Distinct Squares in Circular Words

Mika Amit[1,2] and Paweł Gawrychowski[2,3](✉)

[1] IBM Research, Haifa, Israel
[2] University of Haifa, Haifa, Israel
[3] Institute of Computer Science, University of Wrocław, Wrocław, Poland
gawry@cs.uni.wroc.pl

Abstract. A circular word, or a necklace, is an equivalence class under conjugation of a word. A fundamental question concerning regularities in standard words is bounding the number of distinct squares in a word of length n. The famous conjecture attributed to Fraenkel and Simpson is that there are at most n such distinct squares, yet the best known upper bound is $1.84n$ by Deza et al. [Discr. Appl. Math. 180, 52–69 (2015)]. We consider a natural generalization of this question to circular words: how many distinct squares can there be in all cyclic rotations of a word of length n? We prove an upper bound of $3.14n$. This is complemented with an infinite family of words implying a lower bound of $1.25n$.

Keywords: Squares conjecture · Circular words

1 Introduction

Combinatorics on words is mostly concerned with regularities in words. The most basic example of such a regularity is a square, that is, a substring of the form uu. We might either want to create words with no such substrings, called square-free, or show that there cannot be too many distinct squares for an arbitrary word of length n. Fraenkel and Simpson proved that $2n$ is an upper bound on the number of distinct squares contained in a word of length n, and also constructed an infinite family of words of length n containing $n-\Theta(\sqrt{n})$ distinct squares [12]. Their upper bound uses a combinatorial lemma of Crochemore and Rytter [6], called the Three Squares Lemma. Later, Ilie provided a short and self-contained argument [16]. The Three Squares Lemma is concerned with the rightmost occurrence of every distinct square, and says that, for any position in the word, there do not exist three such rightmost occurrences starting at that position (hence the name of the lemma). It is widely believed that the example given by Frankel and Simpson is the worst possible, and the right bound is n instead of $2n$. The best known upper bound was $2n - \Theta(\log n)$ [17] until recently Deza, Franek and Thierry improved the upper bound to $11/6n$ through a somewhat involved argument [9]. All these bounds are based on the idea of looking at three rightmost occurrences of squares starting at the same position. It is known that two such occurrence already imply a certain periodic structure [2,10,13,18,23], and that it is enough to consider binary words [20].

G. Fici et al. (Eds.): SPIRE 2017, LNCS 10508, pp. 27–37, 2017.
DOI: 10.1007/978-3-319-67428-5_3

Regularities are commonly considered in more general contexts than standard words, such as partial words [1] or trees [5,14]. Another natural generalization of standard words, motivated by the circular structure of some biological data, are circular words (also known as necklaces). A circular word (w) is defined as an equivalence class under conjugation of a word w, that is, it corresponds to all possible rotations of w. Both algorithmic [3,4,15] and combinatorial aspects of such words have been studied. The latter are mostly motivated by an old result of Thue [25], who showed that there is an infinite square-free word over $\{0, 1, 2\}$. This started a long line of research of pattern avoidance. Currie and Fitzpatrick [8] generalized this to circular words, and then Currie [7] showed that for any $n \geq 18$ there exists a circular square-free word of length n (see also a later proof by Shur [22]). Recently, Simpson [24] considered bounding the number of distinct palindromes in a circular word of length n. It is well-known (and easy to prove) that the number of distinct palindromes in a standard word of length n is at most n. Interestingly, this increases to $5/3n$ for circular words. Also equations on circular words have been studied [21].

We consider the following question: how many distinct squares can there be in a circular word of length n? Note that due to how we have defined a circular word, we are interested in squares of length at most n. Recall that the $2n$ bound of Fraenkel and Simpson [12] is based on the notion of rightmost occurrences. The improved $11/6n$ bound of Deza et al. [9] is also based on this concept. For a circular word, it is not clear what the rightmost occurrence might mean, and indeed the proofs seem to completely break. Of course, to bound the number of distinct squares in a circular word w of length n, one can simply bound the number of distinct squares in a word ww of length $2n$, thus immediately obtaining an upper bound of $4n$ (by invoking the simple proof of Ilie [16]) or $3.67n$ (by invoking the more involved proof of Deza et al. [9]). This, however, completely disregards the cyclic nature of the problem.

We start with exhibiting an infinite family of circular words of length n containing $1.25n - \Theta(1)$ distinct squares. Therefore, it appears that the structure of distinct squares in circular words is more complex than in standard words. We then continue with a simple and self-contained upper bound of $3.75n$ on the number of distinct squares in a circular word of length n. Then, by invoking some of the machinery used by Deza et al. [9], we improve this to $3.14n$.

2 Preliminaries

Let $|w|$ denote the length of a string w, $w[i]$ is the i-th character of w, and $w[i..j]$ is a shortcut for $w[i]w[i+1]\ldots w[j]$. A natural number p is a period of w iff $w[i] = w[i+p]$ for every $i = 1, 2, \ldots, |w| - p$. The smallest such p is called the period of w. We say that w is periodic if its period is at most $|w|/2$, otherwise w is aperiodic. The well-known periodicity lemma says that if p and q are both periods of w and furthermore $p + q \leq |w| + \gcd(p, q)$ then $\gcd(p, q)$ is also a period of w [11].

$w^{(i)}$ denotes the cyclic rotation of w by i, that is, $w[i..|w|]w[1..(i-1)]$. A circular word (w) is an equivalence class under conjugation of w, that is, all cyclic rotations $w^{(i)}$. A word uu is called a square, and we say that it occurs in (w) if it occurs in $w^{(i)}$ for some i. We are interested in bounding the number of distinct squares occurring in a circular word of length n.

3 Lower Bound

We define an infinite family of words $f_k = \mathtt{a}(\mathtt{ba})^{k+1}\mathtt{a}(\mathtt{ba})^{k+2}\mathtt{a}(\mathtt{ba})^{k+1}\mathtt{a}(\mathtt{ba})^{k+2}$. See Fig. 1 for an example. Observe that $|f_k| = 8k + 16$. We claim that cyclic rotations of f_k contain many distinct squares.

$\boldsymbol{k}$	$\boldsymbol{f_k}$	#squares / $\lvert \boldsymbol{f_k} \rvert$
1	ababaabababaababaabababa	25/24
2	abababaababababaabababaababababa	36/32
3	ababababaabababababaababababaabababababa	45/40
4	abababababaababababababaabababababaababababababa	56/48
5	ababababababaabababababababaababababababaabababababababa	65/56

Fig. 1. The number of distinct squares in f_k, for $k = 1, 2, 3, 4, 5$.

Lemma 1. *For any $k \geq 0$, the circular word (f_k) contains $10k + 16 - (k \bmod 2)$ distinct squares.*

Proof. To count distinct squares uu occurring in (f_k), we consider a few disjoint cases. We first count uu such that aa occurs at most once inside:

1. Any uu such that $\mathtt{aa}$ does not occur inside must be be fully contained in an occurrence of $\mathtt{a}(\mathtt{ba})^{k+2}$ or $\mathtt{a}(\mathtt{ba})^{k+1}$ in f_k. Thus, to count such uu we only have to find all distinct squares in $\mathtt{a}(\mathtt{ba})^{k+2}$. For any $i = 1, 2, \ldots, \lfloor (k+2)/2 \rfloor$, $(\mathtt{ab})^i(\mathtt{ab})^i$ and $(\mathtt{ba})^i(\mathtt{ba})^i$ appear there, and it can be seen that there are no other squares. Thus, the number of such uu is exactly $2\lfloor (k+2)/2 \rfloor$.
2. Any uu such that $\mathtt{aa}$ occurs exactly once inside must have the property that u starts and ends with $\mathtt{a}$. It follows that such uu must be be fully contained in an occurrence of $\mathtt{a}(\mathtt{ba})^{k+1}\mathtt{a}(\mathtt{ba})^{k+1}$ in f_k. For any $i = 0, 1, \ldots, k+1$, $\mathtt{a}(\mathtt{ba})^i\mathtt{a}(\mathtt{ba})^i$ appears there, and it can be seen that there are no other squares containing exactly one occurrence of $\mathtt{aa}$, so there are exactly $k + 2$ such uu.

Then we count uu such that $\mathtt{aa}$ occurs exactly twice inside. Then, $\mathtt{aa}$ must occur once in u and furthermore, by analyzing the distances between the occurrences of $\mathtt{aa}$ in f_k, we obtain that $|u| = 2k + 5$ or $|u| = 2k + 3$. We analyze these two possibilities:

1. If $|u| = 2k + 3$ then uu appears in an occurrence of $(\mathtt{ba})^k\mathtt{baa}(\mathtt{ba})^k\mathtt{baa}(\mathtt{ba})^k\mathtt{b}$ in f_k. There are $2k + 2$ such uu.
2. If $|u| = 2k + 5$ then uu appears in an occurrence of $\mathtt{a}(\mathtt{ba})^k\mathtt{baaba}(\mathtt{ba})^k$ $\mathtt{baaba}(\mathtt{ba})^k$ in f_k. There are $2k + 2$ such uu.

Finally, we count uu such that aa occurs at least three times inside. By analyzing the distances between the occurrences of aa in f_k, we obtain that in such case $|u| = 4k+8$, so $|uu| = |f_k|$. We claim that there are exactly $|f_k|/2 = 4k+8$ such uu. To prove this, write $f_k = x_k x_k$ with $x_k = \mathtt{a}(\mathtt{ba})^{k+1}\mathtt{a}(\mathtt{ba})^{k+2}$. x_k cannot be represented as a nontrivial power y^p with $p \geq 2$, because aa occurs only once inside x_k, so it would mean that y starts and ends with a, but then $p = 2$ is not possible due to $|\mathtt{a}(\mathtt{ba})^{k+1}| \neq |\mathtt{a}(\mathtt{ba})^{k+2}|$, and $p \geq 3$ would generate another occurrence of a. Clearly, every cyclic shift of f_k is a square occurring in (f_k), because a cyclic shift of a square is still a square. It remains to count distinct cyclic shifts of f_k. Assume that two of these shifts are equal, that is, $(f_k)^{(i)} = (f_k)^{(j)}$ for some $0 \leq i < j < |f_k|$, so $x_k = (x_k)^{(j-i)}$. Then $\gcd(|x_k|, j-i)$ is a period of x_k. But x_k is not a nontrivial power, so $j - i = 0 \bmod |x_k|$. Consequently, every $i = 0, 1, \ldots, |x_k| - 1$ generates a distinct square.

All in all, the number of distinct squares occurring in (f_k) is

$$k + 2 + 2\lfloor (k+2)/2 \rfloor + 2(2k+2) + 4k + 8 = 9k + 16 + 2\lfloor k/2 \rfloor$$

or, in other words, $10k + 16 - (k \bmod 2)$. □

By Lemma 1, for any n_0 there exists a circular word of length $n \geq n_0$ containing at least $1.25n - \Theta(1)$ distinct squares.

4 Upper Bound

Our goal is to upper bound the number of distinct squares occurring in a circular word (w) of length n. Each such square occurs in ww, hence clearly there are at most $4n$ such distinct squares by plugging in the known bound on the number of distinct squares. However, we want a stronger bound.

Recall that the bound on the number of distinct squares is based on the notion of the rightmost occurrence. For every distinct square uu occurring in a word, we choose its rightmost occurrence. Then, we have the following property.

Lemma 2 ([12]). *For any position i, there are at most two rightmost occurrences starting at i.*

Consider the rightmost occurrences of distinct squares of length up to n in ww. We first analyze the rightmost occurrences starting at positions $1, 2, \ldots, \frac{1}{4}n$.

Lemma 3. *If $w[\frac{1}{4}n..\frac{1}{2}n]$ is aperiodic then every rightmost occurrence starting at position $i \in \{1, 2, \ldots, \frac{1}{4}n\}$ is of the same length.*

Proof. Assume otherwise, that is, $w[\frac{1}{4}n..\frac{1}{2}n]$ is aperiodic, but there are two rightmost occurrences uu and $u'u'$ starting at positions $i, i' \in \{1, 2, \ldots, \frac{1}{4}n\}$, respectively, in ww such that $|u| > |u'|$. Then, $i + 2|u| > n$ and $i' + 2|u'| > n$, as otherwise we could have found the same square in the second half of ww. Because $|u|, |u'| \leq \frac{1}{2}n$, this implies $i + |u| > \frac{1}{2}n$ and $i' + |u'| > \frac{1}{2}n$. So $w[\frac{1}{4}n..\frac{1}{2}n]$[1]

[1] Formally, we need to appropriately round both $\frac{1}{4}n$ and $\frac{1}{2}n$. We chose not to do so explicitly as to avoid cluttering the presentation.

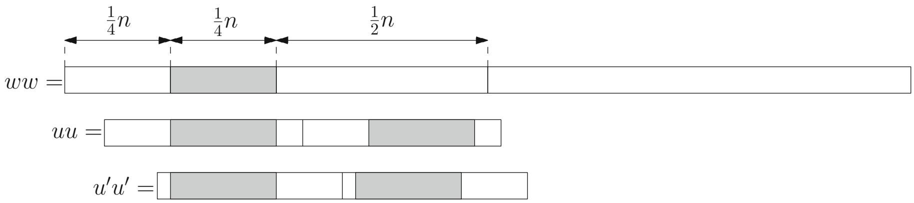

Fig. 2. Two rightmost occurrences of squares uu and $u'u'$ in ww.

is fully inside the first half of both uu and $u'u'$. But then it also appears starting at positions $\frac{1}{4}n + |u|$ and $\frac{1}{4}n + |u'|$, see Fig. 2. The distance between these two distinct (due to $|u| > |u'|$) occurrences is

$$(\frac{1}{4}n + |u|) - (\frac{1}{4}n + |u'|) = |u| - |u'|$$

We know that $|u| \leq \frac{1}{2}n$ and $|u'| > \frac{1}{2}n - i' \geq \frac{1}{2}n - \frac{1}{4}n = \frac{3}{8}n$. Thus, the distance is less than $\frac{1}{2}n - \frac{3}{8}n = \frac{1}{8}n$ and we conclude that the period of $w[\frac{1}{4}n..\frac{1}{2}n]$ is at most $\frac{1}{8}n$, which is a contradiction. □

By Lemm 3, assuming that $w[\frac{1}{4}n..\frac{1}{2}n]$ is aperiodic, for every $i = 1, 2, \ldots, \frac{1}{4}n$ there is at most one rightmost occurrence starting at i. For all the remaining i, there are at most two rightmost occurrences starting at i, making the total number of distinct squares at most $\frac{1}{4}n + 2(2n - \frac{1}{4}n) = 3\frac{3}{4}n$.

It might be the case that $w[\frac{1}{4}n..\frac{1}{2}n]$ is periodic. However, the number of distinct squares occurring in (w) is the same as the number of distinct squares occurring in any $(w^{(i)})$, so we are free to replace w with any of its cyclic shifts. We claim that if, for any $i = 0, 1, \ldots, n-1$, $w^{(i)}[\frac{1}{4}n..\frac{1}{2}n]$ is periodic, then the whole w is a nontrivial power y^p with $p \geq 8$. To show this, we need an auxiliary lemma that is a special case of Lemma 8.1.2 of [19]. We provide a proof for completeness.

Lemma 4. *For any word w and characters a, b, if both aw and wb are periodic then their periods are in fact equal.*

Proof. We assume that the period of aw is $p \leq |aw|/2$ and the period of wb is $q \leq |wb|/2$. Then p and q are both periods of w. By symmetry, we can assume that $p \geq q$. $p + q \leq (|aw| + |wb|)/2 = 1 + |w|$, so by the periodicity lemma $\gcd(p, q)$ is a period of w. We claim that $\gcd(p, q)$ is also a period of aw. To prove this, it is enough to show that $a = w[\gcd(p, q)]$. $\gcd(p, q)$ is a period of w and, for $n \geq 2$, $p \leq |w|$, so this is equivalent to showing that $a = w[p]$. But this holds due to p being a period of aw. Hence $\gcd(p, q)$ is a period of aw, but p is the period of aw and $p \geq q$, therefore $p = q$. □

We observe that the substrings $w^{(i)}[\frac{1}{4}n..\frac{1}{2}n]$ correspond to all substrings of length $\frac{1}{4}n$ of ww. By Lemma 4, if every substring of length $\frac{1}{4}n$ of ww is periodic, then the periods of all such substrings are the same and equal to $d \leq \frac{1}{8}n$. Therefore, d is also a period of the whole ww. But then $\gcd(|w|, d) \leq d \leq \frac{1}{8}|w|$ is also a period of ww. We conclude that $\gcd(|w|, d) \leq \frac{1}{8}|w|$ is period of w, hence $w = y^p$ for some $p \geq 8$, as claimed.

It remains to analyze the number of distinct squares in a circular word (w), where $w = y^p$ for $p \geq 8$. Each such square is a distinct square in y^{p+1}. The number of distinct squares in y^{p+1} is at most $2(p+1)|y| = 2\frac{p+1}{p}n \leq 2.25n$, since $p \geq 8$.

Theorem 5. *The number of distinct squares in a circular word of length n is at most $3.75n$.*

To improve on the above upper bound, we need some of the machinery used by Deza et al. [9]. Two occurrences of squares uu and UU starting at the same position such that $|u| < |U|$ are called a double square and denoted (u, U). If both are the rightmost occurrences, this is an FS-double square. An FS-double square is identified with the starting position of the two occurrences.

Lemma 6 (see proof of Theorem 32 in [9]). *If (u, U) is the leftmost FS-double square of a string x and $|x| \geq 10$, then the number of FS-double squares in x is at most $\frac{5}{6}|x| - \frac{1}{3}|u|$.*

We again consider the rightmost occurrence of every distinct square of length up to n in ww and assume that $w[\frac{1}{4}n..\frac{1}{2}n]$ is aperiodic (as otherwise we already know there are at most $2.25n$ distinct squares). We need to consider two cases: either there are no rightmost occurrences starting at $i = 1, 2, \ldots, \frac{1}{4}n$, or there is at least one such occurrence.

No Rightmost Occurrences Starting at $i = 1, 2, \ldots, \frac{1}{4}n$. In this case, it is enough to bound the number of distinct squares in $\hat{w} = w[(\frac{1}{4}n + 1)..n]w$. Let i be the starting position of the leftmost FS-double square (u, U) in $\hat{w}$. If $i > \frac{3}{4}n$ then the total number of distinct squares is at most $\frac{3}{4}n + 2n = 2\frac{3}{4}n$, so we assume $i \leq \frac{3}{4}n$. Then, the total number of distinct squares can be bounded by applying Lemma 6 on $w[(\frac{1}{4}n + i)..n]w$ to show that the number of FS-double squares is at most

$$\frac{5}{6}(\frac{7}{4}n - i + 1) - \frac{1}{3}|u|$$

We know that $i + 2|u| > \frac{3}{4}n$, as otherwise uu would occur later in w. Therefore, the maximum number of distinct squares is

$$\frac{7}{4}n + \frac{5}{6}(\frac{7}{4}n - i + 1) - \frac{1}{3}\frac{\frac{3}{4}n - i + 1}{2} = (\frac{7}{4} + \frac{35}{24} - \frac{1}{8})n - (\frac{5}{6} - \frac{1}{6})i + \frac{4}{6} \leq 3\frac{1}{12}n \quad (1)$$

At Least One Rightmost Occurrence Starting at $i \in \{1, 2, \ldots, \frac{1}{4}n\}$. We now move to the more interesting case where there are some rightmost occurrences starting at $i = 1, 2, \ldots, \frac{1}{4}n$. We then know by Lemma 3 that they all correspond to squares of the same length 2ℓ. Let $i \in \{1, 2, \ldots, \frac{1}{4}n\}$ be the starting position of one of these rightmost occurrences. Then, $i + 2\ell > n$ as otherwise the square would occur later in the second w, so $\ell > (n - \frac{n}{4})/2 = \frac{3}{8}n$. We also know that $\ell < \frac{1}{2}n$, as otherwise $w = y^2$ and there are only $3n$ distinct squares. To conclude, $\ell \in (\frac{3}{8}n, \frac{1}{2}n)$. Observe that, due to the square starting at position i, the aperiodic substring $s = w[\frac{1}{4}n..\frac{1}{2}n]$ also occurs at position $\frac{1}{4}n + \ell$ in ww. Therefore, we can rotate w by ℓ and repeat the whole reasoning. We either obtain that the number of distinct squares is at most $3\frac{1}{12}n$ (if, in $w^{(\ell)}w^{(\ell)}$, there are no rightmost occurrences starting at $i = 1, 2, \ldots, \frac{1}{4}n$), or there is another occurrence of s at position $\frac{1}{4}n + \ell + \ell' - n$ in w, where $\ell, \ell' \in (\frac{3}{8}n, \frac{1}{2}n)$. Because s is aperiodic and $\ell + \ell' > \frac{3}{4}n$, the other occurrence must actually be at position $\frac{1}{4}n - \Delta$, where $\Delta \in (\frac{1}{8}n, \frac{1}{4}n)$. By repeating this enough times (and recalling that two occurrences of s cannot be too close to each other, as otherwise s is not aperiodic), we either obtain that there are at most $3\frac{1}{12}n$ distinct squares or all occurrences of s in (w) are at positions $\frac{1}{4}n + \sum_{j=1}^{i-1} \Delta_j$ (recall that (w) denotes the circular word, so we calculate positions modulo n) for $i = 1, 2, \ldots, d$, where $\sum_{j=1}^{d} \Delta_j = n$ and $\Delta_j \in (\frac{1}{8}n, \frac{1}{4}n)$ for every $j = 1, 2, \ldots, d$. That is, the whole (w) is covered by the occurrences of s, and because s is aperiodic these occurrences overlap by less than $\frac{1}{8}n$. Observe that there cannot be any other occurrences of s in (w), because the additional occurrence would overlap with one of the already found occurrences by at least $\frac{1}{8}n$, thus contradiction the assumption that s is aperiodic. By the constraints on Δ_j, $d \in \{5, 6, 7\}$. See Fig. 3 for an illustration with $d = 7$. We further consider three possible subcases.

$d = 5$. In such case, we have $\Delta_j \geq \frac{1}{5}n$ for some j. By rotating w, we can assume that $j = 1$. Recall that then all squares starting at $i = 1, 2, \ldots, \frac{1}{4}n$ have the same length 2ℓ (and there is at least one such square), so there is another occurrence of s starting at position $\frac{1}{4}n + \ell$, and then by repeating the reasoning at position $\frac{1}{4}n + \ell + \ell'$, where $\ell + \ell' = n - \Delta_1$ (due to $\ell, \ell' \in (\frac{3}{8}n, \frac{1}{2}n)$). Combining this with $\Delta_1 \geq \frac{1}{5}n$, we obtain that $\min\{\ell, \ell'\} \leq \frac{2}{5}n$. By again rotating w, we can assume that in fact $\ell \leq \frac{2}{5}n$. Let $i \in \{1, 2, \ldots, \frac{1}{4}n\}$ be the starting position of a rightmost occurrence of a square of length 2ℓ. Then $i + 2\ell > n$ as otherwise it would not be a rightmost occurrence, so $i > \frac{1}{5}n$ and we obtain that there are less than $\frac{1}{4}n - \frac{1}{5}n = \frac{1}{20}n$ rightmost occurrences starting at $i = 1, 2, \ldots, \frac{1}{4}n$. By the previous calculation (1) the number of remaining rightmost occurrences is at most $3\frac{1}{12}n$, making the total number of distinct squares at most $3\frac{2}{15}n$.

$d = 6$. We will show that this is, in fact, not possible. Recall that, for every $i = 1, 2, \ldots, 6$, after rotating w by $r = \sum_{j=1}^{i-1} \Delta_j$ we obtain that there is at least one rightmost occurrence starting in the prefix of length $\frac{1}{4}n$ of $w^{(r)}w^{(r)}$, and in fact, by Lemma 3, all such rightmost occurrences correspond to squares of the same length $2\ell_i$, where $\ell_i \in (\frac{3}{8}n, \frac{1}{2}n)$. Thus, for every occurrence of s starting at position $\frac{1}{4}n + \sum_{j=1}^{i-1} \Delta_j$, there is another occurrence at position $\frac{1}{4}n + \sum_{j=1}^{i-1} \Delta_j + \ell_i$ in (w) (recall that the positions are taken modulo n). We claim that $\ell_i = \Delta_i + \Delta_{i+1}$ or $\ell_i = \Delta_i + \Delta_{i+1} + \Delta_{i+2}$, where the indices are taken modulo 6. Certainly, $\ell_i = \Delta_i + \Delta_{i+1} + \ldots + \Delta_{i+k}$ for some k. We cannot have

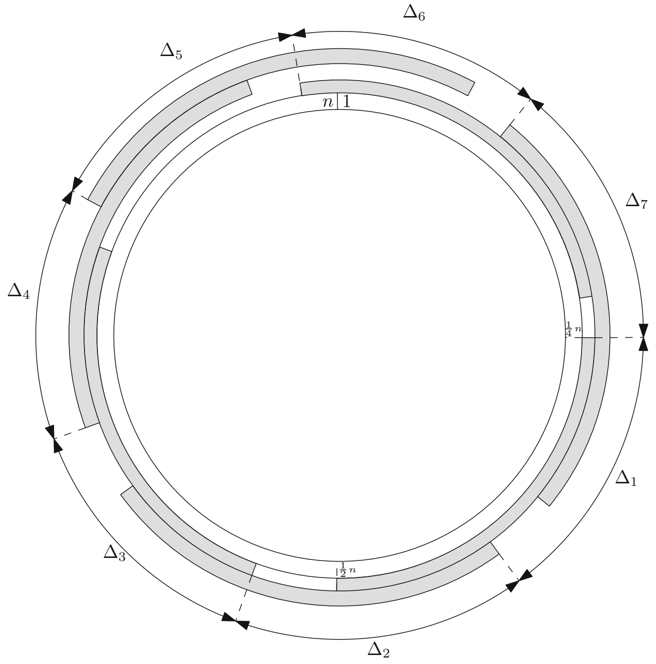

Fig. 3. Seven occurrences of an aperiodic s of length $\frac{1}{4}n$ inside (w).

$k = 0$ because $\ell_i > \frac{3}{8}n$ and $\Delta_i < \frac{3}{8}n$. We also cannot have $k \geq 3$, because $\ell_i < \frac{1}{2}n$ and $\Delta_i + \Delta_{i+1} + \Delta_{i+2} + \Delta_{i+3} > \frac{1}{2}n$. So, $k = 1$ or $k = 2$. For every $i = 1, 2, \ldots, 6$, we define $\mathsf{succ}(i) \in \{1, 2, \ldots, 6\}$ as follows. If $\ell_i = \Delta_i + \Delta_{i+1}$ then we set $\mathsf{succ}(i) = i + 2$, and otherwise (if $\ell_i = \Delta_i + \Delta_{i+1} + \Delta_{i+2}$) $\mathsf{succ}(i) = i + 3$. Intuitively, every occurrence of s in (w) points to another such occurrence. Due to $\ell_i \in (\frac{3}{8}n, \frac{1}{2}n)$ holding for every $i = 1, 2, \ldots, 6$, the difference between the starting positions of the i-th and the $\mathsf{succ}(i)$-th occurrence of s belongs to $(\frac{3}{8}n, \frac{1}{2}n)$, so the difference between the starting position of the i-th and the $\mathsf{succ}(\mathsf{succ}(i))$-th occurrence of s belongs to $(\frac{3}{4}n, n)$. In fact, due to s being aperiodic, the latter difference must belong to $(\frac{3}{4}n, \frac{7}{8}n)$. Consequently, there are no other occurrences of s between the $\mathsf{succ}(\mathsf{succ}(i))$-th and the i-th, so $\mathsf{succ}(\mathsf{succ}(i)) = i - 1$. Now, we consider two cases:

1. $\mathsf{succ}(1) = 3$, then $\mathsf{succ}(3) = 6$, so $\mathsf{succ}(6) = 2$, $\mathsf{succ}(2) = 5$ and $\mathsf{succ}(5) = 1$.
2. $\mathsf{succ}(1) = 4$, then $\mathsf{succ}(4) = 6$, so $\mathsf{succ}(6) = 3$, $\mathsf{succ}(3) = 5$, $\mathsf{succ}(5) = 2$, $\mathsf{succ}(2) = 4$.

In both cases, we obtain that $\mathsf{succ}(i) = \mathsf{succ}(j)$ for some $i \neq j$. But this is a contradiction, because then there are two occurrences of s within distance less than $\frac{1}{8}n$, so s is not aperiodic.

$d = 7$. We define $\mathsf{succ}(i)$ for every $i = 1, 2, \ldots, 7$ as in the previous case. Because $\mathsf{succ}(i) \in \{i+2, i+3\}$ and $\mathsf{succ}(\mathsf{succ}(i)) = i - 1$ still holds, we obtain that in fact

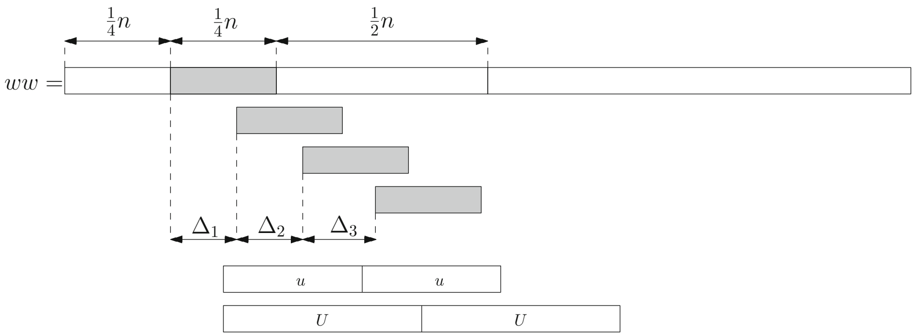

Fig. 4. The leftmost FS-square starting at position $j \leq \frac{1}{4}n + \Delta_1$.

$\mathsf{succ}(i) = i + 3$ for every $i = 1, 2, \ldots, 7$. This means that $\ell_i = \Delta_i + \Delta_{i+1} + \Delta_{i+2}$. Consider all rightmost occurrences starting at $i = 1, 2, \ldots, \frac{1}{4}n$. We must have that $i + 2\ell_1 > n$ for each of them, so $i > n - 2(\Delta_1 + \Delta_2 + \Delta_3)$, making the total number of such occurrences at most $\min\{\frac{1}{4}n, 2(\Delta_1 + \Delta_2 + \Delta_3) - \frac{3}{4}n\}$. Because $\Delta_1 + \Delta_2 + \Delta_3 \leq \frac{1}{2}n$ due to $\Delta_i > \frac{1}{8}n$ holding for every $i = 1, 2, \ldots, 7$ and $\sum_{i=1}^{7} \Delta_i = n$, this number is actually $2(\Delta_1 + \Delta_2 + \Delta_3) - \frac{3}{4}n$.

Now we must account for the remaining distinct squares. Let j be the starting position of the leftmost FS-double square (u, U) in ww. Note that $j > \frac{1}{4}n$ because there is at most one rightmost occurrence starting at $i = 1, 2, \ldots, \frac{1}{4}n$. We lower bound j by considering two possible cases:

1. $j > \frac{1}{4}n + \Delta_1$.
2. $j \leq \frac{1}{4}n + \Delta_1$, then the occurrences of s starting at $\frac{1}{4}n + \Delta_1$ and $\frac{1}{4}n + \Delta_1 + \Delta_2 + \Delta_3$ are disjoint and both fully inside the first w, because $\Delta_1 + \Delta_2 + \Delta_3 \leq \frac{1}{2}n$. Thus, both u and U contain s as a substring. See Fig. 4. Then, because all occurrences of s start at positions of the form $\frac{1}{4}n + \sum_{j=1}^{i-1} \Delta_j$, we conclude that $|u| = \Delta_2 + \Delta_3$ and $|U| = \Delta_2 + \Delta_3 + \Delta_4$. So, $j > n - 2(\Delta_2 + \Delta_3)$.

We now know that $j > \min\{\frac{1}{4}n + \Delta_1, n - 2(\Delta_2 + \Delta_3)\}$. Using $j + 2|u| > n$ we obtain that the number of remaining distinct squares is at most

$$1\frac{3}{4}n + \frac{5}{6}(2n - j) - \frac{1}{3}|u| \leq 3\frac{5}{12}n - \frac{5}{6}j - \frac{1}{3}\frac{n - j}{2} = 3\frac{1}{4}n - \frac{2}{3}j$$

so the total number of squares is

$$\begin{aligned}
&\leq 3\frac{1}{4}n + 2(\Delta_1 + \Delta_2 + \Delta_3) - \frac{3}{4}n - \frac{2}{3}j \\
&\leq 2\frac{1}{2}n + 2(\Delta_1 + \Delta_2 + \Delta_3) - \frac{2}{3}\min\{\frac{1}{4}n + \Delta_1, n - 2(\Delta_2 + \Delta_3)\}
\end{aligned}$$

We rewrite the above in terms of ℓ_1 and Δ_1:

$$2\frac{1}{2}n + 2\ell_1 - \frac{2}{3}\min\{\frac{1}{4}n + \Delta_1, n - 2\ell_1 + 2\Delta_1\} \leq 2\frac{1}{2}n + 2\ell_1 - \frac{2}{3}\min\{\frac{3}{8}n, \frac{5}{4}n - 2\ell_1\}$$

The above expression is increasing in ℓ_1. Because $\sum_{i=1}^{7} \ell_i = \sum_{i=1}^{7}(\Delta_i + \Delta_{i+1} + \Delta_{i+2}) = 3n$, after an appropriate rotation we can assume that $\ell_1 \leq \frac{3}{7}n$, and bound the expression:

$$2\frac{1}{2}n + \frac{6}{7}n - \frac{2}{3}\min\{\frac{3}{8}n, \frac{5}{4}n - \frac{6}{7}n\} = 3\frac{5}{14}n - \frac{1}{4}n = 3\frac{3}{28}n$$

Wrapping Up. We have obtained that either there is an aperiodic substring of length $\frac{1}{4}n$, and thus there are at most $2.25n$ distinct squares, or there are no rightmost occurrences starting at $i = 1, 2, \ldots, \frac{1}{4}n$ and the maximum number of distinct squares is $3\frac{1}{12}n$, or there is at least at least one rightmost occurrence starting at $i \in \{1, 2, \ldots, \frac{1}{4}n\}$. In the last case, either $d = 5$ and there are at most $3\frac{2}{15}n$ distinct squares, or $d = 7$ and there are at most $3\frac{3}{28}n$ distinct squares. The maximum of these upper bounds is $3\frac{2}{15}n$.

Theorem 7. *The number of distinct squares in a circular word of length n is at most* $3.14n$.

5 Conclusions

We believe that it should be possible to show an upper bound of $3n$, possibly without using the machinery of Deza et al., but it seems to require some new combinatorial insights. A computer search seems to suggest that the right answer is $1.25n$, but showing this is probably quite difficult. Another natural direction for a follow-up work is to consider higher powers in circular words.

References

1. Blanchet-Sadri, F., Mercas, R., Scott, G.: Counting distinct squares in partial words. Acta Cybern. **19**(2), 465–477 (2009)
2. Bland, W., Smyth, W.F.: Three overlapping squares: the general case characterized and applications. Theor. Comput. Sci. **596**, 23–40 (2015)
3. Castiglione, G., Restivo, A., Sciortino, M.: Circular Sturmian words and Hopcroft's algorithm. Theor. Comput. Sci. **410**(43), 4372–4381 (2009)
4. Crochemore, M., Fici, G., Mercaş, R., Pissis, S.P.: Linear-time sequence comparison using minimal absent words & applications. In: Kranakis, E., Navarro, G., Chávez, E. (eds.) LATIN 2016. LNCS, vol. 9644, pp. 334–346. Springer, Heidelberg (2016). doi:10.1007/978-3-662-49529-2_25
5. Crochemore, M., Iliopoulos, C.S., Kociumaka, T., Kubica, M., Radoszewski, J., Rytter, W., Tyczyński, W., Waleń, T.: The maximum number of squares in a tree. In: Kärkkäinen, J., Stoye, J. (eds.) CPM 2012. LNCS, vol. 7354, pp. 27–40. Springer, Heidelberg (2012). doi:10.1007/978-3-642-31265-6_3
6. Crochemore, M., Rytter, W.: Squares, cubes, and time-space efficient string searching. Algorithmica **13**(5), 405–425 (1995)
7. Currie, J.D.: There are ternary circular square-free words of length n for $n \geq 18$. Electron. J. Comb. **9**(1), N10 (2002)

8. Currie, J.D., Fitzpatrick, D.S.: Circular words avoiding patterns. In: Ito, M., Toyama, M. (eds.) DLT 2002. LNCS, vol. 2450, pp. 319–325. Springer, Heidelberg (2003). doi:10.1007/3-540-45005-X_28
9. Deza, A., Franek, F., Thierry, A.: How many double squares can a string contain? Discrete Appl. Math. **180**, 52–69 (2015)
10. Fan, K., Puglisi, S.J., Smyth, W.F., Turpin, A.: A new periodicity lemma. SIAM J. Discrete Math. **20**(3), 656–668 (2006)
11. Fine, N., Wilf, H.: Uniqueness theorems for periodic functions. Proc. Am. Math. Soc. **16**, 109–114 (1965)
12. Fraenkel, A.S., Simpson, J.: How many squares can a string contain? J. Comb. Theory Ser. A **82**(1), 112–120 (1998)
13. Franek, F., Fuller, R.C.G., Simpson, J., Smyth, W.F.: More results on overlapping squares. J. Discrete Algorithms **17**, 2–8 (2012)
14. Gawrychowski, P., Kociumaka, T., Rytter, W., Waleń, T.: Tight bound for the number of distinct palindromes in a tree. In: Iliopoulos, C., Puglisi, S., Yilmaz, E. (eds.) SPIRE 2015. LNCS, vol. 9309, pp. 270–276. Springer, Cham (2015). doi:10.1007/978-3-319-23826-5_26
15. Hegedüs, L., Nagy, B.: Representations of circular words. AFL. EPTCS **151**, 261–270 (2014)
16. Ilie, L.: A simple proof that a word of length n has at most $2n$ distinct squares. J. Comb. Theory **112**(1), 163–164 (2005)
17. Ilie, L.: A note on the number of squares in a word. Theor. Comput. Sci. **380**(3), 373–376 (2007)
18. Kopylova, E., Smyth, W.F.: The three squares lemma revisited. J. Discrete Algorithms **11**, 3–14 (2012)
19. Lothaire, M. (ed.): Algebraic Combinatorics on Words, Encyclopedia of Mathematics and its Applications, vol. 90. Cambridge University Press, Cambridge (2002)
20. Manea, F., Seki, S.: Square-density increasing mappings. In: Manea, F., Nowotka, D. (eds.) WORDS 2015. LNCS, vol. 9304, pp. 160–169. Springer, Cham (2015). doi:10.1007/978-3-319-23660-5_14
21. Massé, A.B., Brlek, S., Garon, A., Labbé, S.: Equations on palindromes and circular words. Theor. Comput. Sci. **412**(27), 2922–2930 (2011)
22. Shur, A.M.: On ternary square-free circular words. Electron. J. Comb. **17**(1), R140 (2010)
23. Simpson, J.: Intersecting periodic words. Theor. Comput. Sci. **374**(1–3), 58–65 (2007)
24. Simpson, J.: Palindromes in circular words. Theor. Comput. Sci. **550**, 66–78 (2014)
25. Thue, A.: Über unendliche zeichenreihen. Norske Vid. Selsk. Skr. I Mat.-Nat. Kl. Christiania **7**, 1–22 (1906)