

McMaster University

Advanced Optimization Laboratory



Title:

How many double squares can a string contain?

Authors:

Antoine Deza, Frantisek Franek, and Adrien Thierry

AdvOL-Report No. 2013/1

August 2013, Hamilton, Ontario, Canada

How many double squares can a string contain?

Antoine Deza, Frantisek Franek, and Adrien Thierry

August 2013

Abstract

Counting the types of squares rather than their occurrences, we consider the problem of bounding the number of distinct squares in a string. Fraenkel and Simpson showed in 1998 that a string of length n contains at most $2n$ distinct squares. Ilie presented in 2007 an asymptotic upper bound of $2n - \Theta(\log n)$. We show that a string of length n contains at most $\lfloor 11n/6 \rfloor$ distinct squares. This new upper bound is obtained by investigating the combinatorial structure of double squares and showing that a string of length n contains at most $\lfloor 5n/6 \rfloor$ particular double squares. In addition, the established structural properties provide a novel proof of Fraenkel and Simpson's result.

1 Introduction

A square in a string is a tandem repetition of the form $u^2 = uu$. The repeating part, u , is referred to as the *generator* of the square u^2 . If the generator u is *primitive*, i.e. not a repetition of a string, then the square is called *primitively rooted*. The problem of counting the types of squares in a string of length n – later referred to as *the number of distinct squares problem* – was introduced by Fraenkel and Simpson [4] in 1998 who showed that the number of distinct squares in a string of length n is at most $2n$. Their proof relies on a lemma by Crochemore and Rytter [1] describing the relationship among the sizes of three primitively rooted squares starting at the same position. Not using Crochemore and Rytter's Lemma, Ilie [6] provided an alternative proof of Fraenkel and Simpson's result before presenting in [7] an asymptotic upper bound of $2n - \Theta(\log n)$ for sufficiently large n . A d -step approach to this problem introducing the size d of the alphabet as a parameter in addition to the length n of the string was proposed in [2]. Considering the maximum number $\sigma_d(n)$ of distinct primitively rooted squares over all strings of length n with exactly d distinct symbols, it is conjectured there that $\sigma_d(n) \leq n - d$. Note that the number of non-primitively rooted squares, i.e. squares whose generators are repetitions, is bounded by $\lfloor n/2 \rfloor - 1$, see Kubica et al. [9].

A configuration of two squares u^2 and U^2 starting at the same position and so that $|u| < |U| < 2|u| < 2|U|$ has been investigated in different contexts. For instance, the configuration of such two squares with a third one is investigated in [5, 8] with the intention of providing a position where a third square could not start in order to tackle the maximum number of runs conjecture. Within the computational framework introduced in [3], such configurations are investigated in [11] to enhance the determination of $\sigma_d(n)$. Such configurations of two

squares are unique in the context of rightmost occurrences of squares since at most two such squares can start at the same position as shown by Fraenkel and Simpson. In [10] Lam investigates what he calls double squares, i.e. configurations of two rightmost occurrences of squares starting at the same position, in order to bound their number and thus bound the number of distinct squares.

We present structural properties of double squares arising in various contexts and coinciding with Lam's double squares in the context of rightmost occurrences which we refer to as FS-double squares. The structural properties of double squares presented in this paper not only give a novel proof of Fraenkel and Simpson's result, they allow bounding the number of FS-double squares in a string of length n by $\lfloor 5n/6 \rfloor$, which in turn leads to a new upper bound for the number of distinct squares of $\lfloor 11n/6 \rfloor$.

2 Combinatorics of double squares

2.1 Preliminaries

We deal with finite strings over finite alphabets and index strings starting from 1. Thus $x[1]$ refers to the first symbol of a string x , $x[2]$ to the second etc. We use $..$ as a range symbol, thus $x = x[1..n]$ is a string of length n , and $x[i..j]$ refers to the substring, also often called *factor*, starting at position i and ending at position j . For a substring $y = x[i..j]$, $s(y)$ respective $e(y)$ denotes its starting, respective ending, position, i.e. $(s(y), e(y)) = (i, j)$. A substring $y = x[i..j]$ of $x = x[1..n]$ is called a *prefix* respective *suffix* of x if $i = 1$ respective $j = n$, and is *proper* if $y \neq x$, while we call it *trivial* if y is empty. For a string x , a *non-trivial power* of x is a string x^m for some integer $m \geq 2$, where x^m represents a concatenation of m copies of x . In particular, x^2 is called a *square*, and x^3 a *cube*.

Definition 1. *A string x is primitive if x cannot be expressed as a non-trivial power of any string. For any string x , there is a primitive string y so that $x = y^m$ for some integer $m \geq 1$. Such y and m are unique and y is called the primitive root of x . Two strings x and y are conjugates if there are strings u and v so that $x = uv$ and $y = vu$. Note that x is a trivial conjugate of itself. Often the term rotation is used for conjugates.*

Lemmas 2 and 3 are folklore and presented without proofs.

Lemma 2 (Synchronization principle lemma). *Given a primitive string x , a proper suffix y of x , a proper prefix z of x , and $m \geq 0$, there are exactly m occurrences of x in yx^mz .*

Note that Lemma 2 implies that a primitive string does not equal to any of its conjugates.

Lemma 3 (Common factor lemma). *For any primitive strings x and y , if a non-trivial power of x and a non-trivial power of y have a common factor of length $|x|+|y|$, then x and y are conjugates.*

2.2 Double squares

Definition 4. A configuration of two squares u^2 and U^2 in a string x starting at the same position is referred to as a double square. In case that $|u| < |U|$, we say that (u, U) is a double square, i.e. the smaller generator is listed first.

For a double square (u, U) in a string x , if $|u| < |U| < 2|u|$, we say that the squares u^2 and U^2 are proportional and we call such a double square balanced.

For a double square (u, U) , if moreover u^2 and U^2 are rightmost occurrences in x , we refer to the double square (u, U) as FS-double square of x .

Note that if (u, U) is a double square, respective balanced double square, in x and x is a substring of y , then (u, U) is a double square, respective balanced double square, in y as well. For FS-double square, due to u^2 being a rightmost occurrence in x , $|U| < 2|u|$, as otherwise in x would be a farther copy of u^2 , and so every FS-double square is automatically balanced. If x is a substring of y , (u, U) need not be a FS-double square in y ; on the other hand if x is a suffix of y , then (u, U) is a FS-double square in y as well. We refer to the balanced double squares of rightmost occurrences as FS-double squares in recognition of Fraenkel and Simpson's pioneering efforts in the problem.

In Lemma 6 we shall show that certain types of balanced double squares have a unique factorization consisting of a nearly periodical repetition of a primitive string. The following Lemma 5 is used in Lemma 6 to prove uniqueness of this factorization.

Lemma 5. Let $u_1^p u_2 = v_1^q v_2$ where u_1, v_1 are primitive, u_2 is a non-trivial proper prefix of u_1 , and v_2 is a non-trivial proper prefix of v_1 . If $p \geq 2$ and $q \geq 2$, then $u_1 = v_1$, $u_2 = v_2$, and $p = q$.

Proof. Since $p \geq 2$ and $q \geq 2$, and u_1^p and v_1^q have a common factor of size $|u_1| + |v_1|$, then by Lemma 3, $u_1 = v_1$. Thus, $u_2 = v_2$ and $p = q$. \square

Note that in Lemma 5, $p \geq 2$ and $q \geq 2$ are essential conditions. For instance, $u_1 = aabb$, $u_2 = aa$, and $p = 2$ gives $u_1^p u_2 = aabbaabbaa$, and $v_1 = aabbaabba$, $v_2 = a$, and $q = 1$ gives $v_1^q v_2 = aabbaabbaa$; that is, $u_1^p u_2 = v_1^q v_2$.

As we often need to refer to the various occurrences of the same factor, we use a special subscript [1], [2], etc to distinguish them. For instance, $u_{[1]}$ may refer to the first occurrence of u in u^3 , while $u_{[2]}$ would refer to the second occurrence, etc.

Lemma 6 gives various contexts in which a balanced double square has a unique factorization. While a weaker form of Lemma 6 is proven in [11], and item (c) and the fact the U^2 must be primitively rooted are proven in [10], the uniqueness is not addressed in either.

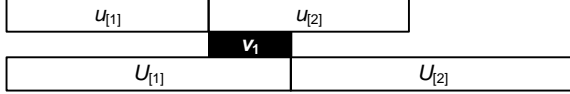
Lemma 6. Let (u, U) be a balanced double square. If one of the following conditions is satisfied

- (a) u is primitive
- (b) U is primitive

(c) u^2 has no further occurrence in U^2

then there is a unique primitive string u_1 , a unique non-trivial proper prefix u_2 of u_1 , and unique integers e_1 and e_2 satisfying $1 \leq e_2 \leq e_1$ such that $u = u_1^{e_1}u_2$ and $U = u_1^{e_1}u_2u_1^{e_2}$. Moreover, U is primitive.

Proof. Let v_1 denote the overlap of $U_{[1]}$ with $u_{[2]}$; that is, $u = v_1\bar{v}_1$ for some \bar{v}_1 and $U = uv_1$, see the diagram below.



Thus, u is a prefix of v_1U and $u = v_1^k v_2$ for some prefix v_2 of v_1 and $k \geq 1$. Let u_1 be the primitive root of v_1 . Then $v_1 = u_1^{e_2}$ for some $e_2 \geq 1$. Therefore $u = u_1^{e_1}u_2$ for some $e_1 \geq ke_2$ and some prefix u_2 of u_1 . The prefix u_2 must be non-trivial, as otherwise:

(a) Let us assume that u_2 is the empty string. If $e_1 \geq 2$, then $u = u_1^{e_1}$ and hence not primitive, a contradiction. If $e_1 = 1$, then $e_2 = 1$ and so $U = u_1^2$ and $u = u_1$ and so $|U| = 2|u|$, a contradiction.

(b) $U = u_1^{e_1+e_2}$ and $e_1+e_2 \geq 2$, hence U would not be primitive.

(c) there would be a farther occurrence of $u^2 = u_1^{2e_1}$ in $U^2 = u_1^{2e_1+2e_2}$.

To prove the uniqueness, consider some primitive w_1 , its non-trivial proper prefix w_2 , and integers $f_1 \geq f_2 \geq 1$ such that $u = w_1^{f_1}w_2$ and $U = w_1^{f_1}w_2w_1^{f_2}$. If $e_1 \geq 2$ and $f_1 \geq 2$, then by Lemma 5, $u_1 = w_1$ and $e_1 = f_1$ and it follows that $u_2 = w_2$ and $e_2 = f_2$. If $e_1 = f_1 = 1$, it follows that $u = u_1u_2 = w_1w_2$. Since $U = uu_1 = ww_1$, $u_1 = w_1$ and so $u_2 = w_2$. The remaining case corresponds to exactly one of the exponents e_1 and f_1 being equal to 1. Without loss of generality, we can assume that $e_1 = 1$ and $f_1 > 1$. We have $u = u_1u_2 = w_1^{f_1}w_2$ and $U = u_1u_2u_1 = w_1^{f_1}w_2w_1^{f_2}$. Thus, $u_1 = w_1^{f_2}$. As u_1 is primitive, $f_2 = 1$, and so $u_1 = w_1$. Therefore, $u_1u_2 = w_1^{f_1}w_2 = u_1^{f_1}w_2$ and so $f_1 = 1$, contradicting $f_1 > 1$.

Let us assume that U is not primitive and derive a contradiction. Thus, $U = v^n$ for some primitive v and some $n \geq 2$. It follows that $|v| \leq \frac{|U|}{2} = \frac{|u_1^{e_1}|+|u_2|+|u_1^{e_2}|}{2} \leq \frac{|u_1^{e_1}|+|u_2|+|u_1^{e_1}|+|u_2|}{2} = |u_1^{e_1}|+|u_2|$. Now consider $U^2 = v^{2n} = u_1^{e_1}u_2u_1^{e_1+e_2}u_2u_1^{e_2}$. It follows that $u_1^{e_1+e_2}u_2$ is a factor of v^{2n} , $2n \geq 2$ of size $\geq |v|+|u_1|$, $e_1+e_2 \geq 2$, and so by Lemma 3, u_1 and v are conjugates, hence $u_1 = v$. Thus $U = v^n = u_1^n = u_1^{e_1}u_2u_1^{e_1}$ and so $n|u_1| = (e_1+e_2)|u_1|+|u_2|$, which is impossible as $0 < |u_2| < |u_1|$. Therefore, U must be primitive. \square

Definition 7 (Notation and terminology). *If a balanced double square satisfies one of three conditions (a), (b), or (c) of Lemma 6, we will refer to such double square as factorizable. We use the following notational convention for factorizable double squares: a double square \mathcal{U} consists of two squares u^2 and U^2 , where $|u| < |U|$ and so we refer to u^2 respective U^2 as the shorter respective longer, square of \mathcal{U} , and to the starting position of u^2 and U^2 as the starting position of \mathcal{U} . The unique exponents are denoted as $u_{(1)}$ and $u_{(2)}$, the repeating primitive part of u is denoted as u_1 , the prefix of u_1 completing u is denoted as u_2 . Thus*

$u = u_1^{u(1)}u_2$ and $U = uu_1^{u(2)} = u_1^{u(1)}u_2u_1^{u(2)}$. Since u_2 is a non-trivial proper prefix of u_1 , there is complement \bar{u}_2 of u_2 in u_1 so that $u_1 = u_2\bar{u}_2$. The conjugate \bar{u}_2u_2 of u_1 is denoted as \tilde{u}_1 , i.e. $\tilde{u}_1 = \bar{u}_2u_2$.

For instance, a factorizable double square \mathcal{V} consists of the shorter square v^2 and the longer square V^2 , and $v = v_1^{v(1)}v_2$ and $V = v_1^{v(1)}v_2v_1^{v(2)}$. We would like to point out that for any factorizable double square \mathcal{U} , $|U^2| = 2((u(1)+u(2))|u_1|+|u_2|) \geq 2((1+1)2+1) = 10$ since $u(1) \geq u(2) \geq 1$, $|u_1| \geq 2$, and $|u_2| \geq 1$. Thus, only strings of length at least 10 may contain a factorizable double square. Note also, that by (c) of Lemma 6, every FS-double square is a factorizable double square. Lemma 8 further specifies the structure of a factorizable double square, i.e. the fact that the shorter and the longer squares must have essentially different structures.

Lemma 8. *If \mathcal{U} is a factorizable double square so that $u = v_1^i v_2$ for some primitive v_1 , some non-trivial proper prefix v_2 of v_1 , and some integer $i \geq 1$; then $U \neq v_1^j v_2$ for any $j \geq 1$.*

Proof. Clearly, $U \neq v_1^j v_2$ for $j \leq i$ since $|U| > |u|$. Thus, consider $j > i$ and assume by contradiction that $U = v_1^j v_2$. Then, for $j = i+1$, $U = uu_1^{u(2)} = v_1^i v_2 u_1^{u(2)} = v_1^{i+1} v_2$ and so $v_2 u_1^{u(2)} = v_1 v_2$. Denote by \bar{v}_2 the complement of v_2 in v_1 , i.e. $v_1 = v_2 \bar{v}_2$. Then $v_2 u_1^{u(2)} = v_2 \bar{v}_2 v_2$, and so $u_1^{u(2)} = \bar{v}_2 v_2$. Since $\bar{v}_2 v_2$ is a conjugate of v_1 and hence primitive, it follows that $u(2) = 1$ and thus $u_1 = \bar{v}_2 v_2$. Thus $U = v_1^{i+1} v_2 = v_2 (\bar{v}_2 v_2)^{i+1} = v_2 u_1^{i+1}$ and also $U = u_1^{u(1)} u_2 u_1$, so $u_1^{u(1)} u_2 u_1 = v_2 u_1^{i+1}$ contradicting Lemma 2 as $|v_2| < |v_1| = |u_1|$. For $j > i+1$, $v_1^i v_2 v_1$ must be a prefix of v_1^j contradicting Lemma 2 \square

Lemma 9 discusses the case when the shorter square of a factorizable double square is not primitively rooted. It shows that the size of U is highly constraint.

Lemma 9. *Let \mathcal{U} be a factorizable double square so that $u = v^k$, for some primitive v and some $k \geq 2$. Then $u(1) = u(2) = 1$ and $U = v^{2k-1} v_1$ for some non-trivial proper prefix v_1 of v . Moreover, $u_1 = v^{k-1} v_1$ and $v_1 u_2 = v$.*

Proof. Let us assume that $u(1) \geq 2$ and derive a contradiction. Then $u = u_1^{u(1)} u_2 = v^k$, giving $|u_1| < |v|$. It follows that $u_1^{u(1)} u_2$ and v^k have a common factor of length $\geq |u_1| + |v|$ and by Lemma 3, u_1 and v are conjugates, and so $u_1 = v$. But then $|u| = u(1)|u_1| + |u_2| = k|u_1|$, which is impossible as $0 < |u_2| < |u_1|$. Therefore, $u(1) = 1$ and so $u(2) = 1$.

Since U is a prefix of v^{2k} , $U = v^t v_1$ where $k \leq t \leq 2k-1$ and v_1 is a proper prefix of v . Since U must be primitive by Lemma 6, v_1 must be a non-trivial proper prefix. If $t = 2k-1$, then we are done and the proof is complete. Let us thus assume that $t < 2k-1$. Then $2k-t \geq 2$ and so the suffix v^{2k-t} of u^2 starts at the same position p as the suffix $v_1 U = v_1 v^t v_1$ of U^2 . Therefore factors v^2 (a subfactor of v^{2k-t}) and $v_1 v$ (a subfactor of $v_1 v^t v_1$) start at the same position p , contradicting Lemma 2 as v is primitive.

Since $U = uu_1$, $U = v^{2k-1} v_1 = v^k v^{k-1} v_1 = uv^{k-1} v_1$, and so $u_1 = v^{k-1} v_1$. Since $u = u_1 u_2$, $v^k = v^{k-1} v_1 u_2$ and so $v_1 u_2 = v$. \square

Definition 10. *A factor $u = x[i..j]$ of x can be cyclically shifted right by 1 position if $x[i] = x[j+1]$. The factor u can be cyclically shifted right by k positions if u can be cyclically*

shifted right by 1 position and the factor $x[i+1..j+1]$ can be cyclically shifted right by $k-1$ positions. Similarly for left cyclic shifts. By a trivial cyclic shift we mean a shift by 0 positions.

Note that if v is a right cyclic shift of u , then u and v are conjugates. Similarly for left cyclic shift.

Let x contain a factorizable double square \mathcal{U} and let $x = y_1 U^2 y_2$. To cyclically shift \mathcal{U} to the right means that both u^2 and U^2 must be cyclically shifted to the right. The maximal right cyclic shift of u^2 is determined by $lcp(u_1, \tilde{u}_1)$, while the maximal right cyclic shift of U^2 is determined by the $lcp(U^2, y_2)$, where $lcp(x, y)$ is the length of the *largest common prefix* of x and y . Similarly, to cyclically shift \mathcal{U} to the left means that both u^2 and U^2 must be cyclically shifted to the left. The maximal left cyclic shift of u^2 is determined by $lcs(u_1, \tilde{u}_1)$, while the maximal left cyclic shift of U^2 is determined by the $lcs(U^2, y_1)$, where $lcs(x, y)$ is the length of the *largest common suffix* of x and y . Thus, $lcs(u_1, \tilde{u}_1)$ represents the maximal potential left cyclic shift of u^2 , while $lcp(u_1, \tilde{u}_1)$ represents the maximal potential right cyclic shift of u^2 .

Lemma 11. *For any factorizable double square \mathcal{U} , $lcp(u_1, \tilde{u}_1) + lcs(u_1, \tilde{u}_1) \leq |u_1| - 2$.*

Proof. If $lcp(u_1, \tilde{u}_1) + lcs(u_1, \tilde{u}_1) \geq |u_1|$, then $u_1 = \tilde{u}_1$ contradicting the primitiveness of u_1 . So $lcp(u_1, \tilde{u}_1) + lcs(u_1, \tilde{u}_1) < |u_1|$. Assume then that $lcp(u_1, \tilde{u}_1) + lcs(u_1, \tilde{u}_1) = |u_1| - 1$. Let $i = lcp(u_1, \tilde{u}_1)$ and let a be the symbol at position i of u_1 , i.e. $u_1[i] = a$. Then $u_1[1..i-1] = \tilde{u}_1[1..i-1]$ as $|\{1..i-1\}| = lcp(u_1, \tilde{u}_1)$, and $u_1[i+1..|u_1|-1] = \tilde{u}_1[i+1..|u_1|-1]$ as $|\{i+1..|u_1|-1\}| = lcs(u_1, \tilde{u}_1)$. Thus, u_1 and \tilde{u}_1 coincide in all positions except possibly i . Therefore $u_1[1..i-1][i+1..|u_1|-1]$ and $\tilde{u}_1[1..i-1][i+1..|u_1|-1]$ must have the same number of a 's. Since u_1 and \tilde{u}_1 are conjugates, they both have to have the same number of a 's. Therefore $\tilde{u}_1[i] = a$ yielding $u_1 = \tilde{u}_1$, and thus contradicting the primitiveness of u_1 . \square

2.3 Inversion factors

A key combinatorial property of factorizable double squares is the highly constrained occurrences of so-called *inversion factors*. The notion of inversion factor is motivated by the two occurrences of the factor $\bar{u}_2 u_2 u_2 \bar{u}_2$ in a double square \mathcal{U} . Even though for the purpose of this paper it would be sufficient to define inversion factor as any cyclic shift of $\bar{u}_2 u_2 u_2 \bar{u}_2$ which would greatly simplify the proof of the correspondingly simplified Lemma 13, we decided to include a more general definition of inversion factor and thus a more general version of Lemma 13.

Definition 12. *Given a factorizable double square \mathcal{U} , a factor of U^2 of length $2|u_1|$ starting at position i is called inversion factor if*

$$\begin{cases} U^2[i+j] = U^2[i+j+|u_1|+|u_2|] & \text{for } 0 \leq j < |\bar{u}_2|, \text{ and} \\ U^2[i+j] = U^2[i+j+|u_2|] & \text{for } |\bar{u}_2| \leq j < |u_2|+|\bar{u}_2|. \end{cases}$$

Note that an inversion factor of \mathcal{U} has a form $\bar{v}_2 v_2 v_2 \bar{v}_2$ where $|v_2| = |u_2|$ and $|\bar{v}_2| = |\bar{u}_2|$.

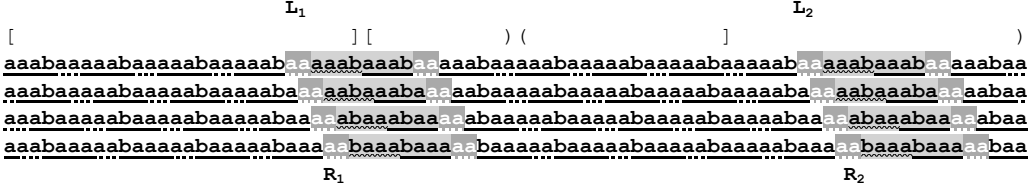


Figure 1: Cyclic shifts of the inversion factor and its environment

In a factorizable double square \mathcal{U} , inversion factors $\bar{u}_2 u_2 u_2 \bar{u}_2$ occur at positions $N_1(\mathcal{U})$ and $N_2(\mathcal{U})$ where

$$N_1(\mathcal{U}) = \epsilon(u_1^{u(1)-1} u_2) + 1 = (u(1)-1)|u_1| + |u_2| + 1$$

$$N_2(\mathcal{U}) = \epsilon(u_1^{u(1)} u_2 u_1^{u(2)+u(1)-1} u_2) + 1 = (2u(1)+u(2)-1)|u_1| + 2|u_2| + 1.$$

Such inversion factors are referred to as *natural*.

Cyclic shifts of the inversion factor $\bar{u}_2 u_2 u_2 \bar{u}_2$ are governed by the values of $lcp(u_1, \tilde{u}_1)$ and $lcs(u_1, \tilde{u}_1)$. A cyclic shift of an inversion factor is again an inversion factor. Thus, at every position of the union of the intervals $[L_1(\mathcal{U}), R_1(\mathcal{U})]$ and $[L_2(\mathcal{U}), R_2(\mathcal{U})]$ there is an inversion factor of \mathcal{U} starting there, where

$$L_1(\mathcal{U}) = \max \{ 1, N_1(\mathcal{U}) - lcs(u_1, \tilde{u}_1) \}$$

$$R_1(\mathcal{U}) = N_1(\mathcal{U}) + lcp(u_1, \tilde{u}_1)$$

$$L_2(\mathcal{U}) = N_2(\mathcal{U}) - lcs(u_1, \tilde{u}_1)$$

$$R_2(\mathcal{U}) = \min \{ \epsilon(U^2) - 2|u_1| + 1, N_2(\mathcal{U}) + lcp(u_1, \tilde{u}_1) \}.$$

If it is clear from the context, we omit the \mathcal{U} designation from $N_1(\mathcal{U})$, $N_2(\mathcal{U})$, $L_1(\mathcal{U})$, $R_1(\mathcal{U})$, $L_2(\mathcal{U})$, and $R_2(\mathcal{U})$. Note that $L_2 - L_1 = R_2 - R_1 = |U|$ and, by Lemma 11, $R_1 - L_1 = R_2 - L_2 \leq |u_1| - 2$. In addition, $L_1 \leq R_1 < \epsilon(u_{[1]}) < s(u_{[2]}) < \epsilon(U^2)$ and $\epsilon(u_{[1]}) < s(u_{[2]}) < L_2 \leq R_2 \leq \epsilon(U^2) - 2|u_1| < \epsilon(U^2)$. A key fact is that besides the intervals $[L_1, R_1]$ and $[L_2, R_2]$, there are no further occurrences of an inversion factor in a factorizable double square \mathcal{U} . In other words, all inversion factors are cyclic shifts of the natural ones.

See Figure 1 for an illustration where $u_2 = aaab$, $\bar{u}_2 = aa$, $u(1) = 4$, and $u(2) = 2$. Consequently, $u_1 = aaabaa$ and $\tilde{u}_1 = aaaaab$, and so $lcp(u_1, \tilde{u}_1) = 3$ and $lcs(u_1, \tilde{u}_1) = 0$. Thus, the inversion factor $\bar{u}_2 u_2 u_2 \bar{u}_2 = aaaaabaaabaa$ has three non-trivial right cyclic shifts and no non-trivial left cyclic shift. Note that there are no other inversion factors besides those highlighted. The configuration of brackets $[] []$ indicates the shorter square while the configuration $[) ()$ indicates the longer square. Also note that the environments of the inversion factors are shifted along: the inversion factor $\bar{v}_2 v_2 v_2 \bar{v}_2$ is always preceded by v_2 (solid underline) alternating with \bar{v}_2 (dotted underline). The leftmost piece of the environment, i.e. starting at the beginning of the string, might just be a suffix of v_2 or \bar{v}_2 . Similarly, the inversion factor $\bar{v}_2 v_2 v_2 \bar{v}_2$ is always followed by v_2 alternating with \bar{v}_2 . The rightmost piece of the environment, ending at the end of the string U^2 , might just be a prefix of v_2 or \bar{v}_2 .

Lemma 13 (Inversion factor Lemma). *An inversion factor of a factorizable double square \mathcal{U} within the string U^2 starts at a position i if and only if $i \in [L_1(\mathcal{U}), R_1(\mathcal{U})] \cup [L_2(\mathcal{U}), R_2(\mathcal{U})]$.*

The rather technical proof of Lemma 13 is given in Section 5.1.

3 Inversion factors and the problem of distinct squares

When computing the number of distinct squares, one must consider just one representative occurrence from all occurrences of each square. Fraenkel and Simpson [4] consider only the last, i.e., the rightmost occurrence. We consider the same context and thus will be investigating FS-double squares. Let us recall that FS-double squares are factorizable which follows from Lemma 6 (c). Fraenkel and Simpson's theorem states that at most two rightmost occurring squares can start at the same position using Lemma 14:

Lemma 14 (Crochemore and Rytter [1], Fraenkel and Simpson [4]). *Let u^2 , v^2 , and w^2 be squares in a string x starting at the same position such that $|u| < |v| < |w|$ and with u primitive, then $|w| \geq |u| + |v|$.*

Though one could prove Lemma 14 using the inversion factor Lemma 13, we follow Ilie [6] and prove Theorem 15 directly.

Theorem 15 (Fraenkel and Simpson [4], Ilie [6]). *At most two rightmost squares can start at the same position.*

Proof. Let us assume by contradiction that three rightmost squares start at the same position: u^2 , U^2 , and v^2 such that $|u| < |U| < |v|$. By item (c) of Lemma 6, u^2 and U^2 form a factorizable double square \mathcal{U} and so $u = u_1^{u(1)}u_2$ and $U = u_1^{u(1)}u_2u_1^{u(2)}$. Since $v_{[1]}$ contains an inversion factor, $v_{[2]}$ must also contain an inversion factor. If the inversion factor in $v_{[2]}$ were from $[L_2, R_2]$, then $|v| = |U|$, a contradiction. Hence $v_{[2]}$ must not contain an inversion factor from $[L_2, R_2]$ and so $u_1^{u(1)}u_2u_1^{u(1)+u(2)-1}u_2$ must be a prefix of v . Therefore $v_{[2]}$ contains another copy of $u_1^{u(1)}u_2u_1^{u(1)}u_2 = u^2$, contradicting the assumption that u^2 is a rightmost square. \square

We often need to investigate the mutual configuration of the shorter squares of two factorizable double squares.

Definition 16. *For two substrings u and v of a string x such that $s(u) < s(v)$, the gap $G(u, v)$ is defined as $s(v) - s(u)$ and the tail $T(u, v)$ is defined as $e(v) - e(u)$. For two factorizable double squares \mathcal{U} and \mathcal{V} such that $s(\mathcal{U}) < s(\mathcal{V})$, the gap $G(\mathcal{U}, \mathcal{V}) = G(u, v)$ and the tail $T(\mathcal{U}, \mathcal{V}) = T(u, v)$.*

Note that $T(u, v)$ could be negative when $e(v) < e(u)$. If $T(u, v) \geq 0$, then $G(u, v)v = uT(u, v)$. If it is clear from the context, we will drop the reference to u and v or \mathcal{U} and \mathcal{V} and use just G and T . Lemma 17 investigates configurations consisting of an FS-double square and a single rightmost square. In essence it says that if we have an FS-double square then the types and starting positions for a possible rightmost square v^2 are highly constraint. Lemma 17 is needed for Lemma 19 discussing configurations of two FS-double squares.

Lemma 17. *Let x be a string starting with an FS-double square \mathcal{U} . Let v^2 be a rightmost occurrence in x . Then*

(a) If $s(v_{[1]}) < R_1(\mathcal{U})$, then there are the following possibilities for v^2 :

- (a₁) $|v| < |u|$: in which case $v = \widehat{u}_1^j \widehat{u}_2$ for some $1 \leq j < u(1)$ where \widehat{u}_2 is a non-trivial proper prefix of \widehat{u}_1 and \widehat{u}_1 respective \widehat{u}_2 is a cyclic shift of u_1 respective u_2 by the same number of positions in the same direction;
- (a₂) $|v| = |u|$: in which case $v = \widehat{u}_1^{U(1)} \widehat{u}_2$ where \widehat{u}_2 is a non-trivial proper prefix of \widehat{u}_1 and \widehat{u}_1 respective \widehat{u}_2 is a cyclic shift of u_1 respective u_2 by the same number of positions in the same direction;
- (a₃) $|u| < |v| < |U|$: is impossible;
- (a₄) $|v| = |U|$: in which case $T(u, v) \geq 0$;
- (a₅) $|v| > |U|$: in which case $T(u, v) \geq 0$ and either $s_1 \bar{u}_2 u_2 u_1^{(U(1)+U(2)-1)} u_2$ is a prefix of v for some suffix s_1 of u_2 , or $s_1 u_1^i u_2 u_1^{(U(1)+U(2)-1)} u_2$ is a prefix of v for some suffix s_1 of u_1 and some $i \geq 1$.

(b) If $e(v_{[1]}) \leq e(u_{[1]})$, then $s(v_{[1]}) < R_1(\mathcal{U})$ and either (a₁) or (a₂) holds.

Definition 18 formalizes the types of relationship implied by Lemma 17.

Definition 18. We say that an FS-double square \mathcal{V} is a mate of an FS-double square \mathcal{U} in a string x , if $s(\mathcal{U}) < s(\mathcal{V})$.

1. \mathcal{V} is an α -mate of \mathcal{U} if $s(\mathcal{V}) \leq s(\mathcal{U}) + \text{lcp}(u_1, \tilde{u}_1)$ and \mathcal{V} is a right cyclic shift of \mathcal{U} .
2. \mathcal{V} is a β -mate of \mathcal{U} if $s(\mathcal{V}) < e(v_{[1]}) < e(u_{[1]})$ and $v = \widehat{u}_1^i \widehat{u}_2$ for some $1 < i < u(1)$ where \widehat{u}_2 is a non-trivial prefix of \widehat{u}_1 and where \widehat{u}_1 respective \widehat{u}_2 is a cyclic shift of u_1 respective u_2 in the same direction by the same number of positions, and V^2 is a right cyclic shift of U^2 by $s(\mathcal{V}) - s(\mathcal{U})$ positions.
3. \mathcal{V} is a γ -mate of \mathcal{U} if $s(\mathcal{V}) < s(\mathcal{U}) + u(1)|u_1|$ and $|v| = |U|$.
4. \mathcal{V} is a δ -mate of \mathcal{U} if $s(\mathcal{V}) < R_1(\mathcal{U})$ and $|v| > |U|$ and either $s_1 \bar{u}_2 u_2 u_1^{(U(1)+U(2)-1)} u_2$ is a non-trivial prefix of v for some suffix s_1 of u_2 , or $s_1 u_1^i u_2 u_1^{(U(1)+U(2)-1)} u_2$ is a non-trivial prefix of v for some s_1 suffix of u_1 and some $i \geq 1$.
5. \mathcal{V} is an ε -mate of \mathcal{U} if $R_1(\mathcal{U}) \leq s(\mathcal{V})$. If, in addition, $e(u_{[1]}) < s(\mathcal{V})$, we will call \mathcal{V} a super- ε -mate.

Note that Definition 18 implies that an α -mate of an α -mate of \mathcal{U} is an α -mate of \mathcal{U} ; an α -mate of a β -mate of \mathcal{U} is β -mate of \mathcal{U} ; a β -mate of a β -mate of \mathcal{U} is a β -mate of \mathcal{U} ; if \mathcal{V} is β -mate of \mathcal{U} , then $|U| = |V|$, $V = \widehat{u}_1^i \widehat{u}_2 \widehat{u}_1^{(U(1)+U(2)-i)}$, and $u(1) - u(2) \geq 2$ since $i \geq u(1) + u(2) - i$. If \mathcal{V} is a γ -mate of \mathcal{U} , then v^2 is right cyclic shift of U^2 .

Lemma 19. Let x be a string starting with an FS-double square \mathcal{U} . Let \mathcal{V} be an FS-double square with $s(\mathcal{U}) < s(\mathcal{V})$, then either

- (a) $\mathfrak{s}(\mathcal{V}) < R_1(\mathcal{U})$, in which case either
- (a₁) \mathcal{V} is an α -mate of \mathcal{U} , or
 - (a₂) \mathcal{V} is a β -mate of \mathcal{U} and $u_{(1)} > u_{(2)}+1$, or
 - (a₃) \mathcal{V} is a γ -mate of \mathcal{U} , or
 - (a₄) \mathcal{V} is a δ -mate of \mathcal{U} ,
- or
- (b) $R_1(\mathcal{U}) \leq \mathfrak{s}(\mathcal{V})$, then
- (b₁) \mathcal{V} is a ε -mate of \mathcal{U} and $\mathfrak{e}(v_{[1]}) > \mathfrak{e}(u_{[1]})$.

The rather technical proofs of Lemmas 17 and 19 are given, respectively, in Sections 5.2 and 5.3.

3.1 Some properties of γ -mates

Let an FS-double square \mathcal{V} be a γ -mate of an FS-double square \mathcal{U} . Then $v = s_2 u_1^{u_{(1)}-t-1} u_2 u_1^{u_{(2)}+t} s_1$ or $v = u_1^{u_{(1)}-t} u_2 u_1^{u_{(2)}+t}$ for some $u_{(1)}-t \geq 1$ and some s_1, s_2 so that $s_1 s_2 = u_1$. Let us define a *type* of \mathcal{V} :

$$\text{type}(\mathcal{V}) = \begin{cases} (u_{(1)}-t, u_{(2)}+t) & \text{if } v = u_1^{u_{(1)}-t} u_2 u_1^{u_{(2)}+t} \\ (u_{(1)}-t, u_{(2)}+t) & \text{if } s_2 u_1^{u_{(1)}-t-1} u_2 u_1^{u_{(2)}+t} s_1 \text{ and} \\ & |s_1| \leq |u_1| - lcs(u_1, \tilde{u}_1) \\ (u_{(1)}-t-1, u_{(2)}+t+1) & \text{otherwise.} \end{cases}$$

Though we do not know exactly what V^2 is like, we can still determine some of its properties.

Lemma 20. *Let an FS-double square \mathcal{V} be a γ -mate of an FS-double square \mathcal{U} of type (p, q) where $p, q \geq 2$ and $p+q \geq 4$. Then $\nu_{(1)} = \nu_{(2)}$ and $|v_2| \leq \min(p, q)|u_1|$. Moreover, either $|v_2| < |u_1|$ or there is a factor $(u_1^q u_2)(u_1^q u_2)$ in V^2 .*

Proof. Let us first assume that $v^2 = [u_1^p u_2 u_1^q][u_1^p u_2 u_1^q]$.

- (a) Let $p \geq q$.

By Lemma 2, the leftmost possible beginning of $V_{[2]}$ can be at $|u_1^p u_2 u_1^{p+q} u_2|+1$ and so $u_1^p u_2$ is a prefix of $v_1^{\nu_{(2)}}$ and v_2 is a factor of u_1^q . First we prove that $|v_1| > (p-1)|u_1|$:

Assume that $|v_1| \leq (p-1)|u_1|$. Then u_1^p contains a factor of size $|v_1|+|u_1|$ and the same factor is also contained in $v_1^{\nu_{(2)}}$ as $u_1^p u_2$ is a prefix of $v_1^{\nu_{(2)}}$. If $\nu_{(2)} \geq 2$, then by Lemma 3, $u_1 = v_1$ and so $u_1^p u_2$ is a prefix of $u_1^{\nu_{(2)}}$ and thus $u_1^p u_2 u_1$ is a prefix of $u_1^{\nu_{(2)}+1}$, which contradicts Lemma 2. Therefore $\nu_{(2)} = 1$ and so $|v_1| \geq p|u_1|+|u_2| > (p-1)|u_1|$, a contradiction with the assumption.

Hence $|v_1| > (p-1)|u_1| \geq q|u_1|$ and since v_2 is a factor in u_1^q , $\nu_{(1)} = \nu_{(2)}$.

If $V_{[2]}$ begins even more to the right, this makes v_2 smaller and $v_1^{\nu_{(2)}}$ bigger, thus the same argument can be applied.

(b) Let $p < q$

By Lemma 2 the leftmost possible beginning of $V_{[2]}$ can be at $|u_1^p u_2 u_1^{p+q} u_2 u_1^{q-p}|+1$ and so $u_1^p u_2 u_1^{q-p}$ is a prefix of $v_1^{\nu(2)}$ and v_2 is a factor of u_1^p . Let $r = \max(p, q-p)$. First we prove that $|v_1| > (r-1)|u_1|$:

Assume that $|v_1| \leq (r-1)|u_1|$. Then either u_1^p or u_1^{q-p} contains a factor of size $|v_1|+|u_1|$ and the same factor is also contained in $v_1^{\nu(2)}$ as $u_1^p u_2 u_1^{q-p}$ is a prefix of $v_1^{\nu(2)}$. If $\nu(2) \geq 2$, then by Lemma 3, $u_1 = v_1$ and so $u_1^p u_2 u_1^{q-p}$ is a prefix of $u_1^{\nu(2)}$, which contradicts Lemma 2. Therefore $\nu(2) = 1$ and so $|v_1| \geq q|u_1|+|u_2| > (r-1)|u_1|$, a contradiction with the assumption.

Hence $|v_1| > (r-1)|u_1| \geq p|u_1|$ and since v_2 is a factor in u_1^p , $\nu(1) = \nu(2)$.

If $V_{[2]}$ begins even more to the right, this makes v_2 smaller and $v_1^{\nu(2)}$ bigger, thus the same argument can be applied.

Let us thus assume that $v^2 = [s_2 u_1^{p-1} u_2 u_1^q s_1][s_2 u_1^{p-1} u_2 u_1^q s_2]$ and $|s_1| \leq |u_1| - \text{lcs}(u_1, \tilde{u}_1)$. Then $|s_2| > \text{lcs}(u_1, \tilde{u}_1)$.

(a) Let $p \geq q$.

By Lemma 2, the leftmost possible beginning of $V_{[2]}$ can be at $|s_2 u_1^{p-1} u_2 u_1^{p+q} u_2 s_1|+1$. If it started to the left of this point, by Lemma 2, s_2 would have to be a suffix of $u_1 u_2$ and so s_2 would be a common suffix of u_1 and \tilde{u}_1 , and so $|s_2| \leq \text{lcs}(u_1, \tilde{u}_1)$, a contradiction. Therefore the same arguments as in the case $v^2 = [u_1^p u_2 u_1^q][u_1^p u_2 u_1^q]$ can be applied.

(b) Let $p < q$

By Lemma 2 and by $|s_2| > \text{lcs}(u_1, \tilde{u}_1)$, the leftmost possible beginning of $V_{[2]}$ can be at $|s_2 u_1^p u_2 u_1^{p+q} u_2 u_1^{q-p} s_1|+1$. Again, if it started to the left of this point, by Lemma 2, s_2 would have to be a suffix of $u_1 u_2$ and so s_2 would be a common suffix of u_1 and \tilde{u}_1 , and so $|s_2| \leq \text{lcs}(u_1, \tilde{u}_1)$, a contradiction. Therefore, the same arguments as in the case $v^2 = [u_1^p u_2 u_1^q][u_1^p u_2 u_1^q]$ can be applied.

If $|v_2| \geq |u_1|$, then a prefix of $V_{[2]}$ must align with the last u_1 of $u_1^p u_2 u_1^{q+p} u_2 u_1^q$ and so $u_1^p u_2 u_1^{q+p} u_2 u_1^q$ is extended for sure by another u_2 , i.e. V^2 contains a factor $u_1^q u_2 u_1^q u_2$. \square

3.2 Some properties of ε -mates of \mathcal{U}

Lemma 21. *Let $\mathcal{U}, \mathcal{V}, \mathcal{W}$ be FS-double squares so that $s(\mathcal{U}) < s(\mathcal{V}) < s(\mathcal{W})$. Let \mathcal{V} be a γ -mate of \mathcal{U} of type $(u(1)-t, u(2)+t)$, $2 \leq p-t$ and $2 \leq q+t$, and let \mathcal{W} be an ε -mate but not a super- ε -mate of \mathcal{V} . Then $G(\mathcal{U}, \mathcal{W}) \geq t|u_1|$ and $T(\mathcal{U}, \mathcal{W}) \geq (u(1)+u(2))|u_1|$.*

Proof. The position of v^2 is:

$u_1^t s_1 [s_2 u_1^{u(1)-t-1} u_2 u_1^{u(2)+t} s_1] [s_2 u_1^{u(1)-t-1} u_2 u_1^{u(2)+t} s_1]$. Since \mathcal{V} is a γ -mate of \mathcal{U} , by Lemma 20 $\nu(1) = \nu(2)$ and so \mathcal{V} cannot have a β -mate, see Lemma 19. Thus $w_{[1]}$ must end past the end of $v_{[1]}$ and thus by Lemma 2, $|w| \geq |v|$. Therefore, $G \geq t|u_1|$ and $T \geq (u(1)+u(2))|u_1|$. \square

Lemma 22. *Let \mathcal{V} be a super- ε -mate of \mathcal{U} . Then either*

- (a) $G(\mathcal{U}, \mathcal{V}) \geq (2u^{(1)}+u^{(2)}-3)|u_1|+2|u_2|$ and
 $T(\mathcal{U}, \mathcal{V}) \geq (u^{(1)}+u^{(2)}-2)|u_1|+|u_2|$, or
- (b) $G(\mathcal{U}, \mathcal{V}) \geq u^{(1)}|u_1|+|u_2|$ and $T(\mathcal{U}, \mathcal{V}) \geq (u^{(1)}+u^{(2)}-1)|u_1|+|u_2|$.

Proof. If v^2 were a factor of $u_1^{u^{(1)}+u^{(2)}-1}u_2$, then there would be a farther copy of v^2 in $u_1^{u^{(1)}+u^{(2)}}u_2$ – just starting $|u_1|$ positions to the right, which is a contradiction as v^2 must be a rightmost occurrence. Hence $\epsilon(v^2) > |u_1^{u^{(1)}}u_2u_1^{u^{(1)}+u^{(2)}-1}u_2|$.

Let us assume that $v_{[1]}$ is a factor in $u_1^{u^{(1)}}u_2u_1^{(u^{(1)}+u^{(2)}-1)}u_2$.

Then $u_1^{(u^{(1)}+u^{(2)})}u_2$ and v^2 both contain a common factor of size $|v|+|u_1|$, and thus by Lemma 3, $v = v_1^k$ for some conjugate v_1 of u_1 and some $k \geq 1$. If $k = 1$, then $s(v_{[1]}) \geq |u_1^{u^{(1)}}u_2u_1^{(u^{(1)}+u^{(2)}-3)}u_2|$ and so $G \geq |u_1^{u^{(1)}}u_2u_1^{(u^{(1)}+u^{(2)}-3)}u_2|$. Moreover $s(v_{[2]}) = s(v_{[1]})+|u_1|$ and

$$\begin{array}{ccc} & \text{so} & T \\ |u_1^{(u^{(1)}+u^{(2)}-2)}u_2|, & & \geq \end{array}$$
i.e. (a) holds.

Let us assume that $k \geq 2$. We will discuss two cases:

- (i) $v_{[1]}$ starts in \bar{u}_2 and ends in \bar{u}_2 . Then there are $s_1s_2 = \bar{u}_2$ so that $v = (s_2u_2s_1)^k$ and so that v^2s_2 is a suffix of $u_1^{u^{(1)}}u_2u_1^{(u^{(1)}+u^{(2)})}$.

- (i₁) Let $|s_2| \leq lcs(u_1, \tilde{u}_1)$.

Then we can assume without loss of generality that $v = u_1^k$ as otherwise we can cyclically shift the whole structure $|s_2|$ positions to the left. By Lemma 9, $V = u_1^{2k-1}t_1$ for some non-trivial proper prefix t_1 of u_1 . Let $t_1t_2 = u_1$. Then the prefix u_1^3 of $V_{[2]}$ must align by Lemma 2 with $t_2u_1u_1$ and hence $t_2u_2 = u_1$. Therefore $|t_2| = |\bar{u}_2|$ and since t_2 is a suffix of $u_1 = u_2\bar{u}_2$, in fact $t_2 = \bar{u}_2$, Hence $u_1 = \bar{u}_2u_2$, a contradiction.

- (i₂) Let $|s_2| > lcs(u_1, \tilde{u}_1)$.

Then by Lemma 9, $V = (s_2u_2s_1)^{2k-1}t_1$ where t_1 is a non-trivial proper prefix of $s_2u_2s_1$. Let $t_1t_2 = s_2u_2s_1$. Then the prefix $(s_2u_2s_1)^3$ of $V_{[2]}$ must align by Lemma 2 with $t_2u_2u_2s_1s_2u_2s_1s_2u_2$ and so either $t_2u_2 = s_2$ or $t_2u_2 = s_su_2s_1s_2$. In either case, s_2 is a suffix of t_2u_2 and since s_2 is a suffix of \bar{u}_2 , s_2 is both a suffix of u_1 and of \tilde{u}_1 . Hence $|s_2| \leq lcs(u_1, \tilde{u}_1)$, a contradiction.

- (ii) $v_{[1]}$ starts in u_2 and ends in u_2 .

Then there are $s_1s_2 = u_2$ so that $v = (s_2\bar{u}_2s_1)^k$ and so that v^2s_2 is a suffix of $u_1^{u^{(1)}}u_2u_1^{(u^{(1)}+u^{(2)})}u_2$.

- (ii₁) Let $|s_2| \leq lcs(u_1, \tilde{u}_1)$.

Then without loss of generality we can assume $v = (\bar{u}_2u_2)^k$ and v^2 is a suffix of $u_1^{u^{(1)}}u_2u_1^{(u^{(1)}+u^{(2)})}u_2$ as otherwise we could cyclically shift the whole structure $|s_2|$ positions to the left. Then a suffix

$(\bar{u}_2u_2)(\bar{u}_2u_2)(\bar{u}_2u_2)(\bar{u}_2u_2)$ of v^2 must align with
 $(\bar{u}_2u_2)(\bar{u}_2u_2)(\bar{u}_2u_2)(u_2\bar{u}_2)(u_2\bar{u}_2)$ giving $\bar{u}_2u_2 = u_2\bar{u}_2$, a contradiction.

- (ii₂) Let $|s_2| > lcs(u_1, \tilde{u}_1)$.

Then $v = (s_2\bar{u}_2s_1)^k$ and by Lemma 9, $V = (s_2\bar{u}_2s_1)^{2k-1}t_1$ and $t_1t_2 = s_2\bar{u}_2s_1$. Then a prefix $(s_2\bar{u}_2s_1)^3$ of $V_{[2]}$ must align by Lemma 2 with $t_2s_1s_2\bar{u}_2s_1s_2\bar{u}_2$ and so $t_2 = s_2\bar{u}_2$. Since $t_1t_2 = s_2\bar{u}_2s_1$, then $t_1t_2s_2 = s_2\bar{u}_2s_1s_2 = s_2\bar{u}_2u_2$, i.e. $t_1t_2s_2 = s_2\tilde{u}_1$ and so s_2

is both a suffix of \tilde{u}_1 and a suffix of u_2 and hence of u_1 , and so $|s_2| \leq lcs(u_1, \tilde{u}_1)$, a contradiction.

Considering the end of v^2 in the next \bar{u}_2 will yield a contradiction using the same argumentation as for (i), and considering the end of v^2 in the next u_2 will yield a contradiction using the same argumentation as for (ii).

Thus, the only remaining case is when $v_{[1]}$ is not a factor in

$u_1^{u(1)}u_2u_1^{(u(1)+u(2)-1)}u_2$, i.e. $\epsilon(v_{[1]}) > u_1^{u(1)}u_2u_1^{(u(1)+u(2)-1)}u_2$ and so $G \geq |u_1^{u(1)}u_2|$ and $T \geq |u_1^{(u(1)+u(2)-1)}u_2|$, i.e. case (b) holds. \square

4 An upper bound for the number of FS-double squares

In this section, we only consider strings containing at least one FS-double square. Let $\delta(x)$ denote the number of FS-double squares in x . We prove by induction that $\delta(x) \leq \frac{7}{8}|x| - \frac{3}{8}|u|$ where u is the generator of the shorter square of the first FS-double square in x . We first need to investigate the relationship between two FS-double squares of x as the induction hypothesis is applied to the substring starting at some FS-double square and extended to the string starting with the first FS-double square.

Lemma 23. *Let x be a string starting with an FS-double square \mathcal{U} and let \mathcal{V} be another FS-double square of x with $\epsilon(u_{[1]}) \leq \epsilon(v_{[1]})$. Let x' be the suffix of x starting at the same position as \mathcal{V} . Let d be the number of FS-double squares between \mathcal{U} and \mathcal{V} including \mathcal{U} but not including \mathcal{V} . Then, $\delta(x') \leq \frac{5}{6}|x'| - \frac{1}{3}|v|$ implies $\delta(x) \leq \frac{5}{6}|x| - \frac{1}{3}|u| + d - \frac{1}{2}|G(\mathcal{U}, \mathcal{V})| - \frac{1}{3}|T(\mathcal{U}, \mathcal{V})|$.*

Proof. As $|G| + |v| = |u| + |T|$, we have $-\frac{1}{3}|v| = -\frac{1}{3}|u| - \frac{1}{3}|T| + \frac{1}{3}|G|$. Thus, $\delta(x) \leq d + \delta(x') \leq d + \frac{5}{6}|x'| - \frac{1}{3}|v| = d + \frac{5}{6}|x'| - \frac{1}{3}|u| - \frac{1}{3}|T| + \frac{1}{3}|G|$. Thus, $\delta(x) \leq \frac{5}{6}(|x'| + |G|) - \frac{1}{3}|u| + d - \frac{5}{6}|G| + \frac{1}{3}|G| - \frac{1}{3}|T| = \frac{5}{6}|x| - \frac{1}{3}|u| + d - \frac{1}{2}|G| - \frac{1}{3}|T|$ since $|x| = |x'| + |G|$. \square

Lemma 23 yields a straightforward induction step whenever $\frac{1}{2}|G| + \frac{1}{3}|T| \geq d$. By Lemma 19, this condition always holds except for the two cases: either \mathcal{V} is a right cyclic shift of \mathcal{U} by 1 position and hence an α -mate of \mathcal{U} , since then $\frac{1}{2}|G| + \frac{1}{3}|T| = \frac{1}{2} + \frac{1}{3} = \frac{5}{6} \neq 1$, or \mathcal{V} is a β -mate of \mathcal{U} and such that $\epsilon(v_{[1]}) < \epsilon(u_{[1]})$ – hence Lemma 23 is not applicable. Therefore the whole group of α -mates and β -mates of \mathcal{U} must be dealt together in the induction rather than carrying it from one FS-double square to another. Since a γ -mate of \mathcal{U} does not provide a sufficiently large tail to offset all of the α -mates and β -mates of \mathcal{U} preceding it, we have to include them in the special treatment as well – this is all precisely defined and explained in Section 4.1. First we need to strengthen the bound on the length of the maximal right cyclic shift of \mathcal{U} when $u(1) = u(2)$.

Lemma 24. *Let x be a string starting with an FS-double square \mathcal{U} such that $u(1) = u(2)$, i.e. $x = U^2y$ for some, possibly empty, y , then $\text{lcp}(u, y) < \min\{|y|, |u_2|\}$.*

Proof. Let f be the size of the \mathcal{U} family, then $f \leq |u_1|$. Let \mathcal{W} be the last member of the α -family of \mathcal{U} . Note that $\mathcal{W} = \mathcal{U}$ when the \mathcal{U} family consists only of \mathcal{U} . We apply Lemma 19 to \mathcal{W} and \mathcal{V} : since \mathcal{V} is neither an α -mate nor a β -mate of \mathcal{W} , then either it is a γ -mate or a δ -mate, or an ε -mate of \mathcal{W} . If it is a γ -mate or a δ -mate, then $|v| \geq |W|$ and so the size of the tail between \mathcal{W} and \mathcal{V} is at least $w^{(2)}|u_1|$. Since $w^{(2)} = u^{(2)} \geq 1$, the size of the tail is at least $|u_1|$. Therefore, the size of the gap G between \mathcal{U} and \mathcal{V} is at least f , the size of the tail T between \mathcal{U} and \mathcal{V} is at least $f + |u_1| \geq 2f$. Therefore, $\frac{1}{2}|G| + \frac{1}{3}|T| \geq \frac{1}{2}f + \frac{1}{3}2f = \frac{7}{6}f > f$. If \mathcal{V} is an ε -mate of \mathcal{W} , then the gap between \mathcal{W} and \mathcal{V} is at least u_1 and the tail exists. Hence, the gap between \mathcal{U} and \mathcal{V} is at least $f + |u_1| \geq 2f$ and the tail exists. Therefore, $\frac{1}{2}|G| + \frac{1}{3}|T| \geq \frac{1}{2}2f = f$. By Lemma 23, $\delta(x) \leq \frac{5}{6}|x| - \frac{1}{3}|u|$. \square

4.1.2 The case \mathcal{U} family consists of both α -mates and β -mates with no γ -mates

A \mathcal{U} family consisting entirely of α -mates and β -mates of \mathcal{U} is called an $(\alpha + \beta)$ -family and has the following structure:

- . The first so-called α -segment consists of \mathcal{U} and possibly its right cyclic shifts, i.e. its α -mates. The size of the segment is $\leq lcp(u_1, \tilde{u}_1) \leq |u_1| - 2$, see Lemma 11. All the FS-double squares in this segments have the first exponent equal to $u^{(1)}$ and the second exponent equal to $u^{(2)}$, thus we say that the type of the segment is $(u^{(1)}, u^{(2)})$.
- . Then there must be a β -mate of \mathcal{U} and possibly its right cyclic shifts. All the FS-double squares in the segment have the first exponent equal to $u^{(1)} - i_1$ and the second exponent equal to $u^{(2)} + i_1$ for some $1 \leq i_1 < (u^{(1)} - u^{(2)})/2$, thus we say that the type of the segment is $(u^{(1)} - i_1, u^{(2)} + i_1)$. This so-called β -segment has size $\leq lcp(u_1, \tilde{u}_1) \leq |u_1| - 2$ if $u^{(1)} - i_1 > u^{(2)} + i_1$, see Lemma 11, or $\leq |u_2| - 1 \leq |u_1| - 2$ if $u^{(1)} - i_1 = u^{(2)} + i_1$.
- . Then there may be another β -segment of type $(u^{(1)} - i_2, u^{(2)} + i_2)$ for some $1 \leq i_1 < i_2 < (u^{(1)} - u^{(2)})/2$, etc.
- . Either there is no other FS-double square in x , or the first FS-double square after the last member of the last β -segment must be either a δ -mate or an ε -mate of \mathcal{U} , since if it were a γ -mate, then the \mathcal{U} family would be an $(\alpha + \beta + \gamma)$ -family discussed in the following section.

There may be t such β -segments where $2t \leq u^{(1)} - u^{(2)}$. Let the last β -segment be of type $(u^{(1)} - t, u^{(2)} + t)$. If $u^{(1)} - t = u^{(2)} + t$ (which implies that $u^{(1)}$ is odd and $u^{(1)} - u^{(2)}$ is even), then $2t = u^{(1)} - u^{(2)}$ and there are $\leq (u^{(1)} - u^{(2)})/2$ segments of size $\leq |u_1|$ and 1 segment of size $\leq |u_2|$ and so the size of the family $f \leq \frac{u^{(1)} - u^{(2)}}{2}|u_1| + |u_2|$. If $u^{(1)} - t > u^{(2)} + t$, there are two cases, either $u^{(2)} = 1$ and then $f \leq \lceil \frac{u^{(1)} - u^{(2)}}{2} \rceil |u_1|$, or $u^{(2)} > 1$ and $f \leq \frac{u^{(1)} - u^{(2)}}{2}|u_1|$.

See Figure 4 for an illustration of an $(\alpha + \beta)$ -family where $u_2 = aaab$, $\bar{u}_2 = aa$, $u^{(1)} = 5$, and $u^{(2)} = 1$. The configuration of square brackets $[\] [\]$ indicates the shorter square while the configuration $[\] (\)$ indicates the longer square. The solid underline indicates u_2 while the dotted underline indicates \bar{u}_2 . The extension of U^2 is the final suffix not in bold. The FS-double square \mathcal{U} can be non-trivially cyclically shifted to the right by at

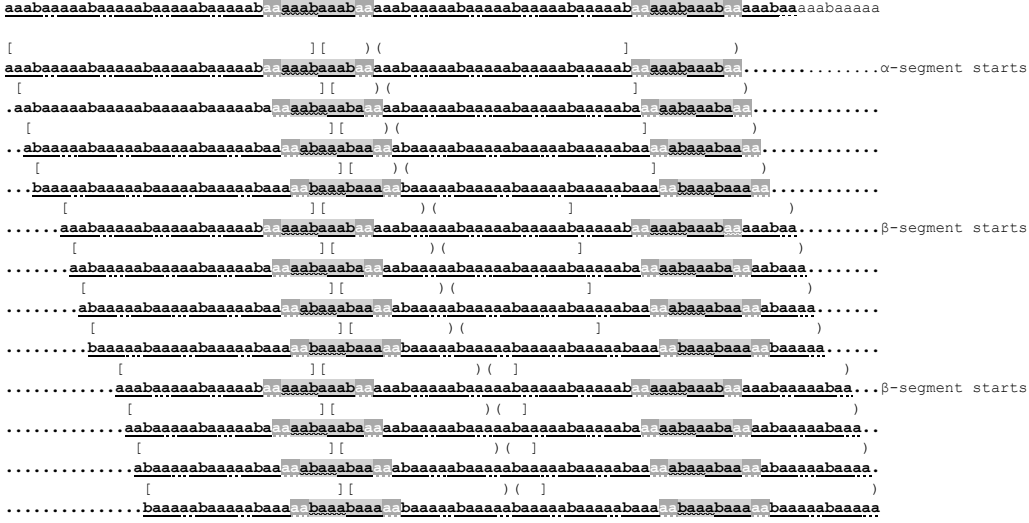


Figure 4: Example of an $(\alpha+\beta)$ -family of \mathcal{U}

most $\text{lcp}(u_1, \tilde{u}_1) = \text{lcp}(u_2 \bar{u}_2, \bar{u}_2 u_2) = \text{lcp}(aaabaa, aaaaab) = 3$ positions, thus every subfamily has at most 4 FS-double squares. Note, however, that the inversion factor $aaaaabaabaa$ – highlighted in Figure 4 – cyclically shifts within a subfamily and then returns to the original position for the first FS-double square of each segment. There is 1 α -segment and 2 β -segments since $(u(1) - u(2))/2 = 2$, t can take the 3 values 0, 1, or 2. For each new segment, the size of the shorter square decreases by a multiple of $|u_1|$ while the size of the longer square remains constant.

Claim 28. *Let x be a string starting with an $(\alpha+\beta)$ -family of an FS-double square \mathcal{U} and let \mathcal{V} be the last member of the \mathcal{U} family. Let every FS-double square \mathcal{W} after \mathcal{V} be so that $R_1(\mathcal{U}) \leq \text{s}(\mathcal{W}) \leq \text{e}(u_{[1]})$. Then $\delta(x) \leq \frac{5}{6}|x| - \frac{1}{3}|u|$.*

Proof. Let the type of \mathcal{V} be $(u(1)-t, u(2)+t)$. Then $2t \leq u(1)-u(2)$. Since every FS-double square \mathcal{W} after \mathcal{V} starts after R_1 but ends before $\text{e}(u_{[1]})$, the total number of FS-double squares in x is the number of FS-double squares in the \mathcal{U} family plus possibly $\leq |u_1|$ additional FS-double squares, i.e. $f \leq (t+2)|u_1|$. Since $|x| \geq |U^2| + f = 2(u(1)+u(2))|u_1| + 2|u_2| + f$, we get

$$\frac{5}{6}|x| - \frac{1}{3}|u| \geq \frac{5}{6}2(u(1)+u(2))|u_1| + \frac{5}{6}2|u_2| - \frac{1}{3}u(1)|u_1| - \frac{1}{3}|u_2| = \frac{4u(1)+5u(2)}{3}|u_1| + \frac{4}{3}|u_2| > \frac{4u(1)-4u(2)}{3}|u_1| + \frac{9u(2)}{3}|u_1| > \frac{8t}{3}|u_1| + 2|u_1| > t|u_1| + 2|u_1| \geq f = \delta(x). \quad \square$$

Claim 29. *Let x be a string starting with an $(\alpha+\beta)$ -family of an FS-double square \mathcal{U} and let there be some FS-double squares in x that are not members of the \mathcal{U} family. Let for any*

\mathcal{V} that is not a member of the \mathcal{U} family, $\delta(x') \leq \frac{5}{6}|x'| - \frac{1}{3}|v|$ where x' is a suffix of x starting at the same position as \mathcal{V} . Then $\delta(x) \leq \frac{5}{6}|x| - \frac{1}{3}|u|$.

Proof. Let the last β -segment be of type $(u_{(1)-t}, u_{(2)+t})$. Then $u_{(1)-t} \geq u_{(2)+t}$ and so $2t \leq u_{(1)} - u_{(2)}$ and the size of the \mathcal{U} family is $\leq (t+1)|u_1|$. By Lemma 19, \mathcal{V} is either a δ -mate, or a γ -mate, or a ε -mate of \mathcal{U} . Since \mathcal{U} family is an $(\alpha+\beta)$ -family, \mathcal{V} cannot be γ -mate of \mathcal{U} . The size of the \mathcal{U} family is $f \leq (t+1)|u_1|$.

Let us first discuss the case when \mathcal{V} is a δ -mate of \mathcal{U} . Then $T(\mathcal{U}, \mathcal{V}) \geq f$, $T(\mathcal{U}, \mathcal{V}) \geq (u_{(1)}+u_{(2)}-1)|u_1|+|u_2|$ and so $\frac{1}{2}|G|+\frac{1}{3}|T| > \frac{1}{2}f + \frac{u_{(1)}+u_{(2)}-1}{3}|u_1| > \frac{1}{2}f + \frac{u_{(1)}-u_{(2)}}{3}|u_1| + \frac{2u_{(2)}-1}{3}|u_1| \geq \frac{1}{2}f + \frac{2t}{3}|u_1| + \frac{1}{3}|u_1| > \frac{1}{2}f + \frac{2t+1}{3}|u_1| > \frac{1}{2}f + \frac{t+1}{2}|u_1| \geq \frac{1}{2}f + \frac{1}{2}f = f$. Thus, by Lemma 23, $\delta(x) \leq \frac{1}{2}|x| - \frac{1}{3}|u|$.

Let assume that \mathcal{V} is an ε -mate of \mathcal{U} .

If there were no super- ε -mate of \mathcal{U} , then by Lemma 28, $\delta(x) \leq \frac{5}{6}|x| - \frac{1}{3}|u|$. So let us assume that there is a super- ε -mate, and let \mathcal{V} be the first super- ε -mate of \mathcal{U} . Between the first ε -mate of \mathcal{U} and \mathcal{V} there are at most $|u_1|$ FS-double squares, $\delta(x) \leq \delta(x') + (t+2)|u_1|$. By the assumption of this lemma, $\delta(x') \leq \frac{1}{2}|x'| - \frac{1}{3}|v|$. By Lemma 22, there are two cases:

(a) $G(\mathcal{U}, \mathcal{V}) \geq (2u_{(1)}+u_{(2)}-3)|u_1|+2|u_2|$ and $T(\mathcal{U}, \mathcal{V}) \geq (u_{(1)}+u_{(2)}-3)|u_1|+|u_2|$.

$$\begin{aligned} \text{Since } u_{(2)} \geq 1 \text{ and } t \geq 2, \text{ then } \frac{1}{2}|G|+\frac{1}{3}|T| &> \frac{2u_{(1)}+u_{(2)}-3}{2}|u_1|+ \\ \frac{u_{(1)}+u_{(2)}-2}{3}|u_1| &= \frac{8u_{(1)}+5u_{(2)}-13}{6}|u_1| = \frac{8u_{(1)}-8u_{(2)}}{6}|u_1| + \frac{13u_{(2)}-13}{6}|u_1| > \\ \frac{16t}{6}|u_1| &= t|u_1| + \frac{10t}{6}|u_1| \geq t|u_1| + \frac{20}{6}|u_1| \geq t|u_1| + 2|u_1| \text{ as } t \geq 2. \end{aligned}$$

(b) $G(\mathcal{U}, \mathcal{V}) \geq u_{(1)}|u_1|+|u_2|$ and $T(\mathcal{U}, \mathcal{V}) \geq (u_{(1)}+u_{(2)}-1)|u_1|+|u_2|$.

$$\begin{aligned} \text{Then } \frac{1}{2}|G|+\frac{1}{3}|T| &> \frac{u_{(1)}}{2}|u_1| + \frac{u_{(1)}+u_{(2)}-1}{3}|u_1| = \frac{5u_{(1)}+2u_{(2)}-2}{6}|u_1| = \\ \frac{5u_{(1)}-5u_{(2)}}{6}|u_1| + \frac{7u_{(2)}-2}{6}|u_1| &\geq \frac{10t}{6}|u_1| + \frac{5}{6}|u_1| = t|u_1| + \frac{4t}{6}|u_1| + \frac{5}{6}|u_1| \geq t|u_1| + \\ \frac{8}{6}|u_1| + \frac{5}{6}|u_1| &= t|u_1| + \frac{13}{6}|u_1| > t|u_1| + 2|u_1| \text{ as } t \geq 2. \end{aligned}$$

□

4.1.3 The case \mathcal{U} -family consists of all three α -mates, β -mates, and γ -mates

We must first estimate the size of the family. We proceed by investigating its structure. Since there must be some β -mates of \mathcal{U} , $u_{(1)} \geq u_{(2)}+2$. The family consists of segments.

The first segment consists of \mathcal{U} and possibly its right cyclic shifts, i.e. its α -mates. The size of the segment is $\leq lcp(u_1, \tilde{u}_1) \leq |u_1|-2$, see Lemma 11. All the FS-double squares in this segments have the first exponent equal to $u_{(1)}$ and the second exponent equal to $u_{(2)}$, thus we say that the type of the segment is $(u_{(1)}, u_{(2)})$.

Then there must be a β -mate of \mathcal{U} and possibly its right cyclic shifts. All the FS-double squares in the segment have the first exponent equal to $u_{(1)}-i_1$ and the second exponent equal to $u_{(2)}+i_1$ for some $1 \leq i_1 < (u_{(1)}-u_{(2)})/2$, thus we say that the type of the segment

\mathcal{V} containing a farther copy of $u_1^r u_2 u_1^r u_2$ and thus implying that though there is a structure of a double square of type (r, r') , it is not an FS-double square, we will say that \mathcal{V} replaces the double square structure of type (r, r') .

Now back to estimating the size f of an $(\alpha+\beta+\gamma)$ -family. We shall show that $f \leq \frac{2}{3}(u_{(1)+1})|u_1|$. There are basically two cases:

- (i) \mathcal{G} , the first member of the γ -segment, is of type $(u_{(1)}-t, u_{(2)}+t)$ and $u_{(1)}-t > 2(u_{(2)}+t)$. Since $u_{(1)}-t > 2(u_{(2)}+t)$, $3t < u_{(1)}-2u_{(2)}$ and so $3t \leq u_{(1)}-2u_{(2)}-1$ and thus $6t \leq 2u_{(1)}-4u_{(2)}-2$. By Lemma 20 and Lemma 24, \mathcal{G} has $\leq (u_{(2)}+t)-1$ cyclic shifts. Thus, we start with \mathcal{U} of type $(u_{(1)}, u_{(2)})$ and end with the last member of the γ -segment that is of type $(u_{(1)}-t-(u_{(2)}+t-1), (u_{(2)}+t+(u_{(2)}+t-1)))$, thus there are at most $(2u_{(2)}+2t-1)-u_{(2)}+1 = u_{(2)}+2t$ members in the $(\alpha+\beta+\gamma)$ -family. Then $3f = 3u_{(2)}+6t \leq 3u_{(2)}+2u_{(1)}-4u_{(2)}-2 = 2u_{(1)}-u_{(2)}-2 \leq 2u_{(1)}-3 < 2u_{(1)}+2 = 2(u_{(1)+1})$ as $q \geq 1$. Thus, $f < \frac{2}{3}(u_{(1)+1})|u_1|$.
- (ii) \mathcal{G} , the first member of the γ -segment, is of type $(u_{(1)}-t, u_{(2)}+t)$ and $u_{(1)}-t \leq 2(u_{(2)}+t)$.

- (ii₁) $u_{(1)}-t \leq u_{(2)}+t$

By Lemma 20, G^2 of \mathcal{G} contains a further copy of $u_1^{u_{(2)}+t} u_2 u_1^{u_{(2)}+t} u_2$ and so \mathcal{G} either “replaces” a possible member of the α -segment or a β -segment, or it “complements” the α -segment or a β -segment. Thus, $f \leq \frac{1}{2}(u_{(1)}-u_{(2)})|u_1| < \frac{2}{3}(u_{(1)+1})|u_1|$.

- (ii₂) $u_{(1)}-t > u_{(2)}+t$.

Either g_2 of \mathcal{G} is small, i.e. $|g_2| < |u_1|$ and then \mathcal{G} has less than $|u_1|$ shifts, and so $f \leq \frac{1}{2}(u_{(1)}-u_{(2)})|u_1|+|u_1| \leq \frac{2}{3}(u_{(1)+1})|u_1|$, or $|g_2| \geq |u_1|$.

Thus assume that $|g_2| \geq |u_1|$. We can further assume by Lemma 20 that the last member of the γ -segment is of type $(u_{(2)}+t, u_{(1)}-t)$, since if it were shifted any further, it would start “replacing” or “completing” the members of the α -segment or the β -segments, so we do not need to count them.

Since $u_{(1)}-t \leq 2(u_{(2)}+t)$, then $u_{(1)}-2u_{(2)} \geq 3t$. Thus $3f =$

$$\begin{aligned} & 3(u_{(1)}-t-u_{(2)}-1)|u_1| = (3u_{(1)}-3t-3u_{(2)}+3)|u_1| \leq \\ & (3u_{(1)}-3u_{(2)}+3+2u_{(2)}-u_{(1)})|u_1| = (2u_{(1)}-u_{(2)}+3)|u_1| \leq \\ & (2u_{(1)}+2)|u_1| = 2(u_{(1)+1})|u_1|. \end{aligned}$$

Therefore, $f \leq \frac{2}{3}(u_{(1)+1})|u_1|$.

Claim 30. *Let a string x start with an $(\alpha+\beta+\gamma)$ -family of an FS-double square \mathcal{U} and let there be no other FS-double squares. Then $\delta(x) \leq \frac{5}{6}|x|-\frac{1}{3}|u|$.*

Proof. The size of the family $f \leq \frac{2}{3}(u_{(1)+1})|u_1|$ and so $\frac{1}{6}f \leq \frac{2}{18}(u_{(1)+1})|u_1|$. $|x| \geq f+|U^2| = f+2(u_{(1)}+u_{(2)})|u_1|+2|u_2|$, and so $\frac{5}{6}|x|-\frac{1}{3}|u| \geq \frac{5}{6}f+$
 $\frac{5}{6}(2(u_{(1)}+u_{(2)})|u_1|+\frac{5}{6}2|u_2|-\frac{1}{3}u_{(1)}|u_1|-\frac{1}{3}|u_2|) = \frac{5}{6}f+\frac{8}{6}p|u_1|+\frac{10}{6}u_{(2)}|u_1|+\frac{3}{6}|u_2| > \frac{5}{6}f+\frac{30}{18}p|u_1| >$
 $\frac{5}{6}f+\frac{2}{18}p|u_1|+\frac{28}{18}p|u_1| \geq \frac{5}{6}f+\frac{2}{18}(p+1)|u_1| \geq \frac{5}{6}f+\frac{1}{6}f = f = \delta(x)$. \square

Claim 31. *Let a string x start with an $(\alpha+\beta+\gamma)$ -family of an FS-double square \mathcal{U} . Let \mathcal{V} be the first FS-double square not in the \mathcal{U} family. Let x' be the suffix of x starting at the same position as \mathcal{V} . Let $\delta(x') \leq \frac{5}{6}|x'|-\frac{1}{3}|v|$. Then $\delta(x) \leq \frac{5}{6}|x|-\frac{1}{3}|u|$.*

Proof. \mathcal{V} can be either a δ -mate or ε -mate of \mathcal{U} . Let \mathcal{G} be the last member of the γ -segment and let its type be $(u^{(1)}-t, u^{(2)}+t)$. Then g^2 has the format $u_1^t s_1 [s_2 u_1^{(u^{(1)}-t-1)} u_2 u_1^{u^{(2)}} s_1] [s_2 u_1^{(u^{(1)}-t-1)} u_2 u_1^{u^{(2)}} s_1]$. If $\varepsilon(v_{[1]}) \leq \varepsilon(g_{[1]})$, then by Lemma 19 \mathcal{U} would be a β -mate of \mathcal{G} , which is impossible as by Lemma 20, $g^{(1)} = g^{(2)}$. Thus $\varepsilon(v_{[1]}) > \varepsilon(g_{[1]})$.

(a) Let \mathcal{V} be a δ -mate.

Then we are assured that $T(\mathcal{U}, \mathcal{V}) \geq (u^{(1)}+u^{(2)}-1)|u_1|$. But a little bit more is true. Clearly, $v_{[1]}$ contains an inversion factor from $[L_1(\mathcal{U}), R_1(\mathcal{U})]$. If $\varepsilon(v_{[2]}) \leq R_2(\mathcal{U})$, then $v_{[2]}$ would contain an inversion factor from $[L_2(\mathcal{U}), R_2(\mathcal{U})]$, giving $|v| = |w|$, a contradiction. Hence $\varepsilon(v_{[2]}) > R_2(\mathcal{U})$ and by Lemma 2, $T(\mathcal{U}, \mathcal{V}) \geq (u^{(1)}+u^{(2)})|u_1|$. Since $G(\mathcal{U}, \mathcal{V}) \geq f$, we have $\frac{1}{2}|G| + \frac{1}{3}|T| \geq \frac{1}{2}f + \frac{1}{3}(u^{(1)}+u^{(2)})|u_1| \geq \frac{1}{2}f + \frac{1}{3}(u^{(1)}+1)|u_1| \geq \frac{1}{2}f + \frac{1}{2}f = f$ as $u^{(2)} \geq 1$ and $\frac{1}{2}f \leq \frac{1}{3}(u^{(1)}+1)|u_1|$.

(b) Let \mathcal{V} be an ε -mate of \mathcal{U} , but not a super- ε -mate.

So $\varepsilon(v_{[1]}) \leq \varepsilon(u_{[1]})$ and $\varepsilon(v_{[1]}) > \varepsilon(g_{[1]})$. By Lemma 2, $T(\mathcal{U}, \mathcal{V}) \geq (u^{(1)}+u^{(2)})|u_1|$ and so $\frac{1}{2}|G| + \frac{1}{3}|T| \geq \frac{1}{2}f + \frac{1}{3}(u^{(1)}+1)|u_1| \geq \frac{1}{2}f + \frac{1}{2}f = f$.

(c) Let \mathcal{V} be a super- ε -mate of \mathcal{U} .

By Lemma 22, there are two possibilities:

- (c₁) $G \geq (2u^{(1)}+u^{(2)}-3)|u_1|$ and $T \geq (u^{(1)}+u^{(2)}-2)|u_1|$
Then $\frac{1}{2}|G| + \frac{1}{3}|T| \geq \frac{6u^{(1)}+3u^{(2)}-9+2u^{(1)}+2u^{(2)}-4}{6}|u_1| = \frac{8u^{(1)}+5u^{(2)}-13}{6}|u_1| = \frac{4u^{(1)}+4u^{(1)}+5u^{(2)}-13}{6}|u_1|$. Since $u^{(1)} \geq 4$ and $u^{(2)} \geq 1$, $\frac{1}{2}|G| + \frac{1}{3}|T| \geq \frac{4u^{(1)}+16+5-13}{6}|u_1| = \frac{4u^{(1)}+8}{6}|u_1| > \frac{4u^{(1)}+4}{6}|u_1| = f$.
- (c₂) $G \geq u^{(1)}|u_1|$ and $T \geq (u^{(1)}+u^{(2)}-1)|u_1|$
 $\frac{1}{2}|G| + \frac{1}{3}|T| \geq \frac{3u^{(1)}+2u^{(1)}+2u^{(2)}-2}{6}|u_1| = \frac{2u^{(1)}+3u^{(1)}+2u^{(2)}-2}{6}|u_1| \geq \frac{2u^{(1)}+16+2-2}{6}|u_1| = \frac{2u^{(1)}+12}{6}|u_1| > \frac{2u^{(1)}+2}{16}|u_1| \geq f$, since $u^{(1)} \geq 4$ and $u^{(2)} \geq 1$.

□

4.2 New upper bounds

Theorem 32. *The number of FS-double squares in a string of length n is bounded by $\lfloor 5n/6 \rfloor$.*

Proof. We prove by induction the following, a slightly stronger, statement: $\delta(x) \leq \frac{5}{6}|x| - \frac{1}{3}|u|$ for $|x| \geq 10$ where u is the generator of the shorter square of the first FS-double square of x . We do not have to consider strings of length 9 or less, as such strings do not contain FS-double squares. Since a string of length 10 contains at most one FS-double square (see the note after Definition 7), the statement is true for strings of size 10. Assuming the statement is true for all $|x| \leq n$, we shall prove it holds for all $|x| \leq n+1$.

If $x = x[1..n+1]$ does not start with an FS-double square, then $\delta(x) = \delta(x[2..n+1]) \leq \frac{5}{6}|x[2..n+1]| - \frac{1}{3}|u| \leq \frac{5}{6}|x[1..n+1]| - \frac{1}{3}|u|$. Thus, we can assume that x starts with an FS-double square \mathcal{U} . If \mathcal{U} is the only FS-double square of x , then $|x| \geq 2|u|$, thus the statement is obviously true. Therefore, we can assume that x starts with a FS-double square \mathcal{U} and $\delta(x) \geq 2$.

Case (a) assume that x starts with an α -family of \mathcal{U} .

If there is no further FS-double square in x , by Claim 26, the assertion is true. Otherwise, we carry out the induction step by Claim 27.

Case (b) assume that x starts with an $(\alpha+\beta)$ -family of \mathcal{U} .

If there is no further FS-double square in x , by Claim 28, the assertion is true. Otherwise, we carry out the induction step by Claim 29.

Case (c) assume that x starts with an $(\alpha+\beta+\gamma)$ -family of \mathcal{U} .

If there is no further FS-double square in x , by Claim 30, the assertion is true. Otherwise, we carry out the induction step by Claim 31. \square

Corollary 33. *The number of distinct squares in a string of length n is bounded by $\lfloor 11n/6 \rfloor$.*

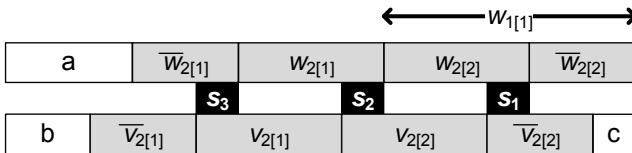
Proof. The number of distinct squares in a string is the sum of the number of FS-double squares plus the number of single rightmost squares. Since, for a string of length n , the number of FS-double squares is bounded by $\lfloor 5n/6 \rfloor$, the number of distinct squares is bounded by $\lfloor (2 \cdot 5/6 + 1/6)n \rfloor$; that is, by $\lfloor 11n/6 \rfloor$. \square

5 Proofs

5.1 Proof of Lemma 13:

Assume, in order to derive a contradiction, that an inversion factor $\bar{v}_2 v_2 v_2 \bar{v}_2$ occurs to the left of L_1 . Consider the inversion factor $\bar{w}_2 w_2 w_2 \bar{w}_2$ starting at the position L_1 . Then w_1 , w_2 and \bar{w}_2 are left cyclic shifts of, respective u_1 , u_2 and \bar{u}_2 . Since $\bar{w}_2 w_2 w_2 \bar{w}_2$ cannot be further cyclically shifted to the left, $lcs(w_2 \bar{w}_2, \bar{w}_2 w_2) = 0$. Since $\bar{v}_2 v_2 v_2 \bar{v}_2$ is occurring to the left of $\bar{w}_2 w_2 w_2 \bar{w}_2$, there are non-empty strings a and c and a string b so that $|a| > |b|$ and $a \bar{w}_2 w_2 w_2 \bar{w}_2 = b \bar{v}_2 v_2 v_2 \bar{v}_2 c$. We split the argument into several cases depending on where the inversion factor $\bar{v}_2 v_2 v_2 \bar{v}_2$ ends.

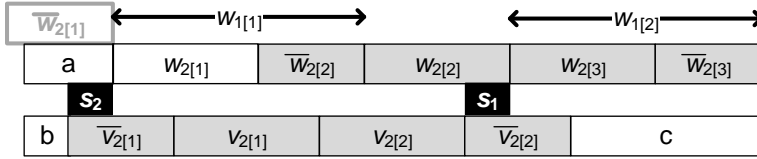
1. Case when $\bar{v}_2 v_2 v_2 \bar{v}_2$ ends in the second copy of \bar{w}_2 in the inversion factor $\bar{w}_2 w_2 w_2 \bar{w}_2$:



Let s_1 be the overlap of $w_{2[2]}$ and $\bar{v}_{2[2]}$. Then s_1 is a non-trivial proper prefix of \bar{v}_2 and a non-trivial proper suffix of w_2 . There is a copy s_2 of s_1 as a suffix of $w_{2[1]}$, and it must

be a prefix of $v_{2[2]}$ as $|w_2| = |v_2|$. Consequently, there is a copy s_3 of s_2 as a prefix of $v_{2[1]}$, and it must be a suffix of $\bar{w}_{2[1]}$ as $|s_3| = |s_1| \leq |\bar{w}_2|$ and $|v_{2[1]}| + |v_{2[2]}| + |\bar{v}_{2[2]}| = |w_{2[1]}| + |w_{2[2]}| + |\bar{w}_{2[2]}|$. Thus, s_1 is a suffix of both w_2 and of \bar{w}_2 , contradicting the fact that $lcs(w_2\bar{w}_2, \bar{w}_2w_2) = 0$.

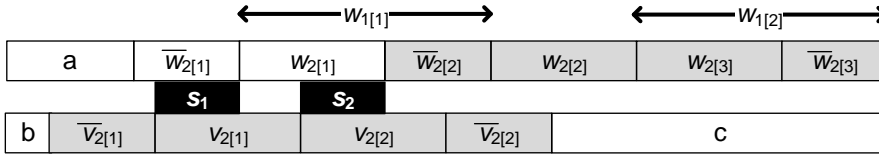
2. Case when $\bar{v}_2v_2v_2\bar{v}_2$ ends in the second copy of w_2 of $\bar{w}_2w_2w_2\bar{w}_2$:



Let s_1 be the overlap of $w_{2[2]}$ and $\bar{v}_{2[2]}$. Then s_1 is suffix of w_2 and a prefix of \bar{v}_2 . There is a copy s_2 of s_1 as a prefix of $\bar{v}_{2[1]}$. Consequently, s_2 must be a suffix of $\bar{w}_{2[1]}$ since $|\bar{v}_{2[1]}| + |v_{2[1]}| + |v_{2[2]}| = |w_{2[1]}| + |\bar{w}_{2[2]}| + |w_{2[2]}|$. Thus, s_1 is a suffix of both w_2 and \bar{w}_2 , contradicting the fact that $lcs(w_2\bar{w}_2, \bar{w}_2w_2) = 0$.

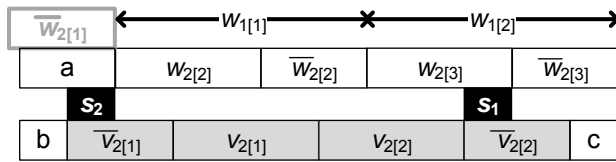
Note that the whole of $\bar{w}_{2[1]}$ might not be a part of the string (and that is why in the diagram it is depicted in gray), in which case a is a non-trivial proper suffix of $\bar{w}_{2[1]}$, and the argument holds.

3. Case when $\bar{v}_2v_2v_2\bar{v}_2$ ends in the first copy of w_2 of $\bar{w}_2w_2w_2\bar{w}_2$:



Let s_1 be the overlap of $\bar{w}_{2[1]}$ and $v_{2[1]}$. Then s_1 is a suffix of \bar{w}_2 and a prefix of v_2 . There is a copy s_2 of s_1 as a prefix of $v_{2[2]}$. It must be a suffix of $w_{2[1]}$ since $|\bar{v}_{2[1]}| + |v_{2[1]}| = |\bar{w}_{2[1]}| + |w_{2[1]}|$. Thus, s_1 is a suffix of both w_2 and \bar{w}_2 , contradicting the fact that $lcs(w_2\bar{w}_2, \bar{w}_2w_2) = 0$.

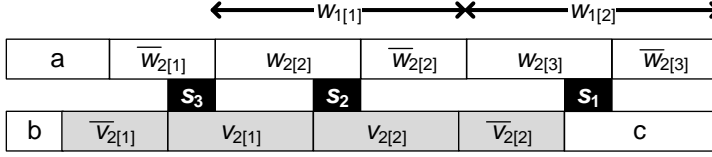
4. Case when $\bar{v}_2v_2v_2\bar{v}_2$ ends in the first copy of \bar{w}_2 of $w_2\bar{w}_2, \bar{w}_2w_2$, or lies completely outside of $w_2w_2\bar{w}_2$ of $\bar{w}_2w_2w_2\bar{w}_2$ and ends in \bar{w}_2 :



Let s_1 be the overlap of $w_{2[3]}$ and $\bar{v}_{2[2]}$. Then s_1 is a suffix of w_2 and a prefix of \bar{v}_2 . There is a copy s_2 of s_1 as a prefix of $\bar{v}_{2[1]}$. It must be a suffix of $\bar{w}_{2[1]}$ as $|w_{2[2]}| + |\bar{w}_{2[2]}| + |w_{2[3]}| = |\bar{v}_{2[1]}| + |v_{2[1]}| + |v_{2[2]}|$. Thus, s_1 is a suffix of both w_2 and \bar{w}_2 , contradicting the fact that $lcs(w_2\bar{w}_2, \bar{w}_2w_2) = 0$.

Note that the whole of $\bar{w}_{2[1]}$ might not be a part of the string (and that is why in the diagram it is depicted in gray), in which case a is a non-trivial proper suffix of $\bar{w}_{2[1]}$, and the argument holds.

5. Case when $\bar{v}_2v_2v_2\bar{v}_2$ lies completely outside of $\bar{w}_2w_2w_2\bar{w}_2$ and ends in w_2 :

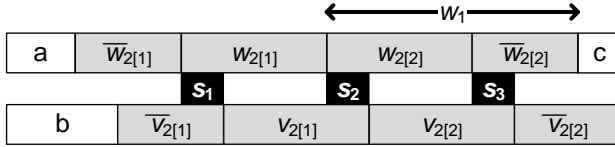


Let s_1 be the offset of $a\bar{w}_2w_2w_2\bar{w}_2$ and $b\bar{v}_2v_2v_2\bar{v}_2$, i.e. $a\bar{w}_2w_2w_2\bar{w}_2 = b\bar{v}_2v_2v_2\bar{v}_2s_1$. Then s_1 is a suffix of w_2 . There is a copy s_2 of s_1 as a suffix of $\bar{w}_2[2]$. It must be a prefix of $\bar{v}_2[2]$ as $|\bar{w}_2[2]|+|w_2[3]| = |v_2[2]|+|\bar{v}_2[2]|$. There is a copy s_3 of s_1 as a prefix of $v_2[1]$. It must be a suffix of $\bar{w}_2[1]$ as $|w_2[2]|+|\bar{w}_2[2]|+|w_2[3]| = |v_2[1]|+|v_2[2]|+|\bar{v}_2[2]|$. Thus, s_1 is a suffix of both w_2 and \bar{w}_2 , contradicting the fact that $lcs(w_2\bar{w}_2, \bar{w}_2w_2) = 0$.

As a second step of the proof, let us investigate whether an inversion factor $\bar{v}_2v_2v_2\bar{v}_2$ can occur to the right of R_1 while ending before L_2 . The proof of this step is essentially the same argumentation as for the first one, so though added for the sake of completion, it is presented in an abbreviated form, i.e. we just present the diagrams and the conclusions.

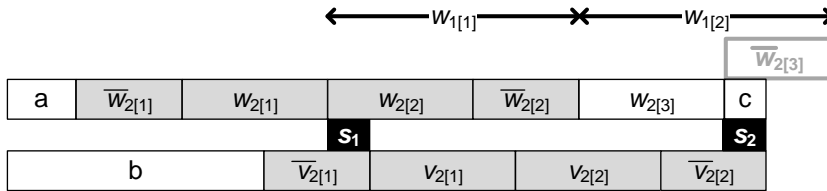
Consider the inversion factor $\bar{w}_2w_2w_2\bar{w}_2$ starting at the position R_1 . Then w_1 respective w_2, \bar{w}_2 are right cyclic shifts of u_1 respective u_2, \bar{u}_2 . Moreover, $lcp(w_2\bar{w}_2, \bar{w}_2w_2) = 0$ as $\bar{w}_2w_2w_2\bar{w}_2$ cannot be shifted right. Since $\bar{v}_2v_2v_2\bar{v}_2$ is occurring to the right of $\bar{w}_2w_2w_2\bar{w}_2$, there are non-empty strings b and c and a string a so that $|a| < |b|$ and $a\bar{w}_2w_2w_2\bar{w}_2c = b\bar{v}_2v_2v_2\bar{v}_2$. We split the argument into several cases depending on where the inversion factor $\bar{v}_2v_2v_2\bar{v}_2$ starts.

1. Case when $\bar{v}_2v_2v_2\bar{v}_2$ starts in the first copy of \bar{w}_2 in the inversion factor $\bar{w}_2w_2w_2\bar{w}_2$:



Then s_1 is both a prefix of w_2 and \bar{w}_2 , contradicting the fact that $lcp(w_2\bar{w}_2, \bar{w}_2w_2) = 0$.

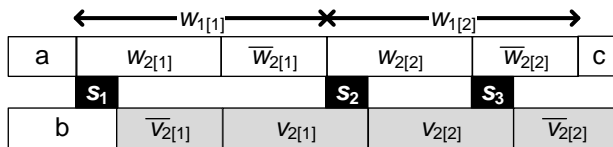
2. Case when $\bar{v}_2v_2v_2\bar{v}_2$ starts in the first copy of w_2 in the inversion factor $\bar{w}_2w_2w_2\bar{w}_2$:



Then s_1 is both a prefix of w_2 and \bar{w}_2 , contradicting the fact that $lcp(w_2\bar{w}_2, \bar{w}_2w_2) = 0$.

3. Case when $\bar{v}_2v_2v_2\bar{v}_2$ starts in the w_2 of w_1 :

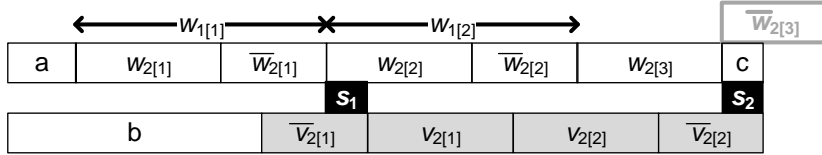
Note that this covers also the case when $\bar{v}_2v_2v_2\bar{v}_2$ starts in the second copy of w_2 in $\bar{w}_2w_2w_2\bar{w}_2$.



Then s_1 is both a prefix of w_2 and \bar{w}_2 , contradicting the fact that $\text{lcp}(w_2\bar{w}_2, \bar{w}_2w_2) = 0$.

4. Case when $\bar{v}_2v_2v_2\bar{v}_2$ starts in the \bar{w}_2 of w_1 :

Note that this covers also the case when $\bar{v}_2v_2v_2\bar{v}_2$ starts in the second copy of \bar{w}_2 in $\bar{w}_2w_2w_2\bar{w}_2$.



Then s_1 is both a prefix of w_2 and \bar{w}_2 , contradicting the fact that $\text{lcp}(w_2\bar{w}_2, \bar{w}_2w_2) = 0$. Note that the whole of $\bar{w}_2[3]$ might not be a part of the string (and that is why in the diagram it is depicted in gray), but then t is a non-trivial proper prefix of \bar{w}_2 and the the argument holds.

5. case when $\bar{v}_2v_2v_2\bar{v}_2$ does not at all overlap with $\bar{w}_2w_2w_2\bar{w}_2$.

That case is argued identically as for an inversion factor occurring to the left of L_2 .

The third step of the proof is to assume by contradiction that an inversion factor occurs to the left of L_2 which follows the same line of argumentation as the first step. The fourth and last step of the proof is to assume that an inversion factor occurs to the right of R_2 which follows the same line of argumentation as for the second step. \square

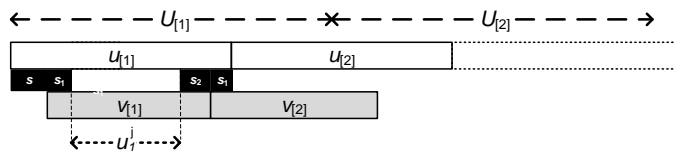
5.2 Proof of Lemma 17:

- (a) Case $\mathfrak{s}(v_{[1]}) < R_1$.

Without loss of generality we can assume that $\text{lcp}(u_1, \tilde{u}_1) = 0$ and hence $R_1 = N_1$. If it is not, instead of doing the argument with $u_1^{u(1)}u_2u_1^{(u(1)+u(2))}u_2u_1^{u(2)}$ we can do the argument with $s_2w_1^{u(1)}w_2w_1^{(u(1)+u(2))}w_2w_1^{(u(2)-1)}s_1$ where w_1 respective w_2 is a right cyclic shift of w_1 respective w_2 by $\text{lcp}(u_1, \tilde{u}_1)$ positions, $s_1s_2 = w_1$, and $|s_1| = \text{lcp}(u_1, \tilde{u}_1)$. Then $\text{lcp}(w_1, \tilde{w}_1) = 0$. The proof is carried out by a discussion of all possible cases of the ending point of $v_{[1]}$.

- (A) Case $\mathfrak{e}(v_{[1]}) \leq \mathfrak{e}(u_{[1]})$

Note that $\mathfrak{e}(v^2) > \mathfrak{e}(U_{[1]}) = \mathfrak{e}(u_1^{u(1)}u_2u_1^{u(2)})$, for otherwise there would be a farther copy of v^2 in $U_{[2]}$. By the inversion factor Lemma 13, $v_{[1]}$ does not contain the whole of any inversion factors. Thus, $v_{[2]}$ cannot contain either the whole of any inversion factors, and in particular cannot contain the inversion factor at N_1 . Therefore, $v_{[1]}$ must end in the suffix \bar{u}_2u_2 of $u_{[1]}$. Let s be the offset of $v_{[1]}$ in $u_{[1]}$ and let s_1 be the overlap between $u_{[1]}$ and $v_{[2]}$, i.e. $svs_1 = u = u_1^{u(1)}u_2$, see the diagram bellow for an illustration.



Then s_1 is both a prefix of v and a suffix of u . Since s_1 is the overlap of u_1 and v_1 , $|s_1| < |u_1|$ and s_1 is a suffix of $\bar{u}_2 u_2$. It follows that $v = t_1 u_1^i t_2$ for some suffix t_1 of u_1 , some prefix t_2 of u_1 , and some $i \geq 0$.

On the other hand, $U_{[1]} = u_1^{u(1)} u_2 u_1^{u(2)} = u u_1^{u(2)}$ is a non-trivial proper prefix of sv^2 , and so $svs_1 u_1^{u(2)}$ is a non-trivial proper prefix of sv^2 , implying that $s_1 u_1^{u(2)}$ is a non-trivial proper prefix of v and, therefore, $v = s_1 u_1^j s_2$ for some prefix s_2 of u_1 and some $j \geq 1$.

Thus, $v = t_1 u_1^i t_2 = s_1 u_1^j s_2$. Since t_1 is a suffix of u_1 and t_2 a prefix of u_1 , by Lemma 2, $t_1 = s_1$ and $t_2 = s_2$. Therefore, s_1 is a suffix of u_1 .

Since $s_2 s_1$ is a suffix of u , then $s_2 s_1 = u_1^i u_2$ for some $i \geq 0$. Since $|s_2| + |s_1| < 2|u_1|$, either $i = 0$ or $i = 1$, which proves that either $s_2 s_1 = u_1 u_2$ or $s_2 s_1 = u_2$.

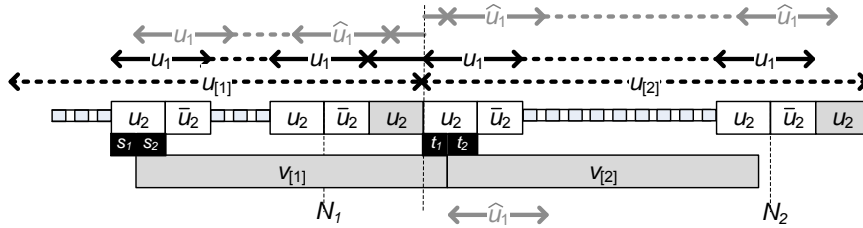
In the former case, $|v| = (j+1)|u_1| + |u_2|$ and so $v = \hat{u}_1^{(j+1)} \hat{u}_2$, while in the latter case $v = \hat{u}_1^j \hat{u}_2$, where in both cases \hat{u}_1 respective \hat{u}_2 is a left cyclic shift of u_1 respective u_2 by $|s_1|$ positions. The left cyclic shift is possible as s_1 is both a suffix of u_1 and a suffix of $\tilde{u}_1 = \bar{u}_2 u_2$. Therefore, $v = \hat{u}_1^j \hat{u}_2$ and $1 \leq j \leq u(1)$ and so when $j < u(1)$, case (a₁) holds true, and when $j = u(1)$, case (a₂) holds true.

(B) Case $e(u_{[1]}) < e(v_{[1]}) \leq e(u_{[1]} u_1)$

We discuss this case in four different configurations based on where $v_{[1]}$ starts and where it ends.

(1) A configuration when $v_{[1]}$ starts in a u_2 and ends in the first u_2 of $u_{[2]}$.

Let s_1 be the offset of $v_{[1]}$ in the u_2 it starts in, let s_2 be the overlap of $v_{[1]}$ and the u_2 it starts in, let t_1 be the overlap of $v_{[1]}$ with the u_2 it ends in, and let t_2 be the overlap of $v_{[2]}$ with the u_2 where $v_{[1]}$ ends. Let $\hat{u}_1 = s_2 \bar{u}_2 s_1$; as a conjugate of u_1 , it is primitive.

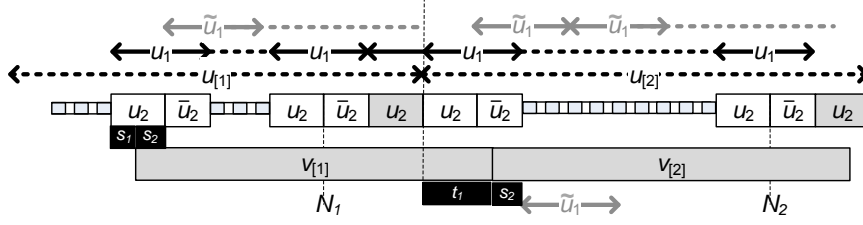


By Lemma 2, $t_1 = s_1$ and $t_2 = s_2$, and so $s_1 v_{[2]}$ is a non-trivial proper prefix of $u_1^{(u(1)+u(2))} u_2$. It follows that the suffix $u_2 s_1$ of v must align with $u_1 u_2 = (u_2 \bar{u}_2) u_2$ of $u_1^{(u(1)+u(2))} u_2$, and so s_1 is prefix of $u_2 \bar{u}_2$. Thus, s_1 is a prefix of both, $u_2 \bar{u}_2$ and $\bar{u}_2 u_2$. Therefore, $|s_1| \leq lcp(u_1, \tilde{u}_1) = 0$, and so s_1 is empty. It follows that $v = u_1^j u_2$ for $1 \leq j \leq u(1)$ and so either (a₁) or (a₂) holds true.

Note that Lemma 2 applies even if $u(1) = 1$, since then $v_{[1]}$ must start in the very first u_2 of $u_{[1]}$.

(2) A configuration when $v_{[1]}$ starts in a u_2 and ends in the first \bar{u}_2 of $u_{[2]}$.

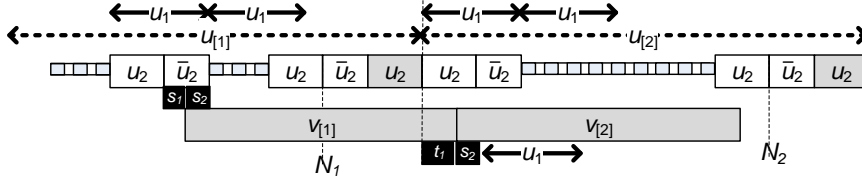
Let s_1 and s_2 be as in the previous case (B)(1). Let t_1 be the overlap of $v_{[1]}$ and $u_{[2]}$.



The factor $u_1^{(u(1)+u(2))}u_2$ has $u_2\tilde{u}_1\tilde{u}_1$ as a prefix as $u(1)+u(2) \geq 2$. The factor v has $s_2\tilde{u}_1$ as a prefix. Thus $u_1^{(u(1)+u(2))}u_2$ has also $t_1s_2\tilde{u}_1$ as a prefix. Since $|t_1s_2| < |u_2|+|u_1|$, this contradicts Lemma 2, as \tilde{u}_1 is primitive being a conjugate of u_1 . Such a configuration is not possible.

- (3) A configuration when $v_{[1]}$ starts in a \bar{u}_2 and ends in the first u_2 of $u_{[2]}$.

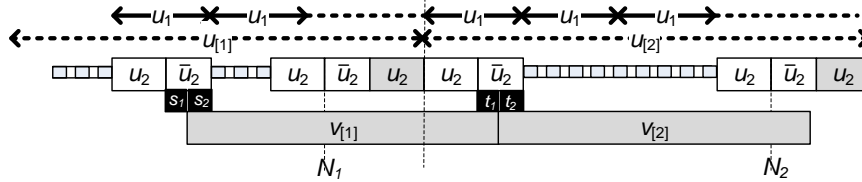
Let s_1 be the offset of $v_{[1]}$ in \bar{u}_2 it starts in, let s_2 be the overlap of $v_{[1]}$ and the \bar{u}_2 it starts in. Let t_1 be the overlap of $v_{[1]}$ with $u_{[2]}$



The factor v has s_2u_1 as a prefix, and so $u_1^{(u(1)+u(2))}$ has as a prefix u_1u_1 and $t_1s_2u_1$. Since $|t_1s_2| < |u_1|$, this contradicts Lemma 2. Such a configuration is not possible.

- (4) A configuration when $v_{[1]}$ starts in a \bar{u}_2 and ends in the first \bar{u}_2 of $u_{[2]}$.

Let s_1 and s_2 be as in (B)(3). Let t_1 be the overlap of $v_{[1]}$ and the \bar{u}_2 it ends in, and let t_2 be the overlap of $v_{[2]}$ with the \bar{u}_2 in which $v_{[1]}$ ends.



By Lemma 2, $t_1 = s_1$ and $t_2 = s_2$. Since $u_2s_1v_{[2]}$ is a prefix of $u_1^{(u(1)+u(2))}u_2$, it follows that the suffix $u_2u_2s_1$ of $v_{[2]}$ must align with u_1u_2 in $u_1^{(u(1)+u(2))}u_2$, and thus $u_2u_2s_1$ is a prefix of $u_2\bar{u}_2u_2$, hence u_2s_1 is a prefix of \bar{u}_2u_2 . Thus, $u_2\bar{u}_2 = u_2s_1s_2$ is a prefix of $\bar{u}_2u_2s_2$, giving $u_2\bar{u}_2 = \bar{u}_2u_2$, which is a contradiction as $u_2\bar{u}_2$ is primitive. Such a configuration is not possible.

- (C) Case $\epsilon(u_{[1]}u_1) < \epsilon(v_{[1]}) < R_2$.

Then $v_{[1]}$ contains the inversion factor at R_1 . Thus, $v_{[2]}$ must contain the inversion factor at R_2 and it must be placed in $v_{[2]}$ in the same position mate to the beginning of $v_{[2]}$ as in $v_{[1]}$, and therefore $|v| = R_2 - R_1 = |U|$. Thus, case (a₄) holds true.

- (D) Case $R_2 \leq \epsilon(v_{[1]})$.

Since $\epsilon(v_{[1]}) \geq R_2 \geq N_2 = u_1^{u(1)}u_2u_1^{(u(1)+u(2)-1)}u_2$, either $s_1\bar{u}_2u_2u_1^{(u(1)+u(2)-1)}u_2$ for some suffix s_1 of u_2 is a prefix of v , or

$s_1 u_1^i u_2 u_1^{(u(1)+u(2)-1)} u_2$ for some suffix s_1 of u_1 and some $i \geq 1$ is a prefix of v , and so case (a₅) holds true.

Case (a₃) is not possible as it never materialized during the discussion of the cases (A) - (D) that cover exhaustively all possible endings of $v_{[1]}$.

(b) Case $\epsilon(v_{[1]}) \leq \epsilon(u_{[1]})$.

If $\mathfrak{s}(v_{[1]}) \geq R_1$, then $|v| < |u_1|$ and so v^2 is a factor of $u_1 u_2 u_1$ and hence of $U_{[1]}$, and thus there is a farther copy of v^2 in $U_{[2]}$, a contradiction. Therefore $\mathfrak{s}(v_{[1]}) < R_1$ and this is the case (A) above, and thus the either the case (a₁) or case (a₂) holds. \square

5.3 Proof of Lemma 19:

Case (a): since $\mathfrak{s}(v^2) = \mathfrak{s}(V^2) = \mathfrak{s}(V) \leq R_1(\mathcal{U})$, applying Lemma 17 to v^2 and V^2 gives the following possibilities:

- (i) $v = \widehat{u}_1^i \widehat{u}_2$ for $1 \leq i < u(1)$ where \widehat{u}_2 is a non-trivial proper prefix of \widehat{u}_1 and where \widehat{u}_1 respective \widehat{u}_2 is a cyclic shift of u_1 respective u_2 in the same direction by the same number of positions (by item (a₁) of Lemma 17 applied to v^2),
- (ii) $v = \widehat{u}_1^{u(1)} \widehat{u}_2$ where \widehat{u}_2 is a non-trivial proper prefix of \widehat{u}_1 and where \widehat{u}_1 respective \widehat{u}_2 is a cyclic shift of u_1 respective u_2 in the same direction by the same number of positions (by item (a₂) of Lemma 17 applied to v^2),
- (iii) $|v| = |U|$ (by item (a₄) of Lemma 17 applied to v^2),
- (iv) $\epsilon(v_{[1]}) - \epsilon(u_{[1]}) \geq (u(1)+u(2)-1)|u_1| + |u_2|$ (by item (a₅) of Lemma 17 applied to v^2),
- (I) $V = \widehat{u}_1^j \widehat{u}_2$ for $1 \leq j < u(1)$ where \widehat{u}_2 is a non-trivial proper prefix of \widehat{u}_1 and where \widehat{u}_1 respective \widehat{u}_2 is a cyclic shift of u_1 respective u_2 in the same direction by the same number of positions (either by item (a₁) or (a₂) of Lemma 17 applied to V^2),
- (II) $|V| = |U|$ (by item (a₄) of Lemma 17 applied to V^2),
- (III) Either $s_1 \overline{u}_2 u_2 u_1^{(u(1)+u(2)-1)} u_2$ for some suffix s_1 of u_2 is a prefix of V , or $s_1 u_1^i u_2 u_1^{(u(1)+u(2)-1)} u_2$ for some suffix s_1 of u_1 and some $j \geq 1$ is a prefix of V (by item (a₅) of Lemma 17 applied to V^2).

We inspect all possible combinations:

- Combining (i) and (I) is impossible: since v is a prefix of V , $\widehat{u}_1 = \widehat{u}_1$ and $\widehat{u}_2 = \widehat{u}_2$. Since $j > i$ as $|V| > |v|$, we can apply Lemma 8 deriving a contradiction.
- Combining (i) and (II) is possible and yields case (a₂): since v is a prefix of V , $\widehat{u}_1 = \widehat{u}_1$ and $\widehat{u}_2 = \widehat{u}_2$ and so \mathcal{V} must be a β -mate of \mathcal{U} . Since $|V| = |U| = (u(1)+u(2))|u_1| + |u_2|$, $V = \widehat{u}_1^i \widehat{u}_2 \widehat{u}_1^{(u(2)+u(1)-i)}$. Since $i \geq u(2)+u(1)-i$ as otherwise there would be a farther copy of v^2 , $2i \geq u(1)+u(2)$. Since $1 \leq i < u(1)$, $i = u(1)-k$ for some $1 \leq k < u(1)$. It follows that $2(u(1)-k) \geq u(1)+u(2)$, so $2u(1)-2k \geq u(1)+u(2)$, and thus $u(1) \geq u(2)+2$.
- Combining (i) and (III) is impossible: since v^2 is a prefix of V^2 , $\widehat{u}_1^i \widehat{u}_2 \widehat{u}_1^i \widehat{u}_2$ is a prefix of V^2 . At the same time either $s_1 u_1^j u_2 u_1^{(u(1)+u(2)-1)} u_2$ is a prefix of V or $s_1 \overline{u}_2 u_2 u_1^{(u(1)+u(2)-1)} u_2$ is a prefix of V . Due to Lemma 2, in both cases, $\widehat{u}_1^i \widehat{u}_2 \widehat{u}_1^{(u(1)+u(2)-1)} \widehat{u}_2$ is a prefix of V

and so $\widehat{u}_1^i \widehat{u}_2^i \widehat{u}_1^i \widehat{u}_2^i$ is a prefix of V . It follows that v^2 is a factor in $V_{[1]}$ and, consequently, it has a farther copy in $V_{[2]}$, a contradiction.

- Combining (ii) and (I) is impossible: as $j \leq u_{(1)}$ implies that $|V| \leq |v|$, hence a contradiction.
- Combining (ii) and (II) is possible and yields that \mathcal{V} is an α -mate of \mathcal{U} , hence case (a₁).
- Combining (ii) and (III) is impossible for the same reasons as for the combination (i) and (III).
- Combining (iii) and (I) or (II) is impossible due to the size of v being bigger than the size of V .
- Combining (iii) and (III) is possible and yields case (a₃) and so \mathcal{V} is a γ -mate of \mathcal{U} .
- Combining (iv) and (I) or (II) is impossible due to the size of v being bigger than the size of V .
- Combining (iv) and (III) yields case (a₄).

Case (b): The FS-double square \mathcal{V} is an ε -mate of \mathcal{U} by definition as $R_1 \leq s(\mathcal{V})$. If $e(v_{[1]}) \leq e(u_{[1]})$, then by Lemma 17, $s(\mathcal{V}) < R_1$, a contradiction. So $e(u_{[1]}) < e(v_{[1]})$. \square

Acknowledgments. The authors would like to thank Nguyen Huong Lam for valuable comments and suggestions which improved the quality of the paper. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada, MITACS, and by the Canada Research Chairs program.

References

- [1] M. Crochemore and W. Rytter. Squares, cubes, and time-space efficient string searching. *Algorithmica*, 13:405–425, 1995.
- [2] A. Deza and F. Franek. A d -step approach to the maximum number of distinct squares and runs in strings. *Discrete Applied Mathematics*, 163:268–274, 2014.
- [3] A. Deza, F. Franek, and M. Jiang. A computational framework for determining square-maximal strings. In J. Holub and J. Žďárek, editors, *Proceedings of the Prague Stringology Conference 2012*, pages 111–119, Czech Technical University in Prague, Czech Republic, 2012.
- [4] A.S. Fraenkel and J. Simpson. How many squares can a string contain? *Journal of Combinatorial Theory, Series A*, 82(1):112–120, 1998.
- [5] F. Franek, R.C.G. Fuller, J. Simpson, and W.F. Smyth. More results on overlapping squares. *Journal of Discrete Algorithms*, 17:2–8, 2012.
- [6] L. Ilie. A simple proof that a word of length n has at most $2n$ distinct squares. *Journal of Combinatorial Theory, Series A*, 112(1):163–163, 2005.

- [7] L. Ilie. A note on the number of squares in a word. *Theoretical Computer Science*, 380(3):373–376, 2007.
- [8] E. Kopylova and W.F. Smyth. The three squares lemma revisited. *Journal of Discrete Algorithms*, 11:3–14, 2012.
- [9] M. Kubica, J. Radoszewski, W. Rytter, and T. Walęń. On the maximum number of cubic subwords in a word. *European Journal of Combinatorics*, 34:27–37, 2013.
- [10] N. H. Lam. On the number of squares in a string. AdvOL-Report 2013/2, McMaster University, 2013.
- [11] M. J. Liu. Combinatorial optimization approaches to discrete problems. Ph.D. thesis, Department of Computing and Software, McMaster University, 2013.