf{ex}(A)$. We have $\alpha = |F_t^k| = 2|A|$ and $|\mathcal{W}_{2,t}| = 3\alpha + 3|AB|$.

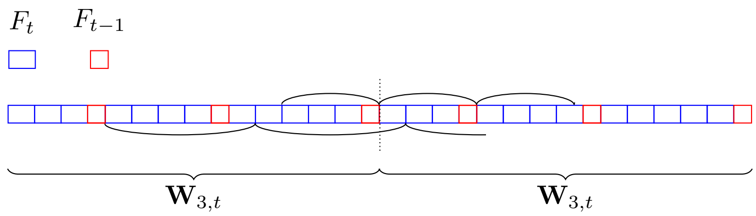

Figure 3 Illustration of Lemma 9: Structure of $\mathcal{W}^2_{k,t}$ for $k=3$. The upper periodic fragment $\mathcal{R}_1 = (F_t^k F_{t-1})^3$ in $\mathcal{W}^2_{3,t}$ does not occur in $\mathcal{W}_{3,t}$. The lower periodic fragment is $\mathcal{R}_2 = (F_t^{k+1}F_{t-1}F_t)^2 F_t^k$. Together, these two periodic fragments contain at least 2α distinct squares of length at most n.

Hence, $\mathbf{cut}((F_t^k F_{t-1} F_t)^3)$ is a prefix of $\mathcal{W}_{k,t}$. Now, $F_t^k F_{t-1} F_t$ is primitive due to Observation 7 and the second part of the statement follows by a direct application of Observation 6. ◀

▶ **Lemma 9.** *If $t \geq 2$ and $k > 0$, $\mathcal{W}^2_{k,t}$ contains more than 2α distinct squares that are rotations of $(F_t^k F_{t-1})^2$ and $(F_t^{k+2} F_{t-1})^2$.*

Proof. We denote $\mathcal{R}_1 := (F_t^k F_{t-1})^3$ and $\mathcal{R}_2 := (F_t^{k+1}F_{t-1}F_t)^2 F_t^k$. We show that both $\mathcal{R}_1$ and $\mathcal{R}_2$ occur in $\mathcal{W}^2_{k,t}$; see Figure 3. Then we obtain the statement using Observation 7 and Observation 6.

We first show that $\mathcal{R}_1$ occurs in $\mathcal{W}^2_{k,t}$. The word

$$\mathcal{W}^2_{k,t} \;=\; F_t^k F_{t-1} F_t^{k+1} F_{t-1} \boxed{\;F_t^{k+2} F_{t-1}\; F_t^k F_{t-1} F_t^{k+1}\;}\; F_{t-1} F_t^{k+2} F_{t-1}$$

has a fragment equal to $F_t^{k+2} F_{t-1} F_t^k F_{t-1} F_t^{k+1}$, which in turn has a fragment equal to $(F_t^k F_{t-1})(F_t^k F_{t-1}) F_t^k F_{t-1} = \mathcal{R}_1$. Thus, $\mathcal{W}^2_{k,t}$ contains more than α distinct squares which are rotations of $(F_t^k F_{t-1})^2$.

Next, we show that $\mathcal{R}_2$ occurs in $\mathcal{W}^2_{k,t}$. The word

$$\mathcal{W}^2_{k,t} \;=\; F_t^k F_{t-1} \boxed{\;F_t^{k+1} F_{t-1} F_t^{k+2} F_{t-1}\; F_t^k F_{t-1} F_t^{k+1}\;}\; F_{t-1} F_t^{k+2} F_{t-1}$$

contains a fragment equal to

$$F_t^{k+1} F_{t-1} F_t\; F_t^{k+1} F_{t-1} F_t\; F_t^{k-1} F_{t-1} F_t^{k+1}$$

which itself has a fragment equal to

$$(F_t^{k+1} F_{t-1} F_t)\; (F_t^{k+1} F_{t-1} F_t)\; F_t^{k-1} F_{t-1} F_t$$

which is equal to $\mathbf{ex}((F_t^{k+1} F_{t-1} F_t)^2\, F_t^k F_{t-1})$, since $F_{t-1} F_t = \mathbf{ex}(F_t F_{t-1})$. Now, observe that $\mathcal{R}_2 = (F_t^{k+1} F_{t-1} F_t)^2\, F_t^k$ is a prefix of this fragment. Thus $\mathcal{W}^2_{k,t}$ contains $|F_t^k| + 1 = \alpha + 1$ distinct squares which are rotations of $(F_t^{k+1} F_{t-1} F_t)^2$.

Since $F_t^k F_{t-1}$ and $F_t^{k+2} F_{t-1}$ have different lengths, all considered square fragments are distinct and this concludes the proof. ◀

▶ **Lemma 10.** *If $k \leq |F_t|$ then $\mathcal{W}_{k,t}$ contains at least $\alpha - \mathcal{O}(|F_t|)$ distinct squares which are rotations of $(F_t^j F_{t-1})^2$ for $0 < j < k$.*

Proof. The word $\mathcal{W}_{k,t}$ contains the fragment $F_t^{k+1}F_{t-1}F_t^{k+2}$; see Figure 4. For each $j \le k$ it contains the fragment

$$F_t^j F_{t-1} F_t^{j+2} = \boxed{F_t^j F_{t-1} F_t^j F_{t-1}\ F_{t-2}F_{t-1}}\, F_{t-2} = F_t^j F_{t-1} F_t^j F_{t-1}\, \mathbf{ex}(F_{t-1}F_{t-2})\, F_{t-2},$$

which contains the fragment $(F_t^j F_{t-1})^2\ \mathbf{cut}(F_t)$.

By Observations 6 and 7, each such fragment, $(F_t^j F_{t-1})^2\ \mathbf{cut}(F_t)$, contains $|\,\mathbf{cut}(F_t)|+1 = |F_t| - 1$ distinct squares that are rotations of $(F_t^j F_{t-1})^2$.

Summing over all $0 < j < k$, we obtain at least $(|F_t| - 1) \cdot (k - 1) = \alpha - \mathcal{O}(|F_t|)$ distinct squares of the claimed form. ◀

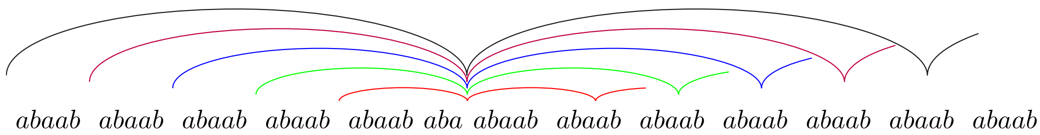

Figure 4 Illustration of Lemma 10. For $k = 4$, $t = 4$, let $A = F_t = abaab$, $B = F_{t-1} = aba$, and $C = \mathbf{cut}(F_t) = aba$. Then $\mathcal{W}_{k,t}$ contains the fragment $F_t^k F_{t-1} F_t^{k+2}$, which in turn contains the fragments $(A^j B)^2 C$, for $1 \le j \le k$. Each of these fragments implies $|C| + 1 = |F_t| - 1$ squares. Altogether (including the case $j = 0$), they contain $(|F_t| - 1) \cdot k + |F_{t-1}|$ squares, but Lemma 10 excludes the squares generated for $j = 0$ and $j = k$ to avoid double counting.

The fact that F_t^{k+2} occurs in $\mathcal{W}_{k,t}$ and Observations 6 and 7 imply the following:

▶ **Lemma 11.** *$\mathcal{W}_{k,t}$ contains at least $\frac{1}{2}\alpha$ distinct squares that are rotations of the words in $\{F_t^{2j} \mid 1 \le j \le \frac{k+2}{2}\}$.*

We are now ready to prove the main result of this section.

▶ **Theorem 12.** *There are infinitely many integers n for which $\mathsf{CS}(n) \ge 1.5n - \mathcal{O}(\sqrt{n})$.*

Proof. Let us set $k := |F_t|$. Then $\alpha = k^2$ and $|F_t| = \mathcal{O}(\sqrt{n})$. From Equation (1), the length of $\mathcal{W}_{k,t}$ is $n = 3\alpha + \mathcal{O}(\sqrt{n})$ and hence $\alpha = \frac{1}{3}n - \mathcal{O}(\sqrt{n})$.

By Lemmas 8–11, the word $\mathcal{W}_{k,t}^2$ contains at least $4.5\alpha - \Theta(|F_t|)$ squares (of length at most n). They are all distinct. Indeed, the squares from Lemmas 8–10 are rotations of $(F_t^j F_{t-1})^2$, for different $j \in [1\mathinner{.\,.} k+2]$, so they have different lengths for different j. The squares from Lemma 11 are of the form $(F_t^j)^2$ for $j \le k$, and we have $|F_t^{j-1}F_{t-1}| < |F_t^j| < |F_t^j F_{t-1}|$.

Therefore, the circular word $\mathcal{W}_{k,t}$ contains

$$4.5 \cdot (\tfrac{1}{3}n - \mathcal{O}(\sqrt{n})) \;=\; 1.5n - \mathcal{O}(\sqrt{n})$$

distinct squares, as required. ◀

4 Upper Bound

Let us fix a word w of length n. We denote $\widetilde{w} = ww$. Our aim is to show that $|\mathbf{S}_n(\widetilde{w})| \le \lceil 1.8n \rceil$.

High-level idea of the proof. If there exists a fragment (a δ-window) of w^∞ of (small) length δ that contains no starting position of a long square, then it is easy to show that there are at most $2n - \delta$ distinct squares in the circular word w. If no such fragment exists, we call w *δ-dense*. Hence, we reduce the problem to showing that primitive δ-dense words do not exist – a simple upper bound of $1.5n$ can be shown for non-primitive words. We view a square uu as a "transporter" which shifts long fragments of w by $|u|$ positions (from the first copy of u to the second one). By choosing a long non-periodic fragment (called a *sample*) s, we consider only those distances $|u|$ that correspond to the distances between occurrences of the sample in ww which start in the first copy of w. The number of these occurrences is very small (bounded by a constant). Then, using samples, we show that if w is primitive, then w is not δ-dense.

▶ **Observation 13.** *If w' is a rotation of w, then $\mathbf{S}_n(ww) = \mathbf{S}_n(w'w')$.*

▶ **Lemma 14.** *If w' is a rotation of w that is not primitive, then $|\mathbf{S}_n(\widetilde{w})| \le 1.5n$.*

Proof. Suppose $w' = u^k$ for some integer $k \ge 2$. Then $\mathbf{S}_n(\widetilde{w}) = \mathbf{S}_n(w'w') = \mathbf{S}_n(uw')$. Consequently, due to [3], we have $|\mathbf{S}_n(\widetilde{w})| = |\mathbf{S}_n(uw')| \le |u| + |w| = (1 + \frac{1}{k})n$. Since $k \ge 2$, it follows that $|\mathbf{S}_n(\widetilde{w})| \le 1.5n$. ◀

▶ **Corollary 15.** *If $\widetilde{w}$ contains a square of length exactly n, then $|\mathbf{S}_n(\widetilde{w})| \le 1.5n$.*

By the above corollary, we may henceforth assume that w is primitive and $\widetilde{w}$ does not contain any square of length n.

▶ Remark 16. Circular words that are themselves squares can still contain many squares. It was shown in [1] that there exists an infinite family of words w_k, each of which is a square, such that $\mathbf{S}_n(w_k w_k) \ge 1.25|w_k| - o(|w_k|)$.

▶ **Definition 17.** *Let w be a word of length n. A square fragment uu starting at position j in w^∞ is called a δ-**square** fragment with respect to an interval $[i \mathinner{.\,.} i+\delta)$ if*

$$i \le j < i + \delta \text{ and } n - (j - i) < |uu| < n. \tag{2}$$

Observe that if $j < \delta$ is the last occurrence in ww of a square uu, then uu is a δ-square with respect to the interval $I = [0 \mathinner{.\,.} \delta)$.

▶ **Definition 18.** *A word w is called δ-**dense** if, for every length-δ fragment I of w^∞, there exists a δ-square fragment with respect to I. (See Figure 5.)*

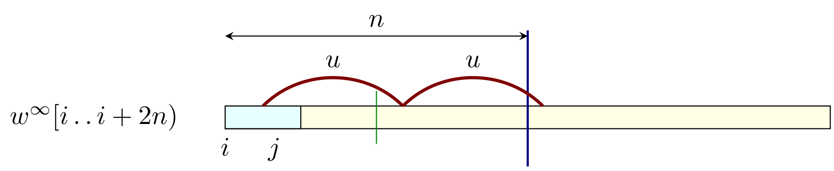

Figure 5 Illustration of the notion of δ-density. The shown word is a prefix of $w^\infty[i \mathinner{.\,.} \infty)$. Here, uu is a δ-square fragment with respect to the interval $I = [i \mathinner{.\,.} i+\delta)$ (shown in blue): there exists a position j in the interval I where a square uu starts and satisfies $|uu| + (j - i) > n$. The word w is δ-dense if such a square uu exists for each interval $[i \mathinner{.\,.} i+\delta)$ with $i \in \mathbb{Z}_+$.

We first consider the easy case when $\widetilde{w}$ is not δ-dense.

▶ **Lemma 19** (Upper bound for words that are not δ-dense)**.** *If $\widetilde{w}$ is not δ-dense, where $0 < \delta \le \frac{1}{2}n$, then $|\mathbf{S}_n(\widetilde{w})| \le 2n - \delta$.*

Proof. By Corollary 15, if $\widetilde{w}$ contains a square fragment of length exactly n, $|\mathbf{S}_n(\widetilde{w})| \le 1.5n$. Henceforth we assume that $\widetilde{w}$ does not have such a fragment. Since $\widetilde{w}$ is not δ-dense, by the definition, there exists an interval $I = [i \mathinner{.\,.} i+\delta)$ in w^∞ such that no δ-square fragment exists with respect to I. That is, for every square fragment uu starting at some position $j \in I$, $|uu| < n - (j-i)$. Equivalently, the longest square starting at any position $q \in I$ has length at most $n-(q-i)$. After a rotation (cf. Observation 13), we may assume that $i = 0$, so that $I = [0 \mathinner{.\,.} \delta)$. Now, let us consider the suffix $\widetilde{w}'$ of $\widetilde{w}$ of length $2n-\delta$. Every square fragment of $\widetilde{w}$ occurs entirely within $\widetilde{w}'$. Therefore, all distinct squares appear in a fragment of length $2n-\delta$. The combination of this fact and the result of [3] implies that $|\mathbf{S}_n(\widetilde{w})| \le 2n-\delta$. ◀

Our aim is to show that no δ-dense words exist when w is primitive and δ is sufficiently small.

The next lemma states that each fragment of $\widetilde{w}$ of length $\frac{1}{2}n-\delta$ has a copy to its right at a well-specified distance d. See Figure 6 for an illustration.

▶ **Lemma 20** (Transporting Lemma). *Assume w is δ-dense with positive integer $\delta \le \frac{1}{2}n$. Let F be a fragment of length at most $\frac{1}{2}n-\delta$ that occurs at some position $i \ge n$ in w^∞. Then, F also occurs in w^∞ at position $i+\ell$ for some $\ell \in (\frac{1}{2}(n-\delta) \mathinner{.\,.} \frac{1}{2}n)$.*

Proof. Since w is δ-dense, there exists a position j in $(i-\delta \mathinner{.\,.} i]$ where a δ-square fragment $x = uu$ occurs in w^∞. By the definition of a δ-square, its total length satisfies $n-(j-(i-\delta)) < |uu| < n$. It follows that the half-length $|u|$ satisfies $\frac{1}{2}(n-\delta) < |u| < \frac{1}{2}n$.

By Equation (2), the square $x = uu$ ends at a position $z \ge i-\delta+n$, so the first half u of the square ends at a position $z' \ge i-\delta+\frac{1}{2}n \ge i+|F|$. Therefore the fragment F is completely contained within u. The second half of x also contains a copy of F shifted by a distance $|u|$. Therefore, F also occurs at position $i+|u|$. ◀

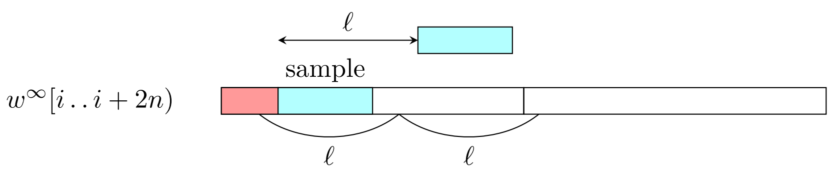

Figure 6 Illustration of the Transporting Lemma (Lemma 20). The sample is shown in blue, and the δ-length interval with respect to which the δ-square occurs is shown in red.

We use the following fact from [1], whose proof relies on the periodicity lemma.

▶ **Fact 21** ([1, Lemma 4]). *Let w be a word and let a and b be letters. If both aw and wb are periodic, then they have the same period.*

▶ **Corollary 22.** *If w is primitive and $\ell \in [1 \mathinner{.\,.} n]$, then $\widetilde{w}$ contains a non-periodic fragment* s *of length ℓ. We refer to this as an ℓ-*sample.

Proof. Let us assume that $\widetilde{w}$ does not contain a non-periodic fragment of length ℓ. Then, by Fact 21, all length-ℓ fragments of ww are periodic with the same period, i.e., the whole $\widetilde{w}$ has a period at most $\frac{1}{2}\ell$. By the periodicity lemma (Lemma 1), $\widetilde{w} = ww$ has a period p such that $p \le \frac{1}{2}n$ and p divides n. Hence, w is non-primitive. ◀

We show that, for every primitive word w, $\widetilde{w}$ is not δ-dense for a certain parameter $\delta \le 0.5n$. Together with Lemmas 14 and 19, this implies that $|\mathbf{S}_n(\widetilde{w})| \le 2n-\delta$.

Simple Upper Bound

First we give a simple and intuitive proof for $\delta = \lfloor \frac{1}{6}n \rfloor$.

▶ **Lemma 23.** *If w is primitive and $n = |w| \geq 6$, then $\widetilde{w}$ is not $\lfloor \frac{1}{6}n \rfloor$-dense.*

Proof. We prove this by contradiction. Assume that $\widetilde{w}$ is $\lfloor \frac{1}{6}n \rfloor$-dense. Let $\mathbf{s}$ be a $\lfloor \frac{1}{3}n \rfloor$-sample that is non-periodic (as guaranteed by Corollary 22). After a suitable rotation of ww, we can assume that $\mathbf{s}$ occurs at positions 0 and n. Due to Lemma 20, $\mathbf{s}$ also occurs at position $i = d + d'$, where $d, d' \in (\frac{1}{2} \cdot \frac{5}{6}n \,..\, \frac{1}{2}n)$; see Figure 7. Consequently, $\mathbf{s}$ occurs at distinct positions n and i that are at distance at most $\lfloor \frac{1}{6}n \rfloor \leq |\mathbf{s}|/2$. Hence, $\mathbf{s}$ is periodic, a contradiction. ◀

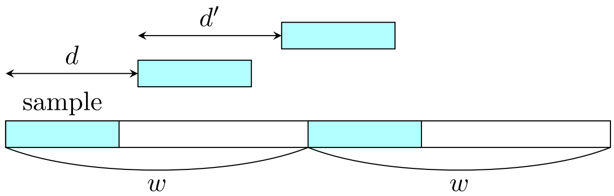

Figure 7 Illustration of the proof of Lemma 23.

Lemmas 14, 19, and 23 imply that $\mathsf{CS}(n) \leq 2n - \lfloor \frac{1}{6}n \rfloor = \lceil \frac{11}{6}n \rceil$, after a trivial verification of the cases when $n < 6$.

Stronger Upper Bound

Next, we strengthen the upper bound using a more refined argument.

▶ **Lemma 24.** *If w is primitive, with $n = |w| \geq 5$, then $\widetilde{w}$ is not $\lfloor 0.2n \rfloor$-dense.*

Proof. We prove this by contradiction. Assume that $\widetilde{w}$ is δ-dense for $\delta = \lfloor 0.2n \rfloor$. Let $\mathbf{s}$ be a non-periodic sample of length $\ell = \lfloor 0.3n \rfloor$, (as guaranteed by Corollary 22, since $\ell \geq 1$ for $n \geq 4$). Let $\mathbf{A}$ denote the set of starting positions of $\mathbf{s}$ in ww.

After a suitable rotation of ww, we may assume that positions 0 and n belong to $\mathbf{A}$. Let $\mathbf{A}' = \mathbf{A} \cap [0\,..\,n)$, $\mathbf{A}' = \{a_0, \ldots, a_{|\mathbf{A}'|-1}\}$ with $a_0 < \cdots < a_{|\mathbf{A}'|-1}$ and define $\ell_i = a_{(i+1) \bmod |\mathbf{A}'|} - a_i$ for $i \in [0\,..\,|\mathbf{A}'|)$.

For each occurrence at position a_i, let d_i be the shift distance given by the Transporting-Lemma (Lemma 20). This lemma applies since $\ell + \delta \leq \frac{1}{2}n$, and it ensures that $d_i \in (\frac{1}{2} \cdot \lceil 0.8n \rceil \,..\, 0.5n)$.

Then, there exists some $i' \in [0\,..\,|\mathbf{A}'|)$ such that $(a_i + d_i) \bmod n = a_{i'}$. We denote $\mathsf{succ}(i) := i'$, we define $\mathsf{succ}(i) := \mathsf{succ}(i \bmod |\mathbf{A}'|)$.

▷ Claim 25. For each $i \in [0\,..\,|\mathbf{A}'|)$, $\mathsf{succ}(\mathsf{succ}(i)) = (i-1) \bmod n$ and $\ell_i \in (0.15n\,..\,\lfloor 0.2n \rfloor)$.

Proof. Let $\mathsf{succ}(i) = i'$ and $\mathsf{succ}(i') = i''$. The sum of two consecutive shift distances $d_i + d_{i'} \in (\lceil 0.8n \rceil\,..\,n)$. Hence, $a_i + d_i + d_{i'} < a_i + n$, and there is no occurrence of $\mathbf{s}$ in w^∞ between $a_i + d_i + d_{i'}$ and $a_i + n$; otherwise, $\mathbf{s}$ would be periodic. Therefore, $\mathsf{succ}(\mathsf{succ}(i)) = (i-1) \bmod n$. Moreover,

$$\ell_{(i-1) \bmod n} \leq n - (\lceil 0.8n \rceil + 1) < \lfloor 0.2n \rfloor, \text{as claimed.}$$ ◁

▷ Claim 26. For each $i \in [0\,..\,|\mathbf{A}'|)$, $\mathsf{succ}(i) = (i+3) \bmod n$.

Proof. Each shift d_i from the Transporting-Lemma lies in $(0.4n\,..\,0.5n)$ (as $\frac{1}{2}\cdot\lceil 0.8n\rceil \geq \frac{1}{2}\cdot 0.8n$), while each ℓ_i lies in interval $(0.15n\,..\,\lfloor 0.2n\rfloor)$ by Claim 25.

Since $2\cdot\lfloor 0.2n\rfloor \leq 0.4n < d_i < 0.5n < 4\cdot 0.15n$, each shift spans exactly three intervals. Hence, $\mathsf{succ}(i) = (i+3) \bmod n$. ◁

The interval bounds from Claim 25 imply that $|\mathbf{A}'| = 6$. But then, by Claim 26, $\mathsf{succ}(\mathsf{succ}(0)) \neq |\mathbf{A}'|-1$, contradicting Claim 25. This completes the proof. ◀

By Lemmas 14, 19, and 24, we obtain the upper bound $\mathsf{CS}(n) \leq 2n - \lfloor 0.2n\rfloor = \lceil 1.8n\rceil$, after a trivial verification of the cases when $n < 5$.

▶ **Theorem 27.** $\mathsf{CS}(n) \leq \lceil 1.8n\rceil$.

▶ Remark 28. If we used the sample size $\lfloor 0.25n\rfloor$ as in [1], we would obtain a weaker upper bound of $\lceil \frac{13}{7}n\rceil$.

5 Limitations of our Approach

It appears that by taking larger values of δ, one could possibly strengthen Lemma 24, showing that every primitive word is not δ-dense, and thereby obtaining a stronger upper bound of $2n-\delta$. However, we show that for $\delta = \frac{1}{3}n+4$ this approach no longer works.

We say that a square uu *kills* a position i in w^∞ if uu is a δ-square fragment with respect to the interval $[i\,..\,i+\delta)$.

▶ **Observation 29.**

(a) *A square uu such that $n-(\delta-1) < |uu| < n$ starting at position j in w^∞ kills all positions contained modulo n in the interval $[j-(\delta-1)\,..\,j+|uu|-n)$.*

(b) *For a periodic fragment R with $n-(\delta-1) < 2\cdot\mathrm{per}(R) < n$ starting at position j in w^∞, all positions contained modulo n in the interval*

$$I_R = [j-(\delta-1)\,..\,j+|R|-n-1] \tag{3}$$

are killed by the $|R|-2\cdot\mathrm{per}(R)+1$ squares of length $2\cdot\mathrm{per}(R)$ that are induced by R; see Figure 8.

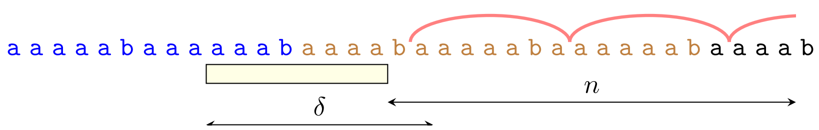

Figure 8 Let $w = a^4ba^5ba^6b$ (fragment shown in brown) and $\delta = 10$. The figure shows a word $(w')^2$ for a rotation w' of w. The square fragments in w^∞ induced by the periodic fragment R (shown in red) together kill all positions contained modulo n in the interval I_R indicated by the yellow rectangle.

▶ **Fact 30.** *There exist arbitrarily long primitive words w that are δ-dense for $\delta = \frac{1}{3}|w|+4$.*

Proof. Let us take $\mathcal{W}_{k,1} = a^k\, b\, a^{k+1}\, b\, a^{k+2}\, b$ of length $n = 3k+6$ and $\delta = k+6 = \frac{1}{3}n+4$. Using Equation (3), it is easy to see that each position of $\mathcal{W}_{k,1}$ is killed by one of the following three periodic fragments:

- The fragment $R_1 = (a^k ba)^2 a^k$ occurring at position 0 kills positions in the interval

$$[0 - (k+5) \mathinner{.\,.} 0 + (3k+4) - n - 1] \equiv_{\bmod n} [2k+1 \mathinner{.\,.} n-3].$$

- The fragment $R_2 = (a^{k+1} ba)^2 a^{k-1}$ occurring at position $k+1$ kills positions in the interval

$$[(k+1) - (k+5) \mathinner{.\,.} (k+1) + (3k+5) - n - 1] \equiv_{\bmod n} [n-4 \mathinner{.\,.} n-1] \cup [0 \mathinner{.\,.} k-1];$$

cf. Figure 8.

- The fragment $R_3 = (a^k b)^2 a^k$ occurring at position $2k+5$ kills positions in the interval

$$[(2k+5) - (k+5) \mathinner{.\,.} (2k+5) + (3k+2) - n - 1] \equiv_{\bmod n} [k \mathinner{.\,.} 2k].$$

The union of the intervals $[0 \mathinner{.\,.} k-1]$, $[k \mathinner{.\,.} 2k]$, $[2k+1 \mathinner{.\,.} n-3]$, and $[n-4 \mathinner{.\,.} n-1]$ covers all positions in $\mathcal{W}_{k,1}$. This implies that every position is killed by a suitable δ-square fragment. Hence, $\mathcal{W}_{k,1}$ is δ-dense. ◀

6 Final Remarks

We have shown that $1.5n - \mathcal{O}(\sqrt{n}) \leq \mathsf{CS}(n) \leq \lceil 1.8n \rceil$, where the lower bound holds for infinitely many n. We conjecture that the lower bound is tight, that is, that $\mathsf{CS}(n) \leq 1.5n$.

Tight bounds related to repetitions and symmetries are usually non-trivial. This was, for instance, the case for bounds on standard squares (see [7, 9, 12, 5, 18, 3]) and powers ([11, 13, 14]), runs (the "runs theorem"; see [2] and references therein), and palindromes in circular words ([17]). For example, the abstract of [17] states: "*In this paper we show, with a very complicated proof, …*").

Possibly, the elegant graph-theoretic proof of the recent upper bound for standard squares in [3] can be adapted to establish an $1.5n$ upper bound for distinct squares in circular words.

References

1 Mika Amit and Pawel Gawrychowski. Distinct squares in circular words. In *String Processing and Information Retrieval - 24th International Symposium, SPIRE 2017*, volume 10508 of *Lecture Notes in Computer Science*, pages 27–37. Springer, 2017. `doi:10.1007/978-3-319-67428-5_3`.

2 Hideo Bannai, Tomohiro I, Shunsuke Inenaga, Yuto Nakashima, Masayuki Takeda, and Kazuya Tsuruta. The "runs" theorem. *SIAM Journal on Computing*, 46(5):1501–1514, 2017. `doi:10.1137/15M1011032`.

3 Srecko Brlek and Shuo Li. On the number of squares in a finite word. *Combinatorial Theory*, 5(1), 2025. `doi:10.5070/C65165014`.

4 James D. Currie. There are ternary circular square-free words of length n for n >= 18. *Electronic Journal of Combinatorics*, 9(1), 2002. `doi:10.37236/1671`.

5 Antoine Deza, Frantisek Franek, and Adrien Thierry. How many double squares can a string contain? *Discrete Applied Mathematics*, 180:52–69, 2015. `doi:10.1016/J.DAM.2014.08.016`.

6 Nathan J. Fine and Herbert S. Wilf. Uniqueness theorems for periodic functions. *Proceedings of the American Mathematical Society*, 16(1):109–114, 1965. `doi:10.2307/2034009`.

7 Aviezri S. Fraenkel and Jamie Simpson. How many squares can a string contain? *Journal of Combinatorial Theory, Series A*, 82(1):112–120, 1998. `doi:10.1006/JCTA.1997.2843`.

8 Aviezri S. Fraenkel and Jamie Simpson. The exact number of squares in Fibonacci words. *Theoretical Computer Science*, 218(1):95–106, 1999. `doi:10.1016/S0304-3975(98)00252-7`.

9 Lucian Ilie. A note on the number of squares in a word. *Theoretical Computer Science*, 380(3):373–376, 2007. `doi:10.1016/J.TCS.2007.03.025`.

10 Costas S. Iliopoulos, Dennis W. G. Moore, and William F. Smyth. A characterization of the squares in a Fibonacci string. *Theoretical Computer Science*, 172(1-2):281–291, 1997. doi:10.1016/S0304-3975(96)00141-7.

11 Marcin Kubica, Jakub Radoszewski, Wojciech Rytter, and Tomasz Waleń. On the maximum number of cubic subwords in a word. *European Journal of Combinatorics*, 34(1):27–37, 2013. doi:10.1016/J.EJC.2012.07.012.

12 Nguyen Huong Lam. On the number of squares in a string. *AdvOL-Report*, 2, 2013.

13 Shuo Li. On the number of k-powers in a finite word. *Advances in Applied Mathematics*, 139:102371, 2022. doi:10.1016/J.AAM.2022.102371.

14 Shuo Li, Jakub Pachocki, and Jakub Radoszewski. A note on the maximum number of k-powers in a finite word. *Electronic Journal of Combinatorics*, 31(3), 2024. doi:10.37236/11270.

15 Gwénaël Richomme, Kalle Saari, and Luca Q. Zamboni. Standard factors of Sturmian words. *RAIRO - Theoretical Informatics and Applications*, 44(1):159–174, 2010. doi:10.1051/ITA/2010011.

16 Arseny M. Shur. On ternary square-free circular words. *Electronic Journal of Combinatorics*, 17(1), 2010. doi:10.37236/412.

17 Jamie Simpson. Palindromes in circular words. *Theoretical Computer Science*, 550:66–78, 2014. doi:10.1016/J.TCS.2014.07.012.

18 Adrien Thierry. A proof that a word of length n has less than 1.5n distinct squares, 2020. arXiv:2001.02996.

19 Axel Thue. Über unendliche Zeichenreihen. *Norske Videnskabers Selskabs Skrifter Mat.-Nat. Kl.*, 7:1–22, 1906.