Avoidance of split overlaps

Daniel Gabric and Jeffrey Shallit
School of Computer Science
University of Waterloo
Waterloo, ON N2L 3G1
Canada
dgabric@uwaterloo.ca
shallit@uwaterloo.ca

Abstract

We generalize Axel Thue's familiar definition of overlaps in words, and show that there are no infinite words containing split occurrences of these generalized overlaps. Along the way we prove a useful theorem about repeated disjoint occurrences in words — an interesting natural variation on the classical de Bruijn sequences.

1 Introduction

In this paper, we are concerned with words over a finite alphabet Σ of cardinality $k \geq 1$; more specifically, avoiding certain kinds of repetitions in them.

Two kinds of repetitions that have been studied for more than a hundred years are squares and overlaps [13, 1]. A *square* is a finite nonempty word of the form xx (such as the English word `murmur`). Another type of repetition is the α-power. We say a word w is an α-power, for $\alpha = p/q$, a rational number, if $|w| = p$ and w has period q. (We say a word w has period $q \geq 1$ if $w[i] = w[i+q]$ for all i for which this makes sense.) Thus `alfalfa` is a (7/3)-power. A word y is a *factor* of a word w if $w = xyz$ for words x, z (possibly empty). When we speak about a word "avoiding α-powers", we mean it has no factor that is a β-power, for all $\beta \geq \alpha$. The smallest period of a word w is sometimes called *the* period, and is written $\text{per}(w)$.

An *overlap* is a finite word of the form $axaxa$ for a a single letter, and x a (possibly) empty word, such as the French word `entente`. An overlap can be viewed as just slightly more than a square: it consists of two repetitions of a nonempty word w, followed by the first letter of w.

The term "overlap" comes from the following "folk" observation: say two distinct occurrences of a length-n factor x in w, say $x = w[i..i+n-1] = w[j..j+n-1]$ with $i < j$, "overlap each other" if $0 < j - i < n$.

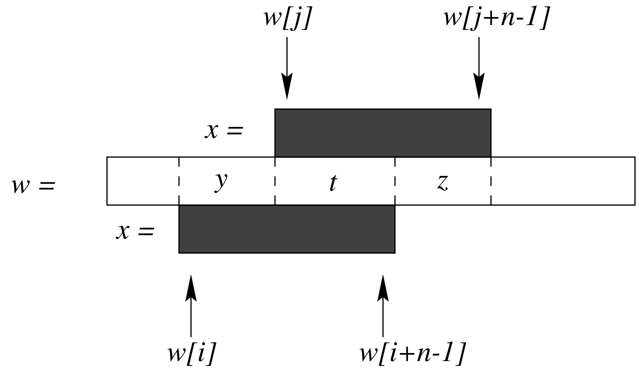

Figure 1: Overlapping factors

Proposition 1. *If w contains two distinct occurrences of x that overlap each other, then w contains an overlap.*

Proof. Define $y = w[i..j-1]$, $t = w[j..i+n-1]$, and $z = w[i+n..j+n-1]$ and examine Figure 1. Each of these words is nonempty, and $x = yt = tz$. By the Lyndon-Schützenberger theorem [9], it follows that there exist words u, v with u nonempty, and an integer $e \geq 0$, such that $y = uv$, $t = (uv)^e u$, and $z = vu$. Thus $w[i..j+n-1] = ytz = (uv)^{e+2}u$, which contains an overlap. □

This suggests the following natural generalization of overlap: a *t-overlap* is a word of the form xxx', where x is a nonempty word of length at least t, and x' is the first t letters of x. For example, the unfamiliar English word `prelinpinpin` contains a suffix that is a 2-overlap. Note that a 0-overlap is a square, and a 1-overlap is an ordinary overlap. Of course, a t-overlap contains a t'-overlap for all $t' < t$.

Thue proved [13, 1] that one can avoid overlaps over any alphabet containing at least two letters. Here by "avoid" we mean "there exists an infinite word containing no overlaps" or, equivalently, "there exist infinitely many finite words containing no overlaps". Since every t-overlap contains a 1-overlap, Thue's construction also shows it is possible to avoid t-overlaps over any alphabet with at least two letters.

So instead, in this paper we consider split occurrences of repetitions. A *split occurrence* of a repetition is a word of the form xyz, where xz forms the repetition. For example, the English word `contentment` contains a split occurrence of the 2-overlap `ntentent`, which arises from the factor xyz, where $x =$ `ntent`, $y =$ `m`, $z =$ `ent`.

It follows from known results that there exist infinite words avoiding split occurrences of α-powers, for any rational number $\alpha > 2$ [10]. To see this, take the alphabet size k sufficiently large that there exists an infinite word $\mathbf{w}$ over $\Sigma_k = \{0, 1, \ldots, k-1\}$ avoiding $\alpha/2$ powers. (By Dejean's theorem [5, 2, 12] this is possible.) Suppose xyz is a factor of $\mathbf{w}$ that is a split occurrence of a β power for $\beta \geq \alpha > 2$. Then clearly either x is a $\geq \beta/2$ power or z is, a contradiction.

In contrast, in this paper we show that no matter what the alphabet size is, there are no infinite words avoiding split occurrences of t-overlaps. Our main theorem is the following.

Theorem 2. *Let w be a word over a k-letter alphabet avoiding split occurrences of t-overlaps. Then w is finite.*

We also investigation the avoidance of *reversed split* repetitions. A reversed split repetition is a word of the form xyz, where zx forms the repetition. For example, the English word `independent` contains a reversed split overlap: it has the factor xyz, where $x =$ `nde`, $y =$ `p`, and $z =$ `ende`, giving the overlap $zx =$ `endende`.

2 Some useful results on primitive words and bordered words

We call a nonempty word w *primitive* if w cannot be written in the form x^k for an integer $k \geq 2$; see, for example, [6].

Lemma 3. *Let $A_k(n,p)$ denote the number of length-n words over Σ_k with smallest period p, and let $\psi_k(n)$ denote the number of primitive words over Σ_k. Then $A_k(n,p) = \psi_k(p)$ for $1 \leq p \leq \frac{n}{2} + 1$.*

Proof. We claim that every length-n word w with shortest period p can be written in the form $w = x^i x'$, where x' is a prefix of x and $|x| = p$ and x primitive. For if x were not primitive, say $x = y^j$ for some $j \geq 2$, then p could not be the shortest period.

We now claim that if x is primitive and $1 \leq p \leq \frac{n}{2} + 1$, then $w = x^{n/p}$ has shortest period p. Suppose to the contrary that w has shortest period $q < p$. Since $n \geq p + q - 1$, by the Fine-Wilf theorem [7], w also has the period $\gcd(p,q)$. If q divides p, then x was not primitive, a contradiction. Otherwise $\gcd(p,q) < q$, a contradiction. $\square$

A *border* of a word w is a nonempty word x, $x \neq w$, such that x is both a prefix and suffix of w. Thus `entanglement` has the border `ent`. If a word has a border, it is called *bordered*, and otherwise it is called *unbordered*. It is easy to see that if a word of length n has a border, it must have a border of length $\leq n/2$.

Lemma 4. *For $k \geq 2, n \geq 1$, there are at least $k^n(1 - 1/k - 1/k^2)$ unbordered words of length n over a k-letter alphabet.*

Proof. Let $u_k(n)$ denote the number of unbordered words of length n over a k-letter alphabet. It follows from the recurrence for $u_k(n)$ given in [11] that $u(k,n)$ is a polynomial of degree n in k. By explicit computation of these polynomials for $n = 1, 2, \ldots, 12$, we can easily verify the inequality $u_k(n) \geq k^n(1 - 1/k - 1/k^2) \geq k^n(1 - 1/k - 1/k^2)$ for $n \leq 12$. In particular, $u_k(12) = k^{12} - k^{11} - k^{10} + k^6 + k^5 - k^2$.

Now assume $n > 12$. For each unbordered word w of length 12, write $w = xz$ with $|x| = |z| = 6$, and consider the words xyz of length n, where y is an arbitrary word of length

$n-12$. There are $u_k(12)k^{n-12}$ such words. Each such word is unbordered, unless it has a border of length i for $6 < i \leq n/2$. But the total number of words with border length i satisfying $6 < i \leq n/2$ is at most

$$k^{n-7} + k^{n-8} + \cdots + k^{n/2} \leq (k^{n-6} - 1)/(k-1).$$

Therefore, there are least

$$u_k(12)k^{n-12} - (k^{n-6}-1)/(k-1) = k^n(1-1/k-1/k^2+1/k^6+1/k^7-1/k^{10}) - (k^{n-6}-1)/(k-1)$$

unbordered words of length n for $n > 12$. Since $k^n/k^6 \geq (k^{n-6}-1)/(k-1)$ and $k^{n-7} \geq k^{n-10}$, the desired bound follows. □

3 Disjoint occurrences

Let $\Sigma_k = \{0, 1, \ldots, k-1\}$ be an alphabet of $k \geq 1$ letters. It is known that for every $k \geq 1$ and $n \geq 1$, there exists a word of length $k^n + n - 1$ that contains every length-n word exactly once as a factor; such words are called *de Bruijn words* of order n; see [3, 4]. This bound of $k^n + k - 1$ is optimal, because from the pigeonhole principle, it follows that if w is a word of length $\geq k^n + n$, then w must contain at least two different occurrences of some word x of length n.

However, these two different occurrences of x could overlap each other in w. If two distinct occurrences do not overlap, we say they are *disjoint*.)

If we insist on having two disjoint occurrences, we get a different bound. For example, there are binary words of length 7 that do not contain two disjoint occurrences of the same length-2 word, such as 0111000. Let us define $C(k, n)$ to be the length of the longest word over Σ_k having the property that there are no two disjoint occurrences of the same word. By considering disjoint occurrences of length-n blocks, the pigeonhole principle easily gives the bound $C(k, n) < n(k^n + 1)$. We now obtain some better bounds on $C(k, n)$.

We need a lemma.

Lemma 5. *Let x, w be words with $|x| = n$. Suppose w contains m occurrences of x, but not two or more disjoint occurrences. Then $m \leq \lceil n/\operatorname{per}(x) \rceil$. Furthermore, for each individual x, this upper bound is achievable.*

Proof. Let w contain the maximum possible number of overlapping occurrences of the length-n word x, and no disjoint occurrences of x. Let d be the shortest distance between two consecutive occurrences of x in w. If there are m overlapping occurrences, then the last occurs at distance at least $d(m-1)$ from the first. If $d(m-1) \geq n$, then the last occurrence does not overlap the first, so $d(m-1) < n$. It follows that $m < n/d + 1$, and since t is an integer, we have $m \leq \lceil n/d \rceil$.

We now show that $d = \operatorname{per}(x)$. Two overlapping occurrences of x with the shortest distance between them correspond to writing $x = yt = tz$ for some y, t, z (with t the overlap), with $1 < |t| < n$, and minimizing $|y|$; see the diagram. Now, from the Lyndon-Schützenberger theorem [9], it follows that there exist u, v with u nonempty and an integer $e \geq 0$ such that that $y = uv$, $t = (uv)^e u$, and $z = vu$. Hence $y = uv$ is a period of x; to minimize y we take y to be the shortest period of x.

We have now shown that $m \leq \lceil n/\operatorname{per}(x) \rceil$. It remains to see that this bound is always achievable. Let y be the shortest period of x, and write $x = y^f u$, where u is a nonempty prefix of y, possibly equal to y itself. Then $y = uv$ for some (possibly empty) v. Consider the word $w = (uv)^{2f} u$; it is easy to see that $x = (uv)^f u$ overlaps itself at least $f + 1$ times in this w. Since $f|y| < n \leq (f+1)|y|$, it follows that $f + 1 = \lceil n/\operatorname{per}(x) \rceil$. □

Theorem 6. *We have*

$$C(k,n) \leq \left(\sum_{w \in \Sigma_k^n} \left\lceil \frac{n}{\operatorname{per}(w)} \right\rceil \right) + n - 1.$$

Proof. Let w be a longest word having no disjoint occurrences of the same length-n factor. Let us now count the number of occurrences of each length-n factor x in w. By Lemma 5, w can contain at most $\lceil n/\operatorname{per}(x) \rceil$ occurrences of x. Thus, in the worst case, w can have at most $\sum_{x \in \Sigma_k^n} \lceil \frac{n}{\operatorname{per}(x)} \rceil$ total occurrences of length-n words. Thus the word can be of length at most $\left(\sum_{x \in \Sigma_k^n} \lceil \frac{n}{\operatorname{per}(x)} \rceil \right) + n - 1$. □

Corollary 7. *For $k \geq 2$ we have $C(k,n) \leq k^n(1+1/k+1/k^2) + n(k^{n/2+1} - 1)/(k-1) + n - 1$.*

Proof. We split the sum $\sum_{x \in \Sigma_k^n} \lceil \frac{n}{\operatorname{per}(x)} \rceil$ into three parts: one where $\operatorname{per}(x) \leq n/2$, one where $n/2 < \operatorname{per}(x) < n$, and one where $\operatorname{per}(x) = n$.

From Lemma 3 above, the number of length-n words x with smallest period $p \leq n/2$ is $\psi(k,p)$, the number of primitive words of length p over a k-letter alphabet. Write $A = \sum_{1 \leq p \leq n/2} \psi(k,p)$ and $B = \sum_{1 \leq p \leq n/2} \psi(k,p) \lceil n/p \rceil$. It is known that $\psi(k,n) = \sum_{d|n} \mu(d) k^{n/d}$, where μ is the Möbius function from number theory (see, e.g., [6, p. 245]), but the much weaker bound $\psi(k,n) \leq k^n$ suffices for our purposes here. Thus $B \leq n(k + k^2 + \cdots + k^{n/2}) \leq n(k^{n/2+1} - 1)/(k-1)$.

The number of words with period n is $u_k(n)$, the number of unbordered words of length n. From Lemma 4 we have $u_k(n) \geq k^n(1 - 1/k - 1/k^2)$. Thus we have

$$\begin{aligned} \sum_{x \in \Sigma_k^n} \left\lceil \frac{n}{\operatorname{per}(x)} \right\rceil &= B + 2(k^n - A - u_k(n)) + u_k(n) \qquad (1) \\ &\leq 2k^n - u_k(n) + B \\ &\leq 2k^n - k^n(1 - 1/k - 1/k^2) + B \\ &\leq k^n(1 + 1/k + 1/k^2) + n(k^{n/2+1} - 1)/(k-1), \end{aligned}$$

from which the result follows. □

Theorem 8.

(a) $C(1, n) = 2n - 1$ *for* $n \geq 1$*;*

(b) $C(k, 1) = k$ *for* $k \geq 1$*;*

(c) $C(k, 2) = k^2 + k + 1$ *for* $k \geq 1$*;*

(d) $C(k, 3) = k^3 + k^2 + k + 2$ *for* $k \geq 1$*.*

Proof.

(a) A unary word of length $2n$ has two disjoint length-n occurrences.

(b) A word of length $k + 1$, by the pigeonhole principle, has two occurrences of a single letter.

(c) Take a de Bruijn word of order 2 over a k-letter alphabet; it has length k^2+1. Replace each occurrence of aa with aaa; such a replacement clearly does not introduce any disjoint occurrences. The resulting word has length $k^2 + k + 1$. This gives the lower bound. For the upper bound, we use Theorem 6. All length-2 words have period 2, except those of the form aa, which have period 1. Then the sum in Theorem 6 gives the upper bound.

(d) For the upper bound, we note that all length-3 words have period 3, except that aaa has period 1 and aba, with $a \neq b$, has period 2. The sum in Theorem 6 then gives $k^3 + k^2 + k + 2$.

To see the lower bound, we need some terminology and a lemma.

A function $f : \Sigma_k^n \to \Sigma_k$ is said to be a *feedback function*. A feedback function f is said to be *non-singular* if the function $F : \Sigma_k^n \to \Sigma_k^n$ defined as $F(a_1a_2 \cdots a_n) = a_2 \cdots a_n f(a_1a_2 \cdots a_n)$ is one-to-one.

A *universal cycle* for a set of words $S \subseteq \Sigma_k^n$ is a length-$|S|$ word that, when considered circularly, contains every word in S as a factor. A non-singular feedback function partitions Σ_k^n into sets $S_1, S_2, \ldots, S_m$, each having a corresponding universal cycle. For each word $w = w_1w_2 \cdots w_n \in S_i$, for some $1 \leq i \leq m$, we have that $w_2w_3 \cdots w_n f(w) \in S_i$ and w has a corresponding word $v = v_1v_2 \cdots v_n \in S_i$ such that $w = v_2v_3 \cdots v_n f(v)$. The *cycle representative* of a set S is the lexicographically least word in S. It is well known that de Bruijn words can be constructed by joining the universal cycles in a specific way, sometimes with use of a successor rule. A *successor rule* is a feedback function that determines the next symbol in a de Bruijn word using the previous n symbols.

Lemma 9. *For all* $k \geq 2$*, there exists a* k*-ary de Bruijn word of order* 3 *that contains either* $abab$ *or* $baba$ *for all* $a \neq b$ *where* $a, b \in \Sigma_k$*.*

Proof. Consider the feedback function $f : \Sigma_k^3 \to \Sigma_k$ defined by $f(a_1a_2a_3) = a_1+a_2-a_3$. We will show that the function $F(a_1a_2a_3) = a_2a_3f(a_1a_2a_3)$ is one-to-one. Suppose there exist two words $a_1a_2a_3$ and $b_1b_2b_3$ such that $F(a_1a_2a_3) = F(b_1b_2b_3)$. Then we would have that $a_2a_3(a_1 + a_2 - a_3) = b_2b_3(b_1 + b_2 - b_3)$. But this implies that $a_2 = b_2$,$a_3 = b_3$, and $a_1 + a_2 - a_3 = b_1 + b_2 - b_3$. These three equations imply $a_1 = b_1$. Now we have $a_1a_2a_3 = b_1b_2b_3$. Therefore F is one-to-one.

We now define a feedback function g based f and we show g is a successor rule. Let $a_1a_2a_3 \in \Sigma_k^3$. Let $\tau(a_2a_3)$ be an increasing sequence of symbols $c \in \Sigma_k$ such that a_2a_3c is a cycle representative of some set in the partition of Σ_k^3 by f. If 0 is in $\tau(a_2a_3)$ and $a_2a_3c \neq 000$, then prepend $f(0a_2a_3)$ to the sequence. If 0 is not in $\tau(a_2a_3)$ and $\tau(a_2a_3)$ is nonempty, then prepend 0 to the sequence. Let $t_0, t_1, \ldots, t_{p-1}$ be the sequence $\tau(a_2a_3)$. Let $g : \Sigma_k^3 \to \Sigma_k$ be a feedback function defined as follows:

$$g(a_1a_2a_3) = \begin{cases} t_{(j+1) \bmod p}, & \text{if } f(a_1a_2a_3) = t_j \text{ for some } j \in \{0, 1, \ldots, p-1\}; \\ f(a_1a_2a_3), & \text{otherwise.} \end{cases}$$

By [8, Theorem 4.3] we have that g is a successor rule. We now argue that $g(aba) = b$ for all $a, b \in \Sigma_k$ with $a < b$. Since $f(aba) = a + b - a = b$, it suffices to show that $\tau(ba)$ is empty. Suppose that $\tau(ba)$ is nonempty. Then there exists a $d \in \Sigma_k$ such that bad is a cycle representative of some set S' in the partition of Σ_k^3 by f. Consider the word $adf(bad) \in S'$. Since $a < b$, we have that $adf(bad)$ is lexicographically smaller than bad. Thus bad cannot be a cycle representative. So $\tau(a_2a_3)$ is empty. $\square$

We can now continue with the proof of the lower bound. Suppose $abab$ is the occurrence; we can then insert ab immediately after its occurrence. We can also insert aa after the unique occurrence of aaa for each letter a. This introduces no disjoint occurrences, but adds $k(k-1) + 2k$ letters to the de Bruijn word of length $k^3 + 2$, thus matching the upper bound.

$\square$

Computing the exact value of $C(k, n)$, even for k and n seems like a difficult problem. In Table 1 below we give the first few values of this function, obtained by brute force of the solution space.

k \ n	1	2	3	4	5	6	7
1	1	3	5	7	9	11	13
2	2	7	16	32	59	110	≥ 192
3	3	13	41				
4	4	21	86				
5	5	31					

Table 1: Values of $C(k, n)$

Words achieving the bounds in Table 1 are given below:

k	n	Word achieving $C(k,n)$
2	2	0001110
2	3	0000010101111100
2	4	0101010010011011011111111100000001
2	5	0000000001000100011001100111010010101010110110111111110000
2	6	00000000000100001000011000110001110011100111101000101001010010110010011011011010101010111011101111111111100000
3	2	0001021112220
3	3	00000101011002020210220121212222211111200
4	2	0001020311121322223330
4	3	00000101011002020210220030303103201203301302311111212122113131321331232323333322222300
5	2	000102030411121314222324333444 0

Table 2: Words achieving the bounds in Table 1

For all of the entries in this table, except $(4,3)$, the word given is guaranteed to be the lexicographically least.

The value $C(2,6) = 110$ and the associated lexicographically least string, and the bound $C(2,7) \geq 192$ were computed by Bert Dobbelaere, who kindly allowed us to quote them here.

4 Split occurrences of t-overlaps

We now can prove Theorem 2, the main result of this paper. We do so by finding an explicit bound on the length of the longest word avoiding split overlaps. Let $n \geq 0$ and $k \geq 1$ be fixed integers. Define $S(k,t)$ (resp., $R(k,t)$) to be the length of the longest word over a k-letter alphabet containing no occurrences of split t-overlaps (resp., reversed split t-overlaps).

Theorem 10. *We have*

(a) $S(k,t) \leq C(k, C(k,t)+1)$*;*

(b) $S(k,0) = k$*;*

(c) $S(k,1) \leq k^{k+1} + k - 1$*;*

(d) $S(1,t) = 3t - 1$ *for* $t \geq 1$*;*

and the same bounds hold for $R(k,t)$*.*

Proof. We prove the results only for split overlaps; exactly the same arguments can be used for reversed split overlaps.

(a) Let $|w| \geq C(k, C(k,t)+1)+1$. Then w contains at least two disjoint occurrences of some factor x of length $C(k,t)+1$. Write $w = pxqxr$. Then x itself contains two disjoint occurrences of some factor y of length t. Write $x = syuyv$. Then $w = psyuyvqsyuyvr$. Now w contains the factor $yuyvqsyuy$ and so the split t-overlap $yuy \cdot uy$. It therefore follows that $S(k,t) \leq C(k, C(k,t)+1)$, as desired.

(b) For $t = 0$, we can take $C(k,t) = k$. For if a word w is of length at least $k+1$, it must contain two repeated letters, say $w = xayaz$, and hence the split square $a \cdots a$.

(c) For $t = 1$, we have $C(k,t) \leq k^{k+1} + k - 1$. We can use the argument in (a), but with a small twist. Consider the factors of length $k+1$ in a word w of length at least $k^{k+1} + k$. There are at least $k^{k+1} + 1$ of these factors, and by the pigeonhole principle, some factor x of length $k+1$ appears at least twice in w. If these two occurrences of x overlap in w, we are already done, because they contain an overlap right there by Proposition 1. Otherwise, write $w = sxtxu$ for some s, t, u. Now x is of length $k+1$, so again by the pigeonhole principle, some letter a is repeated in x. Write $x = paqar$ for some words p, q, r. Putting this all together, we have $w = spaqartpaqaru$. Consider the factor $aqartpaqa$. It has the split 1-overlap $aq \cdots aqa$.

(d) Easy. Left to the reader. □

Table 3 gives the values of $S(k,t)$ we have computed by brute force.

$k \backslash t$	0	1	2	3	4
1	1	2	5	8	11
2	2	4	12	47	
3	3	9	≥ 97		
4	4	31			
5	5	≥ 100			

Table 3: Values of $S(k,t)$

Words achieving the nontrivial bounds in Table 3 are given below:

k	t	Lexicographically least word achieving $S(k,t)$
2	1	0011
2	2	000110100111
2	3	00111010100001010011101000011111000011010001110
3	1	012021012
4	1	0120321301231013210203123021031

Table 4: Lexicographically least word achieving the bounds in Table 3

Table 5 gives the values of $R(k,t)$ we have computed by brute force.

$k \backslash t$	0	1	2	3	4
1	1	2	5	8	11
2	2	4	15	46	≥ 213
3	3	9	≥ 110		
4	4	30			
5	5	≥ 122			

Table 5: Optimal values of $R(k,t)$

Words achieving the nontrivial bounds in Table 5 are given below:

k	t	Lexicographically least word achieving $R(k,t)$
2	1	0011
2	2	010001100111001
2	3	0010100110100011111000111010000011101010001100
3	1	012010210
4	1	012031231032021030231321023013

Table 6: Lexicographically least word achieving the bounds in Table 3

5 Remarks

We currently do not know whether the upper bound in Theorem 6 is tight, or asymptotically tight, except when $n \leq 3$. Improvement of this bound, or construction of examples nearly matching the bound, would be of interest.

It is a challenging computational problem to compute more values of $C(k,n)$, $S(k,t)$, and $R(k,t)$, which we leave to the reader.

We are grateful to Farbod Yadegarian, who computed lower bounds on $R(5,1)$, $R(3,2)$, and $R(2,4)$.

References

[1] J. Berstel. *Axel Thue's Papers on Repetitions in Words: a Translation.* Number 20 in Publications du Laboratoire de Combinatoire et d'Informatique Mathématique. Université du Québec à Montréal, February 1995.

[2] J. Currie and N. Rampersad. A proof of Dejean's conjecture. *Math. Comp.* **80** (2011), 1063–1070.

[3] N. G. de Bruijn. A combinatorial problem. *Proc. Konin. Neder. Akad. Wet.* **49** (1946), 758–764.

[4] N. G. de Bruijn. Acknowledgement of priority to C. Flye Sainte-Marie on the counting of circular arrangements of $2n$ zeros and ones that show each n-letter word exactly once. Technical Report 75-WSK-06, Department of Mathematics and Computing Science, Eindhoven University of Technology, The Netherlands, June 1975.

[5] F. Dejean. Sur un théorème de Thue. *J. Combin. Theory Ser. A* **13** (1972), 90–99.

[6] P. Dömösi and M. Ito. *Context-Free Languages and Primitive Words.* World Scientific, 2015.

[7] N. J. Fine and H. S. Wilf. Uniqueness theorems for periodic functions. *Proc. Amer. Math. Soc.* **16** (1965), 109–114.

[8] D. Gabric, J. Sawada, A. Williams, and D. Wong. A successor rule framework for constructing k-ary de Bruijn sequences and universal cycles. *IEEE Trans. Inform. Theory* **66**(1) (2020), 679–687.

[9] R. C. Lyndon and M. P. Schützenberger. The equation $a^M = b^N c^P$ in a free group. *Michigan Math. J.* **9** (1962), 289–298.

[10] H. Mousavi and J. Shallit. Repetition avoidance in circular factors. In M.-P. Béal and O. Carton, editors, *Developments in Language Theory, 17th International Conference, DLT 2013*, Vol. 7907 of *Lecture Notes in Computer Science*, pp. 384–395. Springer-Verlag, 2013.

[11] P. T. Nielsen. A note on bifix-free sequences. *IEEE Trans. Inform. Theory* **IT-19** (1973), 704–706.

[12] M. Rao. Last cases of Dejean's conjecture. *Theoret. Comput. Sci.* **412** (2011), 3010–3018.

[13] A. Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske vid. Selsk. Skr. Mat. Nat. Kl.* **1** (1912), 1–67. Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, 1977, pp. 413–478.