McMaster University

Advanced Optimization Laboratory

Title:

On the number of squares in a string

Authors:

Nguyen Huong Lam

AdvOL-Report No. 2013/2

August 2013, Hamilton, Ontario, Canada

On the number of squares in a string

Nguyen Huong Lam

August 2013

(written August 2009, revised April 2010)

Abstract

Fraenkel and Simpson (1995) proved that the string of length n contains at most $2n$ distinct squares. Ilie (2007) improved this bound to $2n - \Theta(\log n)$. Here, it is proved that the number of distinct squares is less than $\frac{95}{48}n$. The proof is based on the notion of type, period and residue of squares. The conjectural value is n

1 Introduction

String, or word, is a sequence of symbols, ("letters"), from a finite set, called alphabet; the number of symboled encountered in the string is called the length of it. The set of strings on an alphabet is equipped with the operation of catenation: if u and v are two strings, the catenation uv of them is simply the juxtaposition of v next to v; the string of the form uu is called a square. How many squares can contain a given string as a segment? This kind of questions is typical in the topics of counting repetitions in a string with long history and ample variations. In 1995, A. Fraenkel and J. Simpson [3] proved that a string of length n contains at most $2n$ *distinct* squares; the authors counted the squares by their latest occurrence in the string, based on a subtle observation that there exist at most two squares whose latest occurrences start at the same position of the string, so the bound above readily follows. Fraenkel and Simpson make use of an intriguing result, called Lemma of Three Squares, to prove their observation, although nowadays it can be derived directly in just a few lines without recourse to the lemma of three squares. This result has inspired considerable interests and motivated further research around the problem. In an attempt to improve the result, in 2007, Ilie proved that the number of distinct squares is bounded by $2n-\Theta(\log n)$ by showing that the number of successive positions, at which latest occurrences of *two* squares start, are half the length the string. This result is modest but the approach has been bearing some fruits. The numerical data suggest that the bound may be n; see [4] for the state of the art and the development of the subject. In this paper I try to give an estimate on the number of positions at which two squares begin from which the bound announced in the abstract follows immediately. This bound above is still weak; maybe, I am missing a simple unifying and encompassing idea, but the approach has also some benefits, it shed some light on their structure and how the squares follow each other and how weird the squares of small type are. The crux of the proof is that we make a classification of the squares starting inside a given square according to the position at which they end then try to estimate the number of the squares of each kind, and subsequently use the induction. To do this, in the sections 3 through 6 we develop a machinery to support the counting argument and in the section 7, we process the proof itself. Now we fix a minimum amount of notation, preliminary results and adopt the key notions for the proof, in the next section.

2 Background

Let A be a finite alphabet and A^* be the set of words on A — we use *words* instead of strings. The product of two words u and v, denoted by uv, is the catenation of u and v; if $u = v$ then uu is called square and we sometimes write u^2 for uu and similarly, u^k instead of $uu \cdots u$ (k times), $k = 1, 2, 3, \ldots$. The words of the alphabet are called letters and every word is written in the form $u = a_1 a_1 \cdots a_n$, where a_1, a_2, ..., a_n are letters; the number n is the length of u, denoted by $|u|$. By convention, when $n = 0$, u is the empty word, denoted by 1 and the length $|1| = 0$ and $u^0 = 1$; also $1u = u1 = u$ for every u. The word 1 is thus a unit for the catenation and every word different from 1 is called nonempty. By definition u is nonempty if and only if $|u| > 0$.

A nonempty word is said to be primitive if it is not a power of another word, that is $u = v^k$ implies $k = 1$ ($u = v$). It is easy to see that every word is a power of a primitive word; more than that, it is a power of unique primitive word, which is called the primitive root of it: this claim assumes more labor to verify, we refer to [5] or [6] for a proof.

Two word u and v are said to be conjugate if $u = rs$ and $v = sr$ for some words s, r. The conjugate words are clearly of the same length and they are primitive or not at the same time which is directly verified by definition.

Let now w be an arbitrary word. A factor of w is a word f satisfying $w = ufv$ for some words u, v. Intuitively, factor is a word that occurs within w, and it does so at different positions; for a nonempty factor f, by this relation we say that the factor f occurs at or has an occurrence at the position $|u|$. More specifically, we say that f has an occurrence *starting* at the position $|u|$ and *ending* at the position $|uf| - 1 = |u| + |f| - 1$ (of w). Among the factors, there is a special one called prefix: the factor f for which $w = fv$, or the factor that has the occurrence at the position 1. Symmetrically, we say of suffix, as the factor f for which $w = uf$ or which has an occurrence starting at the position $|u| = |w| - |f|$, or ending at the position $|w|$. A factor f of w is said to be proper if it is not w, or equivalently $|f| < |w|$. By definition 1 is factor of every word.

Let now W be a word whose distinct squares we manage to count and f be factor of W. The occurrence at position j is said to be later than the occurrence at position i if $j > i$; an occurrence is the latest if it is not later than any other one.

The following result of Fraenkel and Simpson [3] is critical to the subject.

PROPOSITION 2.1. *For an arbitrary word, the number of distinct squares of it that have the latest occurrence at the same position does not exceed 2.*

This result motivates the following definition, convenient for our exposition.

DEFINITION 2.2. Let W be an arbitrary word. A *double square* of W is a pair of squares uu and UU with $|u| < |U|$, abbreviated (u, U), of W such that u and U have the latest occurrence at the same position, say, i. uu is called the shorter and UU the longer square, and $|u|$ is called the smaller and $|U|$ the larger length of the double square (u, U).

We then say that that the double square (u, U) *occurs* at the (unique) position i. Thus, it has the usual meaning to say that one double square occurs later than another one, and it is the first, the second, etc. of W according to the position of its occurrence

EXAMLE 2.3. Let $W = abaababaabaababaa$ which has the following ten distinct squares aa, $abab$, $baba$, $baabaa$, $aabaab$, $abaaba$, $abaababaab$, $baababaaba$, $abaababaabaababa$ and $baababaabaababaa$ having their latest occurrence at the positions 16,12, 13, 7,8,9, 1,2,1 and 2, respectively. Thus, W has two double squares $(abaab, abaababa)$ and $(baaba, baababaa)$, occuring at the positions 1 and 2 respectively and $(abaab, abaababa)$ is the first double square of W.

The double square has a special structure as shown in the following assertion.

PROPOSITION 2.4. *Let (u, U) be a double square. Then $u = x^p x_1, U = x^p x_1 x^q$ for some primitive word x, nonempty proper prefix x_1 of x and positive integers p, q with $p \geq q$.*

Proof. Since uu and UU have their latest occurrence at the same position i, u is a proper prefix of U, so let $U = uz$ for a nonempty word z. We have also $|uu| > |U|$, otherwise uu is a prefix of U, hence uu has another occurrence at the position $i + |u|$ which is later than i. Therefore $|z| < |u|$. As uu is a prefix of UU, hence of $Uu = uzu$, we get that u is a prefix of zu. This means $u = z^k z_1$ with $k > 0$ and z_1 is a proper prefix of z. Now let x be the primitive root of z, $z = x^m$, and $z_1 = x^l x_1$, $l \geq 0$ with x_1 a proper prefix of x. So

$$u = x^{km} x_1, U = uz = x^{km} x_1 x^m.$$

Put $p = km, q = m$, indeed, $p \geq q$. The prefix x_1 must be nonempty, otherwise $UU = x^{p+q} x^{p+q} = uux^{2q} = x^{2q} uu$: uu has a later occurrence despite the assumption. This completes the proof.

Let $x = x_1 x_2$ and we denote $\bar{x} = x_2 x_1$, then

$$UU = ux^q ux^q = x^p x_1 x^{q+p} x_1 x^q = uu\bar{x}^q x^q.$$

We shall see that $x \neq \bar{x}$ by Remark 2.6 (ii) below.

The proposition above gives way to the following definition.

DEFINITION 2.5. Let (u, U) be a double square, $u = x^p x_1$, $U = x^p x_1 x^q$. The pair of integers (p, q) is the *type*, the word x is the *period* and the nonempty proper prefix x_1 of x is the *residue* of (u, U).

Abusing languages, now and then we also call the length $|x|$ and $|x_1|$ period and residue and p the type of (u, U) if this causes no confusion. We shall, in this work, exploit the primitivity of the period extensively with the help of the two following special remarks.

REMARK 2.6. The following facts are simple instances or easy consequences of the so-called Periodicity Lemma, of Fine and Wilf [2]. Let two arbitrary words x^m and y^n have a common factor of length $|x| + |y|$ then $x = \mu^k, y = \lambda^l$ for some nonnegative exponents k and l of some conjugate primitive words μ and λ.

(i) If x and y both are primitive then $k = l = 1$ and x and y are conjugates. In particular, $|x| = |y|$.

If x and y are prefixes of each other or suffixes of each other then x and y are copower (that is the power of the same word or, equivalently, have the same primitive root) of μ. In particular, we two following points:

(ii) If $xy = yx$ then x and y are copower;

(iii) If $xx = uxv$, or equivalently $x = uv = vu$, for some nonempty words u, v then u, v and x are copower. Besides, x is not primitive.

See [5] and [6] for a proof of the Periodicity Lemma.

Remark 2.7. This is the so-called synchronizing property of the primitive word. Let p be a primitive word.

(i) If $up^m v = p^n$ for some words u, v and integers $m, n > 0$ then $u = p^k, v = p^l$, $k, l \geq 0$. Consequently, if $w = up = pv$ and $xwy = p^k$ for some words u, v, x, y and an integer $k \geq 0$ then $w = p^l$ for an integer $l \geq 0$.

(ii) If x and y are conjugates of p and $pp = exf = gyh$ for some word e, f, g, h then $e = g, f = h$ and $x = y$.

The present remark can be verified directly by the previous one. See [1] for the general framework. Let (u, U) and (v, V) be two double squares occurring at the positions i and j, respectively. We say that (v, V) is a *u-close* if and only if $i < j < i + |u|$. The following is one of the premier results that has been proved in [4]; we recast it in our way, but first a definition.

Definition 2.8. Let (u, U) be a double square of type (p, q), period x and residue x_1. A *replica* of (u, U) is a u-close double square having the same longer and larger lengths as those of (u, U), respectively. The (replicating) *capacity* of (u, U) is the quantity $\eta(u, U) = |\mathrm{lcp}\,(x, \bar{x})| + 1$ if $p > 1$ and $\eta(u, U) = \min\{|\mathrm{gcp}\,(x, \bar{x})| + 1, |x_1|\}$ if $p = 1$.

By Remark 2.6 (iii), $\eta(u, U) \leq |x_1|$ if $p = 1$ and $\eta(u, U) \leq |x|$ (even more, $\eta(u, U) \leq |x| - 1$, if the full Periodicity Lemma is in effect, but we do not need that much) if $p > 1$.

We will see in the next theorem that every replica of a double square has the structure similar to that of the "original".

Proposition 2.9. *Suppose that (u, U) is a double square of type (p, q), period x, residue x_1, occurring at the position i and (v, V) is a replica of it, occurring at the position $i + |a|$ for a nonempty proper prefix a of u. Then $|a| < \eta(u, U)$. Moreover, every replica of (u, U) has the same period length and the residue length, hence, the same type as (u, U). Consequently, the number of replicas of (u, U) does not exceed $\eta(u, U)$.*

Proof. By assumption $|u| = |v|$ we have $u = ab, v = ba$ for some nonempty word b, therefore

$$avv = uua.$$

This shows first that x and a are prefixes of each other, and next, as

$$UU = uu\bar{x}^q x^q$$

and avv and UU are prefixes of each other, we see that a and $\bar{x}$ are prefixes of each other. For x is primitive, $|\mathrm{gcp}\,(x,\bar{x})| < |x|$ by Periodicity Lemma, therefore in case $|a| > |\mathrm{gcp}\,(x,\bar{x})|$, the two words x and $\bar{x}$ have a common prefix of length $|\mathrm{gcp}\,(x,\bar{x})| + 1$, which is a nonsense. Thus $|a| \leq |\mathrm{gcp}\,(x,\bar{x})|$.

As for the case $p = 1 = q$, we must have $|a| < |x_1|$, otherwise, first, x_1 is a prefix of a and second, $U = ac$, $V = ca$ for some word c. Expanding

$$aVV = acaca = UUa = xx_1xxx_1xa$$

we see then that the square xx_1xx_1 occurs at the position $i + |xx_1x|$ later than i, as a has the prefix x_1: a contradiction.

For the last claim, let (v, V) be a replica of type (p', q'), period y and residue y_1 of (u, U) such that $u = ab, U = ac$ and $v = ba, V = ca$. We see that

$$aVV = avy^{q'}vy^{q'} = abay^{q'}vy^{q'} = uay^{q'}vy^{q'}.$$

On the other hand

$$aVV = acaca = UUa = ux^{q+p}x_1x^qa$$

and $|y^{q'}| = |V| - |v| = |U| - |u| = |x^q|$. Since $|a| < \eta(u, U) \leq |x|$ we have $|ay^{q'}| \leq |x^{q+p}x_1|$, therefore $y^{q'}$ is a factor of $x^{q+p}x_1$ and $y^{q'} = z^q$ for some conjugate z of x. That both x and y are primitive gives us $y = z$ and $q' = q$, hence $|x| = |y|$ and $p' = p$ and $q = q'$ that concludes the proof.

The aim of this paper is to prove the following bound from above.

Theorem 2.10. *Let W be an arbitrary word and (u, U) be its first double square. Then for the number of distinct double squares $\gamma(W)$ of W,*

$$2\gamma(W) < -|u| + C|W|,$$

where $C = \frac{47}{24}$.

Corollary 2.11. *The number of distinct squares of W is less than $C_0|W|$, where $C_0 = \frac{95}{48}$.*

Proof. Let $s(W)$ be the number of "single" (that is, those which are not double) squares of W. Clearly, the number of distinct squares of $|W|$ is $2\gamma(W) + s(W)$ for which

$$2\gamma(W) + s(W) = \gamma(W) + s(W) + \gamma(W) \leq |W| + \frac{C}{2}|W| = \frac{C+2}{2}|W| = C_0|W|$$

which is what we claimed.

Example 2.12. Since the period is a primitive word, the residue is a nonempty proper prefix of the period and the type is a pair of positive integers, each double square has the longer length at least 5, hence W has a double square only if it is at least of length 10. Thus $\gamma(W) = 0$ if $|W| \leq 9$ and $\gamma(W) \leq 1$ if $|W| = 10$. So for W of length not exceeding 10, we have $\gamma(W) \leq 1$ and $2\gamma(W) \leq 2 < -5 + \frac{47}{24}|W|$.

Now consider the word $W = abaababaabaababaa$. We see that $|W| = 17$, $\gamma(W) = 2$, the shorter length of the first double square is 5. The bound given by Theorem 2.10 is

$-5 + \frac{47}{24} \cdot 17 < 29$, the true value is $\gamma(W) = 4$. The total number of distinct squares is 10, while the bound given by Corollary 2.11 is $\frac{95}{48} \cdot 17 \leq 33$!

We stop here to bid adieu to the preliminary part by saying some caveats. The rest of the paper is devoted to the proof of Theorem 2.10. Unless the reader is endowed with a mental agility of manipulation on words (the author is not), we urge him or her to accompany proofs with suitable drawings describing the situation (as the author did) then the proofs will look intuitive and transparent despite its wordy and lengthy formalism.

3 Coperiod Squares

Let (u, U) and (v, V) be double squares. As the term suggests, we begin with the following.

Definition 3.1. The double square (v, V) is said to be *coperiod* with (u, U) if (v, V) starts inside u and they have the same period length. Moreover, two coperiod double squares are *coresidue* , if in addition, they have the same residue length.

When a double that follows (u, U) is coperiod with it? In this section we answer this question in some detail. First, for those double squares that end inside u inside u.

Proposition 3.2. *Let (u, U) and (v, V) be two distinct double squares. If (v, V) starts and ends inside u then (u, U) has the type large than 2 and (v, V) is coperiod, and more than that, coresidue with (u, U).*

Proof. Let (u, U) has type (p, q), period x and residue x_1. By assumption, we have $u = avt$ with $|a| > 0$, $0 \leq |t|$. Since the square vv starts at the position $|a| + 1$ inside u, we have for the second occurrence of v:

$$v = tx^j s'$$

with s' prefix of x and t suffix of $u = x^p x_1$. Here, we must have $j > 0$ for vv to avoid occurring later, as $UU = ux^{q+p}x^q$, $q > 0$. As for the first occurrence of v, we have

$$v = t'x^i s$$

with t' suffix of x and s prefix of x and $i \geq 0$.

Since $j > 0$, x primitive, $v = tx^j s'$ is a factor of $x^p x_1$, by Synchronization Remark we get $s = s'$ and $t = t'x^n$ for some $n \geq 0$. But we should have $|t| < |x|$ otherwise t has the suffix x that means $x = \bar{x}$ contradicting primitivity of x. So $n = 0$ and $t = t'$ either which gives $i = j$.

We distinguish between two cases.

Case 1. $0 \geq |t| \geq |x_1|$.

Then $st = x_1$ and $|v| = |t'x^i s| = |x^i x_1|$ with $i < p$ as $|v| < |u|$. Here we get $p > 1$.

Let d be the word of length $|x|$ such that $avvd$ is a prefix of UU. Comparing lengths (notice that $|av| < |u|$)

$$|ux^{q+p}x_1| - |avv| \geq |x^{q+p}x_1| - |v| = |x^{q+p}x_1| - |x^i x_1| = |x^{q+p-i}| \geq |x^{q+1}| \geq |xx|$$

lets us see that d is a factor of xx. Now that $v = tx^j s'$, we have $d = t''s'$ and vd has $v = t'x^j s$ as a suffix, because $s' = s$ is a suffix of x.

Now suppose that (v, V) has type (p', q'), period y and residue y_1. Also, $v\bar{y}^{q'}$ has the suffix v. We see that $VV = vv\bar{y}^{q'}y^q$ occurs at the position $|a| + 1$ inside u and that v is a suffix of $v\bar{y}^{q'}$ hence $\bar{y}^{q'}$ and d are prefixes of each other.

If $|\bar{y}^{q'}| < |x| = |d|$, then $\bar{y}^{q'}$ is a prefix of d we set $d = \bar{y}^{q'}e$ with $|e| > 0$. So v is a suffix of ve and of $v\bar{y}^{q'}$ at the same time. Thus v is a suffix of both $\bar{y}^*$ and e^* with $|d| = |x| = |\bar{y}| + |e|$. As $|v| > |x|$, we get that $\bar{y}$ and e are copower which is impossible because x is primitive.

If $|\bar{y}^{q'}| \geq |x| = |d|$ then we get that d is a prefix of $\bar{y}^{q'}$. We set $\bar{y}^{q'} = de$, where $|e| > 0$. Analogously, this also leads to the conclusion that d and e are copower of $\bar{y}$ since $\bar{y}$ is primitive. As d is also primitive (as a conjugate of x), $d = \bar{y}$. Altogether, this means $|x| = |y|$ and, consequently $|y_1| = |x_1|$ that is, v is coperiod and more than that, coresidue with u.

We show further that $q' > 1$. Suppose on the contrary $q' = 1$. Then $|\bar{y}y| = |xx|$. Now that $aVV = avv\bar{y}y$, in view of the comparing length inequality above we get $\bar{y}y$ is a factor of xx implying that $\bar{y} = y$: a contradiction.

Now $p > i = p' \geq q' > 1$ shows that $p > 2$ that completes the Case 1.

Case 2. $|x_1| \leq |t|$.

This time $st = xx_1$ and $|v| = |t'x^i s| = |x^{i+1}x_1|$ implying that $p > i + 1$. We get the same inequality comparing lengths and we proceeds just as in Case 1, without anomaly. The proof is complete.

Squares considered in the proposition above play a part in our enumeration later, so we prescribe them a name.

Definition 3.3. Let (u, U) be a double square. A double square that both starts and ends inside u is called *inner* (coperiod) double square of (u, U).

By Proposition 3.2 a double square has an inner double square only if it has type at least 3. The inner double square has a special feature as follows.

Proposition 3.4. *A double square and every of its inner double squares have the same larger length.*

Proof. Let (v, V), of type (p', q'), period y and residue y_1, be an inner double square of (u, U). As before, we retain the notation of Proposition 3.2, (u, U) has type (p, q), period x and residue x_1. We have to prove that

$$|x^{q+p}x_1| = |y^{p'+q'}y_1|.$$

We have $u = avt$ and $v = t'x^i s$, $v = tx^j s'$, where t' is a proper suffix of x and s, s' are proper prefixes of x. We have seen in the proof of Proposition 3.2 that $t' = t$, $s' = s$ and $|p'| < |p|$, as $|a| > 0$. From the equalities

$$aVV = avy^{q'}vy^{q'} = avy^{q'+p'}y_1y^{q'}$$

and

$$aUU = ux^{q+p}x_1x^q = avtx^{q+p}x_1x^q$$

we see that $y^{q'+p'}y_1$ and $tx^{q+p}x_1$ are prefixes of each other.

In case $|y^{q'+p'}y_1| \leq |tx^{q+p}x_1|$ then $tx^{q+p}x_1 = y^{q'+p'}y_1d$ for some word d. Since $|x| = |y|, |x_1| = |y_1|$ and t is a prefix of x^px_1, we get $d = z^lt$ for some word z such that $|z| = |x|$ and some integer $l \geq 0$.

If $l > 0$ then $z = y_2y_1 = \bar{y}$ and

$$tx^{q+p}x_1 = y^{q'+p'}y_1\bar{y}^lt$$

or

$$avtx^{q+p}x_1 = avy^{q'}y^{p'}y_1\bar{y}^lt = aVv\bar{y}^lt.$$

However

$$avtx^{q+p}x_1 = ux^{q+p}x_1 = Uu$$

is a prefix of UU, so, of $aVv\bar{y}^l$, and indeed, of $aVv\bar{y}$, and thus this word is a prefix of UU. But aVV and UU are prefixes of each other, which implies that $aVv\bar{y}$ is a prefix of $aVV = aVvy^{q'}$, which in turn implies $\bar{y} = y$, as $q' > 0$: a contradiction. Thus, we must have $l = 0$ and the equality $|U| = |x^{q+p}x_1| = |y^{q'+p'}y_1| = |V|$ follows.

For the remaining case $|y^{q'+p'}y_1| > |tx^{q+p}x_1|$, set $y^{q'+p'}y_1 = tx^{q+p}x_1e$ for some nonempty word e. We have $|t| + |e| = l|x|$ for some integer $l > 0$. Since t is a suffix x^px_1 and te is a suffix of $y^{q'+p'}y_1$ we get $te = (y_2y_1)^l$. But e and x^q (note that $UU = ux^{q+p}x_1x^q$) are prefixes of each other that means e and x has a common prefix of length at least $|x| - |t|$, so tx and te has a common prefix of length $|x|$. Consequently, te and v has a common prefix of length $|x| = |y|$ that is, y, a contradiction with $te = (\bar{y}^l)$. This concludes the proof.

We have seen in the proof of Proposition 3.2 that for an inner double square (v, V) of (u, U) then $u = avt$, $v = t'x^is$, where t' is a proper suffix of x and $t = t'$ which shows that t is a common suffix of x and $\bar{x}$. The structure of an inner square is thus described in full. What about the coperiod square that starts inside u but ends outside it? They happen in which cases? Note that if (v, V) starts inside u and ends outside u then we can write $u = ab$, $v = bs$ for some words a, b, s such that $|a| > 0, |s| > 0$. The following simple assertion give an answer.

Proposition 3.5. *Let (u, U) be a double square of type (p, q), period x, residue x_1. Suppose that (v, V) is a double square, coperiod with (u, U), of period y and starting inside u and ends outside u with $u = ab$, $v = bs$ then either $|s| < |x|$ or $|s| > |x^{q+p-1}x_1|$.*

Proof. Assumes on the contrary that $|x| \leq |s| \leq |x^{q+p-1}x_1|$. Set $x^{q+p}x_1 = sr$ for some word r, for which $|r| \geq |x|$ by assumption. Note also that r and v are prefixes of each other. Let d be the suffix of s of length $|x|$ and e be the prefix of r of length $|x|$. Because both d and e are factors of length $|x|$ of $x^{q+p}x_1$, therefore, of x^*, they are equal: $d = e$. But d is a suffix of $v = bs$, hence $d = \bar{y}$ and e is a prefix of v, hence $e = y$ that leads to $y = \bar{y}$, a contradiction which completes the proof.

Thus coperiod double squares that end outside x come in two kinds, one with "short" s and the other with "long" s. The latter does not further interest us specially, but the former will take part in our enumerative argument later.

Definition 3.6. Let (u, U) be a double square of period x. The double square (v, V) that starts inside u and ends outside u with $u = ab, v = bs$ is said to be *outer* (coperiod) double square of (u, U) if $0 < |s| < |x|$.

Nevertheless, we shall call the coperiod double square of the latter kind, that is, when $|s| > |x^{q+p-1}x_1|$, a *marginal* coperiod one. We will be back to marginal double squares in a later section.

The outer double squares, like the inner ones, have the following remarkable property.

Proposition 3.7. *Let (u, U) a double square of type (p, q), $p > 1$, and (v, V) be an arbitrary outer double square of its, of type (p', q'). Then also $p' > 1$ and (u, U) and (v, V) are coresidue and, moreover, they have the same larger length.*

Proof. We prove first for the case $p > 1$. Let (u, U) and (v, V) have periods x and y, residues x_1 and y_1, respectively. As before, we write $u = ab, v = bs$, $0 < |s| < |x|$.

First, we show that

$$|avv| < |ux^{q+p}x_1|.$$

If it is not so, then $avv = ux^{q+p}x_1s'$, or, $v = tx^{q+p-1}x_1s'$ for some $t, s' \in A^*$ and $x = st$. In particular, $|v| > |x^px_1|$, hence $|s'| < |s|$ and s' is a prefix of s; also $|v| > |xx|$ as $p > 1$. By setting $s = s'd$, we see that v is a suffix of vd, hence, of d^*, or, equivalently, v is a prefix of e^* for a conjugate e of d. But that $v = y^{p'}y_1$ is a prefix of e^* with $|e| < |x| = |y|$, by Periodicity Remark, implies $p' = 1$, therefore $|v| < |xx|$, contradicting $|v| > |xx|$.

Next, we show $p' > 1$. Suppose that $p' = 1$, hence $q' = 1$. At this moment we have $|avv| < |uxxx|$ because $|v| < |xx|$. So $avv = uxxs'$ for some word s', $0 \le |s'| < |x|$. We should have $0 < |ux^{q+p}x_1| - |avv| < |xx|$, or equivalently, $|s'| > |x_1|$, since, otherwise, $|aVV| = |avv\bar{y}^{q'}y^{q'}| = |avv\bar{y}y| \le |ux^{q+p}x_1|$ which shows that $\bar{y}y$ is a factor of xx, so $\bar{y} = y$, a contradiction.

Now, looking back to the first occurrence of v, we set $v = t's$. For the second occurrence of v, we set $v = txs'$, where t' is a suffix of xx_1, $x = st$, s' is a proper prefix of x. If $|t'| \ge |xx_1|$, comparing the two occurrences and Synchronization Remark give us $s' = x_1s$ and $v = txx_1s$ which shows that $|v| = |x^2x_1|$ contradicting the assumption $p' = 1$. Suppose now $|t'| < |xx_1|$ and $xx_1 = et'$ for some nonempty word e. As $|s'| > |x_1|$, x_1 should be a prefix of s': $s' = x_1d$, for some word d, $|d| > 0$, therefore $v = txx_1d$ and

$$xx_1xx_1 = et'stx_1 = evtx_1 = etxx_1dtx_1.$$

By Synchronization Remark, $xx_1 \in \lambda^*$, where λ is the primitive root of dtx_1 and v is a factor of λ^*. Now that txs' is a factor of x^3 and $|dtx_1| = |ts'|$, $|v| = |txs'|$, by Periodicity Remark, x is a power of a conjugate of λ, a contradiction as $|\lambda| \le |ts'| < |x|$. Thus $p' > 1$.

Therefore, for the second occurrence of v, we have $v = tx^is'$, $i > 0$, s' is a proper prefix of x. Due to the presence of x^i and Synchronization Remark, we have two possibilities: $|s'| < |s|$ and $|s'| \ge |s|$, or equivalently, $s' = x_1s$.

Case 1. $|s'| < |s|$.

We see that s' is a suffix of s: $s = ds'$ for a word d. We should have more, $i > 1$, because $i = 1$ implies $|v| = |txs'| < |txs| = |xx| = |yy| < |y^{p'}y_1| = |v|$, as $p' > 1$. Hence, by

Synchronization Remark, $x = x_1 d$ that means $d = x_2$ and $|x_1| = |s'| + |t|$ and $|v| = |x^i x_1|$, or just the same $v = y^i y_1$: (v, V) is coresidue with (u, U).

For the last claim, set $x = s't'$ then $\bar{y} = t's'$. Moreover, if we set $x_1 = s't''$ then $|t''| = |t|$ and $|t''s| = \bar{y}$, hence $\bar{y} = t''s$. Since $avv\bar{y}^{q'}y^{q'}$ and $ux^{q+p}x_1x^q$ and $avv = ux^{p'+1}s'$ are prefix of one another, the same is true for $\bar{y}^{q'}y^{q'}$ and $t'x^{q^p-p'-2}x_1x^q$. However,

$$t'x^{q+p-p'-2}x_1x^q = \bar{y}^{q+p-p'-1}t''x^q$$

which shows that $q' \geq q+p-p'-1$. Further, $|t''| = |t|$ should imply $q' > q+p-p'-1$, otherwise, $t''s = y$. Thus, we have, on one hand, $q' \geq q+p-p'$.

On the other hand, notice that $y = ts$. If $q' > q+p-p'$ then t is a prefix of $\bar{y}$, hence $t = t''$ which leads to $\bar{y} = t''s = ts = y$, a contradiction. So we get $q' \leq q+p-p'$ and, certainly, the equality sign prevails.

Case 2. $s' = x_1 s$.

From $v = tx^is' = tx^ix_1s$ and $x = st$ we get at once $|y_1| = |x_1|$, which is already the first claim, and $p' = i+1$.

Further, set also $x = s't'$, so $|t| = |t'| + |x_1|$ and $|t'x_1s| = |x| = |\bar{y}|$, we have $t'x_1s = \bar{y}$ as both of them are suffixes of v. Note that $y = ts$ and $\bar{y} = t's'$. Further,

$$avv = ustx^is' = ux^{i+1}s'$$

and, consequently

$$aUU = ux^{q+p}x_1x^q = avvt'x^{q+p-i-2}x_1x^q$$

$$avv(t's')^{q+p-i-2}t'x_1stx^{q-1} = avv\bar{y}^{q+p-i-2}\bar{y}tx^{q-1}.$$

Now we see that t is not a prefix of $\bar{y} = t'x_1s$, otherwise $t = t'x_1$ yielding $\bar{y} = ts = y$, which is impossible. This, together with the fact that $\bar{y}^{q'}y^{q'}$ and $\bar{y}^{q+p-i-1}tx^{q-1}$ are prefixes of each other, forces $q' = q+p-i-1 = q+p-p'$, or $|V| = |U|$.

For the case $p = 1$, the argument is more sophisticated and we defer the proof until Theorem 5.1. The proof of Proposition 3.7 is finished here.

We see that the turning point in the proof above is the derivation of that if $p > 1$ then $p' > 1$; the reasoning may be a bit messy and roundabout. However, later, we will intensify the argument and attempt to make it more lucid for the case $p = 1$, therein we shall show that more than just the equality of the larger lengths the squares are in fact isomorphic.

The next assertion will come handy later in estimating the number of squares starting inside u. The following proposition is about the so-called synchronization delay, in disguise; see [1] for the general treatment.

Lemma 3.8. *Let t be a common suffix of x and $\bar{x}$ and s be a common prefix of x and $\bar{x}$. Then $|t| + |s| < |x|$.*

Proof. Suppose for the contrary that $|t| + |s| \geq |x|$ and let s_0 be the prefix of s of length $|x| - |t|$. Since $dt = x_1x_2$ and $s_0e = x_1x_2$ for some words d and e, we have

$$dts_0e = x_1x_2x_1x_2$$

and, at the same time, $d't = x_2x_1$ for some word d' such that $|d| = |d'|$. Now, suppose, for instance, $|t| \geq |x_1|$, we have $|s_0| \leq |x_2|$ and thus $s_0e' = x_2$ for some word e'. This amounts to

$$x_1d'ts_0e' = x_1x_2x_1x_2$$

which contradicts Synchronization Remark as $|d| \neq |x_1d'|$ modulo $|x|$ and $|ts_0| = |x|$. The other possibilities are treated analogously and we should have $|ts| < |x|$ which completes the proof.

We stress the following fact, tacitly mentioned in previous proofs, which will be very useful later, with the Lemma 3.8; to wit

PROPOSITION 3.9. *Let (u, U) be a double square of period x and (v_1, V_1) be an outer double square of (u, U) with $u = ab, v = bs$ then s is a common prefix of x and $\bar{x}$.*

Proof. We retain the notation of the proof of Proposition 3.7. Indeed, we see in the proof of Proposition 3.2 that $t = t'$, where t' is a prefix of x in the representation $v = t'x^is$ and obviously, t is a suffix of u, hence of $x_2x_1 = \bar{x}$. As for s, first, evidently, s is a prefix of x; second, from the proof of Proposition 3.7 we see that always $v = tx^is'$ with s' a prefix of x, $i > 1$ when $|s'| < |s|$, or equivalently, $x_2s' = s$, while, the other alternative, $|s'| \geq |s|$ and $i > 0$, is equivalent to $s' = x_1s$. In both cases, s is a prefix of $x_2x_1 = \bar{x}$. The proof is completed.

The next two Sections are instrumental in promoting the approach of this paper.

4 Non-Coperiod Squares

In this section we study the u-close double squares (v, V) that are not coperiod with (u, U). We formulate the results separately for the cases $p > 1, |v| < |x|$; $p = 1, |v| < |x|$; and $p \geq 1, |v| \geq |x|$, but $|y| \neq |x|$. Throughout, we assume that (u, U) occurs at the position 0.

PROPOSITION 4.1. *Let (u, U) be a double square of type (p, q), $p > 1$, period x and residue x_1 and let (v, V) be a double square of type (p', q'), period y and residue y_1 such that $|v| < |x|$, not coperiod with (u, U), occurring at the position $|a|$ for a nonempty proper prefix a of u with $av = us'$. Then $|s'| > |x|$.*

Proof. Suppose on the contrary that $0 < |s'| \leq |x|$ and set

$$x = s't'.$$

Since vv and $UU = x^{q+p}x_1x^q = ux^qux^q$ both have an occurrence at the position 0, we should have $|avv| > |ux^q| \geq |ux|$ to avoid having vv to occur again, at the position $|ux^qa|$ and, moreover, we can write

$$avv = uxs''$$

for a nonempty proper prefix s'' of x. Set further

$$x = s''t''$$

then

$$v = t's''$$

we also see that $|s'| > |s''|$, hence $|t''| > |t'|$ and t' is a proper suffix of t'' and we can write

$$t'' = ft', s' = s''f$$

for a nonempty word f. By the fact that $aVV = avv\bar{y}^{q'}y^{q'}$ and $UU = ux^{q+p}x_1x^q = ux^3x^{q+p-3}x_1x^q$ both occur at the position 0 , $\bar{y}^{q'}y^{q'}$ and $t''x$ have an occurrence at the same position $|avv|$, it is evident that $\bar{y}^{q'}y^{q'}$ and $t''x = ft's''t'' = fvt''$ are prefixes of each other. We have to handle the following possibilities.

Case 1. $|f| < |y|$.

If $q' > 1$ then $fy_1y_2y_1$, being a prefix of fv, is a prefix of $\bar{y}^2 = y_2y_1y_2y_1$ or has the prefix $y_2y_1y_2y_1$. Then, by $|f| < |y|$, fy_1y_2 is a prefix of $y_2y_1y_2y_1$ or $y_1fy_1y_2$ is a prefix of $y_1y_2y_1y_2$ and by Remark 2.7 (i) we get $f = y_2$. Consequently,

$$fv = y_2(y_1y_2)^{p'}y_1 = (y_2y_1)^{p'+1}$$

which shows that $(y_2y_1)^{p'+1}$ and $(y_2y_1)^{q'}(y_1y_2)^{q'}$ are prefixes of each other. In view of $q' \leq p' < p' + 1$, we get $y_2y_1 = y_1y_2$ which is impossible. So we must have $q' = 1$.

Further, as s' and s'' both are suffixes of v, hence both of them and y_2y_1 are suffixes of one another. We further have to distinguish several possibilities.

Subcase 1.1. $|s'| \geq |y|$. Then y_2y_1 is a suffix of s', and f is a nonempty proper suffix of y_2y_1, as f is a nonempty suffix of s' and $|f| < |y|$ by assumption. If we set $y_2y_1 = ef$, then

$$ev = ay_1y_2y_1 = ay_1ef = us' = us''f,$$

which implies $ay_1e = us''$ and e is a suffix of s'' because, by assumption, $|ef| = |y| \leq |s'| = |s''f|$, or, $|e| \leq |s''|$, therefore, e is a suffix of y_2y_1. At the same time, f is a prefix of t'', therefore, a prefix of y_2y_1, too, because of the assumption $|f| < |y|$. Consequently, $y_2y_1 = fe$ and $fe = ef$. By Remark 2.6 (ii), y_2y_1 cannot be primitive, a contradiction.

Subcase 1.2. $|s'| < |y|$. This time s' is a suffix of y_2y_1 and f is also suffix of y_2y_1. We see that $\bar{y}y_1 = y_2y_1y_1$, being a prefix of $\bar{y}y$, is a prefix of $t''x$ (note that $q' = 1$ and $|yy_1| \leq |v| < |x| < |t''x|$). That is, $ft'x = ft's''t''$ has the prefix $y_2y_1y_1$, and similarly, $t'x$ has the prefix $y_1y_2y_1$. At the same time, as $y_1y_2y_1$ is a prefix of $v = t's''$ we see that $y_2y_1y_1$ is a prefix of $fy_1y_2y_1$, which is a prefix of $ft's$, all the more, of $fy_1y_2y_1y_1$, that is, $y_2y_1y_1$ is a prefix of $(fy_1)^m$ for some positive integer m. But f is a suffix of y_2y_1, hence fy_1 is a suffix of $y_2y_1y_1$. By Remark 2.7 (i), we have

$$y_2y_1y_1 = \mu^n, fy_1 = \mu^i, \quad n > 1, i > 0$$

for a primitive word μ. We should have, by Remark 2.6, that

$$|y_1| < |\mu| < |y|$$

for y is primitive.

Suppose that $p' > 1$. Then $|v| \geq |yyy_1|$, from which $|t''x| > |t's''| = |v| > |yy|$ and we have, as both $\bar{y}y$ and $t''x$ have an occurrence at the position $|avv| = |uxs''|$ that $\bar{y}y = y_2y_1y_1y_2$ is a prefix of

$$t''x = ft'x = fvt'' = f(y_1y_2)^{p'}y_1t''.$$

Thus, $y_2y_1y_1y_2$ is a prefix of $fy_1y_2y_1y_2$. From $0 < |f| < |y|$ it follows that y_1y_2 is a factor of $y_1y_2y_1y_2$ with an occurrence that does not coincide with the first or the last one in yy, so by Remark 2.6 (iii) yields that y_1y_2 is not primitive, which is impossible.

If, otherwise, $p' = 1$ then

$$v = yy_1 = y_1y_2y_1 = \bar{\mu}^n$$

for a conjugate $\bar{\mu}$ of μ. Now that

$$vv = \bar{\mu}^n\bar{\mu}^n = \bar{\mu}^{2n}$$

and xx have the common factor xs'' we have $|s''| < |\mu|$, otherwise $|x| = |\mu|$, by Remark 2.6, which is clearly a contradiction, for, $|x| > |v| = |\bar{\mu}^n| > |\bar{\mu}|$. Further, remind that f is a prefix of y_2y_1 by the assumption $|f < |s'| < |y|$. Since $fy_1 = \mu^i$ and y_1f is a preffix of $v = y_1y_2y_1 = \bar{\mu}^n$, we have $y_1f = \bar{\mu}^i$, which implies, as f is a proper prefix of y_2y_1, that

$$y_2y_1 = f\bar{\mu}^j, \quad j > 0.$$

Set now

$$f = \bar{f}\bar{\mu}^k$$

for $k \geq 0$ and $|\bar{f}| < |\bar{\mu}|$. So, as $|y_1| < |\bar{\mu}|$,

$$\bar{\mu} = y_1\bar{f}$$

and

$$y_2y_1 = \bar{f}\bar{\mu}^{k+j}.$$

In view of the relation $y_1y_2y_1 = \bar{\mu}^n$, we can write

$$y_1y_2 = \bar{\mu}^{k+j}f'.$$

and we have

$$f'y_1 = \bar{\mu}$$

for a word f' with $|f'| = |\bar{f}|$. Note that $n = k + j + 1$. Further, by the relation $y_2y_1y_1 = \mu^n$ it follows

$$\bar{f}(y_1\bar{f})^{k+j}y_1 = \mu^n.$$

However, $|\bar{f}y_1| = |y_1\bar{f}| = |\bar{\mu}| = |\mu|$, so

$$\bar{f}y_1 = \mu.$$

Now consider the two following words

$$fvt'' = t''x$$

and

$$\bar{y}y = y_2y_1y_1y_2$$

which are, as we have observed, prefixes of each other. By substituting the expression for y_2y_1 and y_1y_2 we have,

$$y_2y_1y_1y_2 = f\bar{\mu}^j\bar{\mu}^{k+j}f'.$$

But

$$fy_1y_2y_1 = f\bar{\mu}^n.$$

hence, as $n = k + j + 1$,

$$y_2y_1y_1y_2 = fv\mu^{j-1}f'.$$

Further, note that

$$|t''| = |x| - |s''| > |v| - |s''| = |\mu^n| - |s''| > |\mu^n| - |s''| > |\mu|,$$

as $n > 1$.

If $j > 1$, the prefix $fv\bar{\mu}$ of $y_2y_1y_1y_2$ is also a prefix of $t''s''t'' = fvt''$. But then $\bar{\mu}$ is a prefix of t'', also a prefix of fv, hence, a prefix of

$$fy_1 = \bar{f}\bar{\mu}^k y_1 = \bar{f}(y_1\bar{f})^k y_1$$

which implies $\bar{\mu} = \bar{f}y_1 = \mu$, a contradiction.

If $j = 1$ then $|f'| = |f| < |\bar{\mu}|$ shows that f' is, in its turn, a prefix of t'', and of f, hence $f' = \bar{f}$. This implies $\bar{m}u = f'y_1 = \bar{f}y_1 = \mu$, a contradiction again. This terminates Subcase 1.2 and with it, Case 1.

Case 2. $|f| \geq |y|$.

Now, the fact that fvt'' and $\bar{y}^{q'}y^{q'}$ are prefix of each other shows that y_2y_1 is a prefix of f. While f is a suffix of $v = y^{p'}y_1$, by Remark 2.7 (i), we get $f = (y_2y_1)^n$. Moreover, $fv = t''s''$ is a prefix of $t''x$, we get immediately $n = q'$. Consequently,

$$|V| = |vy^{q'}| = |v| + |y^{q'}| = |f| + |v| = |f| + |t'| + |s''| = |t''| + |s''| = |x|.$$

Since V, occurring at the position $|aV|$ and being of the same length as x, is a conjugate of x, VV is a factor of xxx, hence a factor of $xxx^{q+p-2} = x^{q+p}$. This means that VV occurs again, later (than the obvious one at the position $|a|$), a contradiction. There is no more possibilities and Proposition 4.1 is proved.

Now, the next issue, $p = 1, |v| < |x|$. The proof resembles the previous one but it is more subtle, however, the analogy allows us to make some shortcuts.

Proposition 4.2. *Let the double square (u, U) be of type $(1, 1)$, period x and residue x_1 and (v, V) be a double square of type (p', q'), period y and residue y_1 such that $|v| < |x|$, not coperiod with (u, U), starting inside u and satisfying $av = us$. Then $|s| > |x_1|$.*

Proof. We assume the contrary that $|s| \leq |x_1|$ and seek a contradiction. We set $avv = uxt'$ and

$$x = st, \quad v = ts', \quad x = s't'.$$

We see that

$$|s'| < |s|.$$

Put further

$$s = s'f$$

for appropriate words s', t, t', f. Surely, $|x_1| > |s'|$. Observe that fv is a prefix of $t'x_1$ because $|t'x_1| > |fv| = |fts'| = |t's'|$. Also, for $aVV = avv\bar{y}^{q'}y^{q'}$ we see that $t'x_1$ and $\bar{y}^{q'}y^{q'}$ are prefixes of each other, hence $\bar{y}^{q'}y^{q'}$ and fv are prefixes of each other, too.

Case 1. $|f| < |y|$.

We proceed just as in the proof of Proposition 4.1 for $p > 1$, with the only warning that we have now to dispose of the fact that fv and $\bar{y}^{q'}y^{q'}$ are prefixes of each other, but the assumption $|s| \leq |x_1|$ permits us to handle instead with $t'x_1$, fv and $\bar{y}^{q'}y^{q'}$ and the fact that fv is a prefix of $t'x_1$.

We show that $q' = 1$. Suppose that it is not so, $q' > 1$. First, $fy_1y_2y_1$ is a prefix of $fv = t's'$, hence a prefix of tx_1 (note that s' is a prefix of x_1) and, next, as $t'x_1$ and $y_2y_1y_2y_1$ (by $q' > 1$) are prefixes of each other, $fy_1y_2y_1$ and $y_2y_1y_2y_1$ are prefixes of each other either.

If $|y_2y_1y_2y_1| < |fy_1y_2y_1|$ then

$$fy_1y_2y_1 = y_2y_1y_2y_1d$$

for some nonempty word d. By $|f| < |y|$ we get

$$ey_2y_1d = y_1y_2y_1$$

for some word e. Now by Synchronization Remark, $d \in (y_1y_2)^*$, which is absurd by length equality.

If $|y_2y_1y_2y_1| > |fy_1y_2y_1|$, similarly,

$$fy_1y_2y_1d = y_2y_1y_2y_1.$$

Again by Synchronization Remark $d \in (y_2y_1)^*$, which is impossible as $|d| > 0$.

If $y_2y_1y_2y_1 = fy_1y_2y_1$ then

$$fv = fy_1y_2y_1(y_2y_1)^{p'-1} = y_2y_1y_2y_1(y_2y_1)^{p'-1} = (y_2y_1)^{p'+1}$$

and $\bar{y}^{q'}y^{q'}$ cannot be prefixes of each other, unless $\bar{y} = y$, which overly impossible. Consequently, $q' = 1$.

Subcase 1.1. $|s| \geq |y|$. The same contradiction is obtained by just the same reasoning as in the proof of Proposition 4.1 Subcase 1.1 for $p > 1$.

Subcase 1.2. $|s| < |y|$. We have

$$|t'x_1| > |t's'| = |fv| > |v| \geq |y_2y_1y_1|$$

so that $t'x_1$ has $\bar{y}y_1 = y_2y_1y_1$ as a prefix. Likewise,

$$|tx_1| \geq |ts'| = |v|$$

so that tx_1 has $y_1y_2y_1$ as a prefix. Now by the same reasoning as in the proof of Proposition 4.1 Subcase 2.2 for $p > 1$, we have

$$y_2y_1y_1 = \mu^n, fy_1 = \mu^i$$

for $n > 1$, $i > 0$ and μ primitive.

Suppose that $p' > 1$ then

$$|t'x_1| \geq |fv| = |fy_1y_2y_1y_2|$$

that implies that $\bar{y}y = y_2y_1y_1y_2$ is a prefix of $t'x_1$ and, hence a prefix of fv and $fy_1y_2y_1y_2$. Analogously, as in Proposition 4.1, Subcase 1.2 for $p > 1$, this leads also to a contradiction that y is not a primitive.

For the remaining possibility $p' = 1$, we write $v = yy_1 = \bar{\mu}^n$ for a conjugate $\bar{\mu}$ of μ and $fy_1 = \mu^i$.

Subcase 1.2.1. $|s| < |\mu|$. Proceed just as in the proof of Proposition 4.1 for $p > 1$, summarizing, we get

$$fy_1 = \mu, y_1f = \bar{\mu}\, y_1y_2y_1 = \bar{\mu}^n;$$
$$y_2y_1 = (fy_1)^{n-1}f, y_1y_2 = f(y_1f)^{n-1}.$$

Set

$$y_1y_2 = \mu^{n-1}f'$$

for a word f' satisfying

$$|f'| = |f|, f'y_1 = \bar{\mu}.$$

Analogously, we can prove

$$|s'| < |\mu|$$

and

$$|t'| = |x| - |s'| > |v| - |s'| > |f|.$$

Further, we consider

$$\bar{y}y = y_2y_1y_1y_2 = f\bar{\mu}^{n-1}\bar{\mu}^{n-1}f' = f\bar{\mu}^n\bar{\mu}^{n-2}f' =$$
$$= fvf'(y_1f')^{n-2}.$$

Like in Proposition 4.1 Subcase 1.2.1 for $p > 1$, but with $t'x_1$ in place of $t''x$, we have

$$|t'x_1| > |t's| = |t's'| + |f| = |fv| + |f'|.$$

Since $t'x_1$ and $\bar{y}y$ are prefixes of each other, the prefix of length $|fvf'|$ of $\bar{y}y$, that is fvf', is a prefix of $t'x_1$ that means that f' is a prefix of t'. Due to $|f'| = |f|$ we get $f' = f$ that means $y_1y_2 = y_2y_1$, which is impossible.

Subcase 1.2.2. $|s| > |\mu|$. Here we have first

$$y_1f = \bar{\mu}^i, y_2y_1 = f\bar{\mu}^j.$$

Set

$$f = \bar{f}\bar{\mu}^k, k \geq 0, |\bar{f}| < |\bar{\mu}|.$$

Therefore

$$\bar{\mu} = y_1\bar{f}.$$

and

$$y_2y_1 = \bar{f}\mu^{k+j}.$$

We can write

$$y_1y_2 = \bar{\mu}^{k+j}f'$$

with $|f'| = |f|$ for a word f'; so $f'y_1 = \bar{\mu}$. Note that $n = k + j + 1$ and $\bar{f}y_1 = \mu$. Reasoning as in Proposition 4.1 Subcase 1.2.2 for $p > 1$, we get

$$y_2y_1y_1y_2 = f\bar{\mu}^j\mu^{k+j}f' = f\bar{\mu}^n\bar{\mu}^{j-1}f'$$

and

$$fv = fy_1y_2y_1 = f\bar{\mu}^n.$$

That is

$$y_2y_1y_1y_2 = fv\bar{\mu}^{j-1}f' = fvf'(y_1f)^{j-1}.$$

Reasoning just as at the end of Subcase 1.2.1 we get that f' is a prefix of t', therefore $f' = \bar{f}$, or $\bar{\mu} = \mu$, or $y_1y_2 = y_2y_1$, a contradiction.

Case 2. $|f| \geq |y|$.

Completely analogous to the proof of Proposition 4.1 Case 2, we have that

$$f = (y_2y_1)^n, \quad n > 0$$

and due to the fact that fv and $\bar{y}^{q'}y^{q'}$ are prefixes of each other

$$n = q' : f = (y_2y_1)^{q'} = \bar{y}^{q'}.$$

Therefore, as $V = vy^{q'}$, we have

$$|V| = |v| + |f| = |v| + |s| - |s'| = |t| + |s'| + |s| - |s'| = |t| + |s'| = |x|.$$

Furthermore, from the relation $aV = avy^{q'}$, and by the fact that $|s| \leq |x_1|$ and $|tx_1| \geq |x| > |v| > |y^{q'}|$, and as $aVV = aVvy^{q'}$ again, we have

$$aVv = avy^{q'}v = avv\bar{y}^{q'} = usts'f = uxs.$$

This implies that $y^{q'}$ is a prefix of tx_1, from which

$$aVV = uxs'y^{q'}.$$

Put now $tx_1 = y^{q'}V'$ and notice that $aVV = avy^{q'}V = usy^{q'}V$ that show that $y^{q'}V$ and tx, hence $y^{q'}V$ and tx_1, are prefixes of each other. This means that V' and V are prefixes of each other.

If $|V'| \geq |V|$ then V is a prefix of V' and V is a factor of tx_1 and more than that VV is a factor of

$$txx_1 = tstx_1$$

as $|V| = |x|$ that evidently shows that VV occurs later, a contradiction.

If $|V'| < |V|$ then V' is a prefix of V. By the equality

$$aVVV' = uxxx_1$$

we calculate

$$|VV'| = |xxx_1| - |s| - |y^{q'}|$$

for $usy^{q'} = avy^{q'} = aV$, so

$$|VV'| = 2|st| + |x_1| - |s| - |f| = 2|t| + |s'| + |x_1| > 2|t| + |s'| + |s'| = 2|ts'| = 2|v|.$$

Consequently, this means, as vv is a prefix of VV', that this time vv occurs later! A contradiction. Proposition 4.2 is proved.

Finally, the last issue $p \geq 1$, $|v| \geq |x|$, but $|y| \neq |x|$.

PROPOSITION 4.3. *Let the double square (u, U) be of type (p, q), $p \geq 1$, period x and residue x_1 and (v, V) be a double square of type (p', q'), period y and residue y_1 such that $|v| \geq |x|$, $|y| \neq |x|$, not coperiod with (u, U), starting inside u and satisfying $av = us$. Then $|s| > |x|$.*

Proof. Again we suppose on the contrary that

$$|s| \leq |x|$$

which we show lead to contradictions. We write

$$x = st.$$

Since $aVV = avyvy = (us)(yv)y$ (meaning that yv occurs at the position $|us|+1$ of aVV, or the same, of UU), we must have

$$|yv| > |tx^{q+p-1}x_1|.$$

In fact, otherwise, x^* and y^* would have a common factor of length $|yv|$, which is no less than $|y| + |x|$, so by Periodicity Remark y and x, being both primitive, are conjugates, $|y| = |x|$ which is against the assumption.

Now we can write

$$yv = tx^{q+p-1}x_1s''$$

for some word s''.

If $s = s''$, then s'' is also a prefix of x. Since

$$|tx^{q+p-2}x_1| < |y|$$

also by Periodicity Remark, we can write

$$y = tx^{q+p-2}x_1s'$$

for some word s'. Now that $aVV = (avyv)(y)$ (meaning, likewise, that this last occurrence of y starts at the position $|avyv|+1$ of aVV, or of UU), we have

$$avyvy = x^p x_1 x x^{q+p-1} x_1 s'' t x^{q+p-2} s' = x^p x_1 x^{q-1} x^p x_1 x^p (x_1 x_2)^{q-1} s' =$$

$$= x^p x_1 x^{q-1} (x^p x_1)(x^p x_1)(x_2 x_1)^{q-1} s'$$

that shows that $x^p x_1 x^p x_1 = uu$ occurs later, a contradiction.

If $|s''| > |s|$ then set

$$s'' = sd, \quad |d| > 0.$$

Further, set $u = ab$, where $0 < |b| < |u|$, then $v = bs$. Now that

$$yv = v\bar{y} = tx^{q+p-1} x_1 s'' = tx^{q-1} x^p x_1 sd = tx^{q-1} abx_1 sd$$

which implies

first, $|y| = |\bar{y}| = |tx^{q-1}a| + |d|$;
second, v is a suffix of vd and
third, v is a proper prefix of $tx^{q-1}av$.

Putting all this together, we have $\bar{y} = ed$, that is

$$ve = tx^{q-1}av$$

for some nonempty word e, hence v is a suffix of ve. Consequently, v is a common suffix of d^* and e^*. Because of $|v| > |y| = |d| + |e|$ we get that d and e copower, hence $\bar{y}$ is not primitive as both of them are nonempty, a contradiction.

We assume from now on that

$$0 < |s''| < |s|$$

or

$$|tx^{q+p-1}x_1| < |yv| < |tx^{q+p-1}x_1 s|.$$

Under this assumption, we have

$$|y| < |tx^{q+p-1}x_1|$$

for, otherwise, $|v| < |x|$ contradicting the hypothesis. Further, we write yv as

$$yv = tx^{q+p-1}x_1 s''$$

for a nonempty proper prefix s'' of x and set $x = s''t''$. We distinguish three cases.

Case 1. $|y| \le |tx^{q+p-2}x_1|$.

As has been noted above, $y^{p'+2}$ and x^{q+p+1} have the common factor $tx^{q+p-1}x_1$ whose length is $|tx^{q+p-2}x_1| + |x_2 x_1|$ which is equal to or greater than $|y| + |x|$, therefore x and y are (conjugates) of the same length, in spite of the assumption.

Case 2. $|tx^{q+p-2}x_1| < |y| \le |tx^{q+p-1}|$.

Then we can write y in the form

$$y = tx^{q+p-2}x_1 s'$$

for a nonempty prefix s' of x_2 and set thus $x_2 = s't'$. Set further $s = s''d$.

As v is a suffix of vd we see that v is a suffix of $d^* = \mu^*$, where μ is the primitive root of d, $d = \mu^i$, $i > 0$. We stress that d is a suffix of v. Again $aVV = (avyv)(y)$, which implies that t'' and y are prefixes of each other. Concerning y and μ, we have three issues.

Subcase 2.1. $|\mu| \leq |y|$. Then μ is a prefix of y, but y is a prefix of v, by synchronization $v = \mu^n$, $n > 1$ (as $|v| \geq |x| > |d| \geq |\mu|$). Therefore $p' = 1$, $v = yy_1$.

Suppose that
$$|y| \geq |t''|$$
which is equivalent to
$$|x^{q+p-2}s'| \geq |d|.$$
In fact, from
$$|s''| + |y| \geq |s''| + |t''| + |x|$$
it follows
$$|s''| \geq |x| - |y| = |s''| + |t''| - |x_1 s'| - |t| - |x^{q+p-2}|.$$
Therefore
$$|x^{q+p-2}s'| \geq |t''| - |t| = |d|$$
as claimed.

Since vv has the prefix vy, which in its turn has the factor $x^{q+p-1}x_1 s'$, and at the same time, x^{q+p} has also the factor $x^{q+p-1}x_1 s'$, which verifies
$$|x^{q+p-1}x_1 s'| = |x| + |x^{q+p-2}x_1 s'| \geq |x| + |d| \geq |x| + |\mu|$$
we get that x is a power of a conjugate of μ, a contradiction with the primitivity of x.

Suppose, else, that
$$|y| < |t''|.$$
That x is a factor of yv, y is a prefix of t'' and $ts''y$ is a prefix of yv, by synchronization, gives us
$$ts'' = y^m, \quad m > 0$$
and, consequently,
$$s'' = x^{q+p-2}x_1 s' y^{m-1}.$$
It follows $q + p - 2 = 0$, $t' = y^{m-1}t''$ and $|s''| > |y^{m-1}|$. As $v = yy_1$ and
$$|v| > |t'| + |s''| > |y^{m-1}| + |y| + |y^{m-1}| = |y^{2m-1}|$$
so $m = 1$, that is $s'' = x_1 s'$ and $t' = t''$. Further, for $t'' = dt$ and
$$|t'' x_1 s'| = |t' x_1 s''| = |v|$$
we have
$$|dy| = |dts''| = |t''s''| < |v|.$$

From $dv = vd$ we see that dy is a proper prefix of v. By synchronization, $d \in y^*$, hence $\mu = y$, or $v = yy_1 = y^n$, a nonsense.

Subcase 2.2. $|y| < |\mu|$. Under this condition y is a prefix of t'', and furthermore, a proper prefix of $d = \mu^i$. Besides, note that d is a suffix of v, therefore $d = y^j y_1$, $j > 0$ by synchronization that implies $|d| \geq |yy_1| \geq |v|$, a contradiction.

Case 3. $|tx^{q+p-1}x_1| < |y| < |tx^{q+p-1}x_1|$.

THen we write y as

$$y = tx^{q+p-1}s'$$

for a nonempty proper prefix s' of x_1 and denote $x_1 = s't'$ for some word t'. We maintain the notation

$$s = s''d, x = s''t''$$

and now

$$y = tx^{q+p-1}s'$$

and

$$v = t's''.$$

Further, as $aVV = (avyva)(y)$ we see that d is a prefix of y. Also, easily, from the assumption we get $|t'| \geq |t''| > |d|$. Denote by d'' the prefix of t', hence of y, of length d, $|d''| = |d|$. This ensures that $d = d''$. Setting $t' = d''d'$ for the suffix d' of t', we have $d's = v$.

Now we see that

$$dv = d''v = d''d's = t's = t's''d = vd$$

which implies

$$v = \mu^n, d = \mu^i, \quad n > 1, i > 0$$

for the primitive root d of μ. Note that $|\mu| < |x|$.

Consider now the word xx which is a factor of vy, that is, of μ^*. By Periodicity Remark $|x| = |\mu|$ as both of them are primitive, a contradiction. This case is done and Proposition 4.3 is proved.

Summarizing, we reformulate all we have proven in the three propositions above in the following theorem.

THEOREM 4.4 *Let the double square (u, U) be of type (p, q), $p \geq 1$, period x and residue x_1 and (v, V) be a double square, not coperiod with (u, U), starting inside u and satisfying $av = us$. Then $|s| > |x|$ if $p > 1$ and $|s| > |x_1|$ if $p = 1$.*

5 Coperiod Squares of Low Type

We continue in this section the study of outer double squares, now for the case of low type, namely, when the double square (u, U) is of type 1 or 2. The case of type at least 3 has been described in the Proposition 3.7 that they are coresidue with (u, U); the outer squares of type 1 and 2 will be shown here, more than that, to be replicas of (u, U). First, the harder case.

THEOREM 5.1. *Let the double square (u,U) be of type $(1,1)$ and (v,V) be an outer double square of (u,U). Then (v,V) is a replica of (u,U).*

Proof. Assume that (u,U) occurs at the position 1 and suppose that (v,V) occurs at the position $|a|$ for a proper nonempty prefix a of u. Let (u,U) have period x and residue x_1 and (v,V) have type (p',q'), period y and residue y_1 with $u=ab$, $v=bs$, so $av=us$, where $0<|s|<|x|$. Set $x=st$ so that $y=ts$. First, observe that $|y^{p'}y_1|=|v|<|us|<|xx_1x|<3|x|$ which shows that $p'\le 2$.

We first prove that $q'=1$. Suppose the contrary that $q'=2$, then $p'=2$. Because the equalities

$$aVV = avyyvyy, UU = uxxx_1x$$

showing that y and tx_1x both have an occurrence at the position $|avy|=|usts|=|uxs|$ we see that y is a prefix of tx_1x. Note that from

$$|xx| = |yy| < |yyy_1| = |v| = |bs| < |b| + |x|$$

we get $|b|>|x|$ and $|a|=|xx_1|-|b|<|x_1|<|x|$, a is a prefix of x: $x=ad$ for a word d. Now consider the following three occurrences of v: the occurrence at the position $|av|=us$, the occurrence at the position $|uxa|=|usta|$ and the occurrence at the position $|avyy|=|uxxs|=|uxads|$; v has these occurrences because first $aVV=ava\bar{y}^2y^2$; second, $UU=uxux=uxabst=uxavt$; and third, $aVV=avyyvyy=uststsvyy=uxxsvyy$ and because aVV and UU both occur at the position 1. This shows that, v is a prefix of tav and is a prefix of dsv at the same time. In turn, these relations imply, as $|ta|>0$ and $ds>0$, that v is simultaneously a prefix of $(ta)^l$ and $(ds)^m$ for positive integer l and m. But $|ta|+|ds|=|xx|=|yy|<|v|$, by Remark 2.6 ta and ds are copower. However, $tads=txs=yy$ forcing $ta=y$ as y is primitive, or just the same, $a=s$. But then $|b|=|xx_1|-|a|=|tx_1|$ and $|v|=|bs|=|tx_1s|=|xx_1|<|xx|=|yy|<|v|$, a contradiction. So $q'=1$. Further, we have two cases to consider.

Case 1. $|yv|\le|txx_1|$.

Then $3|x|=3|y|>|yv|$ that means $v=yy_1$, $p'=1$. Note that $|yv|>2|y|>|tx|$. Since the latest occurrence of $VV=vyvy$ starts at the position $|a|$ and that of $UU=uxux=ustxx_1x$ starts at the position 1, we see that tx, yv and txx_1 all have an occurrence at the same position $|us|=|av|$, we can write $yv=txs'$, or

$$v = ts' = yy_1 = tsy_1,$$

then $s'=sy_1$ for a prefix s' of x_1. Set $x_1=s't'$. From $|s'|>|s|$ we get that $|t'|=|x_1|-|s'|<|x|-|s'|<|x|-|s|=|t|$ and $|t's|<|ts|=|y|$.

As v is a suffix of xs', hence a factor of $u=xx_1$, we see that v has an occurrence ending at the position $|xs'|-1$; on the other hand v has an occurrence, by assumption, ending at the position $|us|-1=|xx_1s|-1=|xs't's|-1$, hence v is a suffix of $vt's$ and, as $|t's|>0$, v is a suffix of $(t's)^m$ for a nonnegative integer m. Now observe that $t's$ has an occurrence starting at the position $|uxxs'|$ and y has an occurrence starting at the position

$|a| + |vyv| = |avyv| = |ustxs'| = |uxxs'|$, so $t's$ is a prefix of y, hence of v. Now, by Remark 2.7 (ii), we have

$$t's = \mu^i, v = y_1y_2y_1 = \mu^n, \quad i > 0, n > 1$$

for the primitive root μ of $t's$. Further $tstsy_1 = yyy_1 = yv = txs' = tsts'$ implies $sv = xs'$, while

$$vt's = \mu^{i+n}$$

which shows that

$$xs't' = svt' = \lambda^{i+n}$$

for another conjugate λ of μ, that is

$$xx_1 = \lambda^{i+n}.$$

However, $|x_1| = |s't'| > |st'| = |\mu^i|$ and x_1 is a prefix of x, we conclude that $x = \lambda^j$ for $j > 1$, contradicting the primitivity of x. Case 1 does not take place.

Case 2. $|yv| > |txx_1|$.

As before, $aVV = avyvy = usyvy = xx_1syvy$ and $UU = xx_1xxx_1x = xx_1stxx_1x$ we see that yv, txx_1 and txx_1x all have an occurrence starting at the same position $|xx_1s|$, so we write $yv = txx_1s'$, or $v = tx_1s'$, for a nonempty word s' which is a prefix of s if $|s'| \leq |x|$ or has x as a prefix in the opposite case.

Subcase 2.1. $|s'| < |s|$. Set $s = s'd$, note that $0 < |d| < |y| = |x|$. Evidently, v is a suffix of vd so that v is a suffix of d^m for an integer $m > 0$ and we have again $p' = 1$ and $v = yy_1$. On the other hand, d is is prefix of t, hence of y and of v, so by Remark 2.7 (ii) $v = \mu^n$ for the primitive root μ of d, $n > 1$ and $d = \mu^i$, $i > 0$. Observe further that xs is a prefix of xx and $xs = sy$ is a factor of $vv = \mu^{2n}$, whose length is $|x| + |s| > |x| + |d| \geq |x| + |\mu|$. It follows that x is a proper power of a conjugate of μ, a contradiction. Subcase 2.1 does not happen.

Subcase 2.2. $|s| < |s'|$. We first treat the case $|s'| > |x|$. Then x is a prefix of s': $s' = xs''$ and $v = tx_1xs''$ for a word s''. On the other hand $v = bs$ where, by assumption, b is a suffix of xx_1; the equality $tx_1xs'' = bs$ implies by Remark 2.6 (iii) that s'' is a proper suffix of s, because of the primitivity of x. Now

$$x_1atx_1xs''t = x_1abst = x_1xx_1x$$

with $|s''t| > 0$, $|x_1at| > 0$ implies, by Remark 2.6 (ii), that $x_1x = \mu^n, n > 1$ and $s''t = \mu^i$, $i > 0$ for a primitive word μ. We must have $|s''| < |x_1|$, otherwise x_1 is a prefix of s'' and the occurrence of $uu = xx_1xx_1$ starting at the position ux is later than that at the first one. So s'' is a prefix of x_1, hence a proper prefix of s, $s = s''d$ for a nonempty word d. By the equality

$$vd = tx_1xs''d = tx_1xs = tabs = tav$$

we see that v ia a suffix of vd, thus v is a suufix of d^m for an integer $m > 0$. But $|d| < |x| = |y|$ and y is primitive from which it follows that $p' = 1$, and $|y_1| = |t| + |x_1| + |s''|$.

Now that vv has an (the latest, in fact) occurrence at the position $|a|$ as well as $bxxx_1$ does, we get

$$|avv| = |usv| = |usyy_1| = |u| + |s| + |ts| + |t| + |x_1| + |s''| = |uxxx_1| + |s''|$$

$$> |uxxx_1| = |abxxx_1|$$

or, equivalently

$$|vv| > |bxxx_1|$$

which says that xxx_1 is a factor of vv. Hence, xxx and vv have the common factor xxx_1 of length $|xxx_1| = |x| + |v|$; by Remarks 2.6 and by the primitivity of x, v is a power of a conjugate z of x which is impossible because $0 < |y_1| < |y| = |x| = |z|$.

Now suppose $|s'| \leq |x|$. Set now $s' = sd$ for some word d with $0 < |d| < |x|$. Like Subcase 2.1, as t, hence d, is a prefix of v, we get that

$$v = \mu^n, d = \mu^i, \quad n > 1, i > 0$$

for a primitive word μ. Next, as $v = tx_1s'$ has an occurrence starting at the position $|av|$ and txx_1 has an occurrence starting at the position $|us| = |av|$ we see that tx_1s' is a prefix of txx_1, in view of $|s'| \leq |x|$. Therefore, stx_1s' is a prefix of $stxx_1 = xxx_1$ which ia a prefix of xxx; on the other hand $stx_1s' = sv$ is a suffix of vv, that is $vv = \mu^{2n}$ and xxx have the common factor $stx_1s' = xx_1s'$ of length

$$|x| + |x_1| + |s'| > |x| + |s'| > |x| + |d| \geq |x| + |\mu|.$$

Therefore x is a proper power of a conjugate of μ, a contradiction. Altogether, Subcase 2.2 does not happen either.

Subcase 2.3. $|s'| = |s|$. This time, $|v| = |tx_1s'| = |tx_1s| = |xx_1| = |u|$ and $|V| = |vy^{q'}| = |vy| = |ux| = |U|$ meaning that (v, V) is a replica of (u, U) which finishes the proof.

For the analogous result for $p = 2$, we have already prepared all the prerequisites so that the proof is immediate.

THEOREM 5.2. *Let the double square (u, U) have type 2 then every outer coperiod double square (v, V) of (u, U) is isomorphic to it.*

Proof. Let (u, U) have type (p, q), $p = 2$, period x and residue x_1 and (v, V) have type (p', q'), period y, $|y| = |x|$, and residue y_1 with $av = us$, where $|s| < |x|$. By Proposition 3.7, first, $p' > 1$, second, (v, V) is coresidue with (u, U) and third, $|V| = |U|$.

The second point, together with $|s| < |x|$ guarantees that $p' < 3$, which in turn, together with the first point, guarantees that $p' = 2 = p$. Finally, this, together with the third point, shows that $q' = q$. All together, this means that (u, U) and (v, V) are isomorphic and Theorem 5.2 is proved.

6 More on Non-coperiod Squares: U-mean and U-size Squares

We consider in this section the squares that start inside u and ends outside, but not too "close" to u and not too "far" from u. We make all this precise in the following definition.

DEFINITION 6.1. Let (u, U) be a double square of type (p, q), period x and residue x_1 and vv be a square starting inside u at the position $|a| + 1$ for a proper prefix a of u, such that $u = ab$, for a nonempty word b. Let the first occurrence of v end outside u, so that $v = bs$ for some word s with $|x| \leq |s| \leq |x^{q+p-1}x_1|$. Then vv is a U-*mean* square if and only if $|b| < |x|$ and a U-*size* square if and only if $|b| \geq |x|$. A double square (v, V) is said to be U-*mean* if vv is a U-mean square, and similarly, U-*size* if vv is a U-size square.

By the assumption $|x| \leq |s| \leq |x^{q+p-1}x_1|$, clearly, U-mean, as well as U-size double squares both are not coperiod with (u, U). Also, easily, it can be shown later that if vv is a U-size square then $|avv| > |ux^{q+p}x_1|$. In retrospect, if a double square (v, V) starts inside u with

$$u = ab, v = bs, \quad |s| \leq |x^{q+p}x_1|$$

and if $p > 1$, then it is coperiod with (u, U) if and only if $|s| < |x|$, by Theorem 3.5 and 4.4. Thus the Definition 6.1 furnishes a classification of the double squares starting inside u with $|s| \leq |x^{q+p-1}x_1|$ that they consist of (u, U) itself, the outer and the inner, the U-mean and the U-size double squares.

We first analyze the U-mean squares. We distinguish two kinds of them: the first with $|avv| \leq |ux^{q+p}x_1|$ and the second with $|avv| > |ux^{q+p}x_1|$. Intuitively, the U-mean squares of first kind end inside the interval $UU[|u| + 1, |ux^{q+p}x_1|]$ of UU and those of second kind end outside it.

The following two assertions justify in some sense the name.

PROPOSITION 6.2. *Let (u, U) be a double square of type (p, q), period x and residue x_1, and vv be a U-mean square of the first kind, with $u = ab$, $v = bs$, $svc = x^{q+p}x_1$. Then $v = z^n$ for some conjugate z of x and $v = \frac{q+p+1}{2}$ if $q + p$ is odd and $v = \frac{q+p}{2}$ if $q + p$ is even. Moreover, $|c| < |x|$, $|b| = |x_1| + |c|$ if $q + p$ odd and $|b| + |x_1| = |c|$ if $q + p$ even.*

Proof. We notice at once that vv and x^* have the common factor sv, which is no shorter than $|x| + |v|$, therefore by Periodicity Remark, v is a power of a conjugate z of x, $v = z^n$ that is the first claim. The last one is equally simple: if $|c| \geq |x|$ then c has a prefix z and vvz is still a factor of $ux^{q+p}x_1$, hence $vvy = z^{2n+1}$ has a suffix vv despite the assumption on the latest occurrence of vv. So $|c| < |x|$. Finally, as $|b| + |x^{q+p}x_1| = |vv| + |c| = |x^{2n}| + |c|$, an easy computation will show the second claim and the proof is complete.

As for the U-mean squares of the second kind, the reasoning is more lengthy. We assume $p > 1$ and leave the case $p = 1$ open for the sake of completeness.

PROPOSITION 6.3. *We assume the same hypothesis on (u, U) as in Proposition 6.2. Let vv be a U-mean square of the second kind, $avv = ux^{q+p}x_1s'$, for some nonempty word s'. Then $|s'| < |x|$ and $v = z^n$ for some conjugate z of x. Besides, $q + p$ is odd and $n = \frac{q+p+1}{2}$.*

Proof. Set

$$x^{q+p}x_1 = sc.$$

Suppose, otherwise, that $|s'| \geq |x|$. Then s' has x as a prefix, because $UU = ux^{q+p}x_1x^q$ and avv are prefixes of each other. Next, $|s'| < |s|$ because $|b| < |x|$ and $|x| \leq |c|$ by definition, which implies that s' is a proper suffix of s, a factor of x^*, which, therefore, implies further that $s = x^i s'$, for some $i > 0$, by Synchronization Remark. Comparing then the two occurrences of v yields $c = bx^i$. This is a contradiction $x = \bar{x}$ as c is a suffix of x^*x_1. Therefore $|s'| < |x|$.

Now we write $s = x^i s''$, where s'' is a proper prefix of x, and we shall show that $|s'| \leq |s''|$. Suppose the contrary that $|s'| > |s''|$.

If $i > 1$, because c is a suffix of x^*x_1, the Synchronization Remark gives us $s' = x_2s''$ and $s = x^i s'' = x^{i-1}x_1s'$ from which we get $s'c = (x_2x_1)^n$ and $cs' = z^n$ for some conjugate z of x. Therefore $|s'b| = |v| - |x^{i-1}x_1| = |x_2x^j|$ for some j and $s'b = x_2x^j$. Now that $vv = z^{2n}$ is a factor of UU, we should have $|vv| < |U| + |x$, or, $|s'b| < |x|$ forcing $s'b = x_2$. But $s' = x_2s''$, hence $|b| = 0$ despite the definition.

If, else, $i = 1$ then put $s't' = x$ for some nonempty word t'. By $cs' = v$ we see that

$$vt's'' = cs't's'' = cxs''$$

has the prefix bxs'', hence v is a prefix of $(t's'')^*$, hence of $(\lambda)^*$, for the primitive root λ of $t's''$ which is shorter than $|x|$ because $|s'| > |s''|$. Then the fact that λ^* and x^* have the common factor c imply, by Periodicity Remark, that $|c| < |\lambda| + |x|$ in order for x to avoid being a power of a conjugate of λ. That means $|c| = |x^{q+p}x_1| - |xs''| < |\lambda| + |x|$, or, due to $p > 1$, $|s''| + |\lambda| > |xx_1|$, that implies $|s''| > |x_1|$ and x_1 is a prefix of s''. At the same time, also by $p > 1$ we get $|c| > |xx_1|$ and xx_1 is a suffix of c.

Now put $c = bd$, then $ds' = xs'' = s't's''$ and $|d| = |t's''| < |x|$. Put further $x = de$ and consider the word bxx_1, which is, surely, a factor of v. To wit

$$bxx_1 = bdex_1 = cex_1$$

obviously has the suffix xx_1ex_1 as xx_1 is a suffix of c. Thus xx_1 is a suffix of xx_1ex_1, or xx_1ex_1 is a suffix of $(ex_1)^*$. On the other hand, xx_1ex_1 is a factor of λ^*, with $|xx_1| > |\lambda|$ which guarantees again by Synchronization Remark that ex_1 is a power of a conjugate of λ. In particular $|ex_1| \geq |\lambda|$, and in turn

$$|xx_1| = |d| + |ex_1| = |t's''| + |ex_1| \geq 2|\lambda|.$$

Consequently,

$$2|\lambda| \leq |xx_1| < |s''| + \lambda$$

gives us $|s''| > |\lambda|$.

Finally, the fact that xs'' is simultaneously a factor of λ^* as a factor of v and a factor of x^*, of length $|x| + |s''|$ more than $|x| + |\lambda|$ leads us to the contradiction that x is a power of a conjugate of λ which totally shows that $|s'| \leq |s''|$. For this instance we subsequently determine which U-mean squares may be available.

Since $v = bx^i s'' = cs'$, we see that bx^i is a prefix of c and we can write $c = bx^i d$ for some word d such that $s''d = s'$ and $0 < |d| < |s''|$, as $|s'| > 0$. Further, c being a suffix of x^*x_1

and $x^i d$ being a suffix of c imply that $d = x^* x_1$, by Synchronization Remark, hence $d = x_1$ as $|d| < |x|$. Now from

$$x^{q+p} x_1 = x^i s'' c = x^i d s' c = x^i x_1 s' c$$

it follows $s'c = (x_2 x_1)^n$, for an integer $n > 0$ and $s'b = x_2 x^*$.

Now the fact that $UU = x^p x^{q+p} x_1 x^q$ has the factor $bx^i s'' cs' = vv$ (note that s' is a prefix of x) and that $v = cs'$ is a conjugate of $(x_2 x_1)^j = s'c$ let us see that x^* also has vv as a factor. Consequently, $|vv| < |U| + |x|$, if not, $U \in z^*$, for some conjugate z of x which is a contradiction with $0 < |x_1| < |x|$. This means $|s'b| < |x|$ and, actually, $s'b = x_2$. Sum up, we have

$$vv = bx^i s'' cs' = bx^{q+p} x_1 s',$$

or

$$|vv| = |x^{q+p} x_1 s' b| = |x^{q+p+1}|,$$

$q + p + 1 = 2n$, as desired to prove.

In the definition of a U-mean square we require that $|b| < |x|$, but as matter of fact, it satisfies a stronger inequality which will take part in our enumerative argument later. Let denote γ_m the number of U-mean (both first and second kind) squares of (u, U)

Proposition 6.4. *Let (u, U) be a double square of type (p, q) with period x and residue x_1 and (v, V) be an outer double square of (u, U) with $u = ab$, $v = bs$, and $v'v'$ be a U-mean square with $u = a'b'$, $v' = b'x^i s'$ as usual. Then $|b'| + |s| < |x|$. As a consequence, $\gamma_m + |s| < |x|$.*

Proof. By contradiction, we suppose that $|b'| + |s| \geq |x|$. We take e as the suffix of length $|x|$ of $b's$. Since (v, V) is coperiod with (u, U), say, of period y we see that $e = \bar{y}$ as a suffix of length $|x|$ of v'. As $|v'v'| = |b'x^i s' v'| > |b'sy|$we get that $b'sy$, hence, ey is a factor of $v'v' = z^{2n}$. This implies $e = y$, or $\bar{y} = y$ which is impossible.

Now, observe that at a given position of u could start at most one U-mean square of either kind, but not both and $|b'| > 0$ by assumption, from which the claimed inequality immediately follows. The proof is complete.

Now consider the case of U-mean double squares. Let (v, V) be a U-mean double square with, as before, $v = z^n$ and $v = ds'$ for the first occurrence of v and $u = a'd$. Observe that we always have $n > 1$ because $|v| = |ds'| > |x|$. So every U-mean double square (v, V) must have type $(1, 1)$ and the equality $v = yy_1 = z^n$ relating the period y, the residue y_1 of (v, V) and v shows that $|y_1| < |z| = |x| < |y|$.

Now we attempt to estimate the replicating capacity of a U-mean square. We use the notation $|\mathrm{gcp}\,(y, \bar{y})|$ to denote the greatest (longest) common prefix of y and $\bar{y}$.

Proposition 6.5. *Suppose that (v, V) is U-mean double square of period y and residue y_1. Then $|y_1| + |\mathrm{gcp}\,(y, \bar{y})| < |x|$. Consequently, $2\eta(v, V) \leq |x|$ for the replicating capacity $\eta(v, V)$ of (v, V).*

Proof. Let (v, V) be of the first kind. Set as before $u = a'd$, $v = ds'$. We know that $v = yy_1$ and $|d| < |x|$ which shows that y_1 is a suffix of s'.

Suppose on the contrary that $|y_1| + |\mathrm{gcp}\,(y, \bar{y})| \geq |x|$ and let e be the prefix of $\mathrm{gcp}\,(y, \bar{y})$ of length $|x| - |y_1|$. Since e is a prefix of $\bar{y} = y_2y_1$ we get that y_1e is also a prefix $y_1y_2 = y$. At the same time e is a prefix of y that means y_1y and y has the same prefix y_1e of length x. Now that y_1y is a factor of $s'v$, that is, of x^*, the Synchronization Remark guarantees that $y_1 = 1$ but this against the definition of the residue. So the former claim is proved. The later claim follows trivially from the Proposition 2.9.

For the case of the second kind, the proof proceeds exactly in the same manner, only take notice that the first occurrence of v by definition ends always in the interval $UU[|u| + |x|, |x^{q+p-1}x_1|]$. This concludes the proof.

Inquiries into the structure of U-mean squares are finished here and we shall further attend to the U-size squares. They have the following remarkable property.

Proposition 6.6. *Let (u, U) be a double square of type (p, q), period x and residue x_1 and let vv be a U-size square. Then $|avv| > |ux^{q+p}x_1|$ and furthermore, the length of v and the larger length of (u, U) are the same, that is, $|v| = |U|$.*

Proof. We set $u = ab$, $av = ux^is''$ that is $v = bx^is''$ for $a, b, s'' \in A^*$ and $i > 0$. The first claim is straightforward. If $|avv| \leq |ux^{q+p}x_1|$ then by the same argument as for U-mean squares, we get that $v = z^n$ for a conjugate z of x. However, as $|b| \leq |x|$ and $|v| \geq |z| = |x|$, it follows from Synchronization Remark that $|v| = |x^lx_1|$ for some integer l that is a contradiction with $|v| = |z^n|$. So $avv = ux^{q+p}x_1s$, or, $v = cs$ for a nonempty word s. We need to show that $|x^is''| = |s|$, or equivalently $|v| = |U|$, by ruling out all other possibilities excerpt this one.

By definition, $|b| \geq |x|, |c| \geq |x|$ and as b and c both are factors of x^* and prefixes of v, we get $b = dx^kx_1$ and $c = dx^lx_1$ (Synchronization Remark) with $x = de$ for some nonnegative integers k, l and a word e.

If $|s| < |x^is'|$ then $|c| > |b|$ that implies $x_2x_1 = x$, as c is a prefix of $bx^is' = v$, $i > 0$.

If, otherwise, $|s| > |x^is''|$ then, on one hand s has the prefix x (note $q > 0$) and on the other hand, $|b| > |c|$. Both facts imply that cx is a prefix of b, again that means $x_2x_1 = x$, a contradiction. So it remains $|s| = |x^is''|$ and the assertion follows.

We now turn to U-size double squares, the main concern of this section. Let (v, V) be a U-size double square of type (p', q'), period y and residue y_1. Interestingly enough, the type of U-size double squares must not be too high, which just comes handy for our handling later. Remind that a U-mean double square is of type one.

Proposition 6.7. *Let (u, U) be a double square of type at least 3. Then every U-size double square is of type no more than 2 and of period greater than that of (u, U).*

Proof. We resume the usual notation for (u, U). Let $v = bs$ and we have $x^{q+p}x_1 = sb$, in virtue of the previous proposition. We know that (u, U) and (v, V) are not coperiod, that is, $|x| \neq |y|$. Since b is a common factor of x^* and y^*, we must have $|b| < |x| + |y|$. Likewise, $|s| < |x| + |y|$. So $2|x| + 2|y| > |b| + |s| = |v| = |x^{q+p}x_1| > 4|x|$, or $|y| > |x|$. Hence $|b| < 2|y|, |s| < 2|y|$, which immediately implies $p' < 4$.

Suppose $p' = 3$. It should be $|b| > |y|$, otherwise yy is a factor of s in spite of $|s| < 2|y|$. Analogously, $|s| > |y|$ that means b and s each contains an occurrence of y, where b the first and s the last, and the middle one overlaps both b and s. By Synchronization Remark, $|yy| =$

$|x^l x_1|$, $l > 0$, then $2|y| > |x^{q+p}x_1| - 2|x| = |x^{q+p-2}x_1|$ meaning that $2|y| \geq |x^{q+p-1}x_1| = |v| - |x|$, or $|v| < 3|y|$, which is obviously a contradiction. Thus $p' < 3$ as desired to prove.

Further we try to estimate the replicating capacity of the U-size double squares, separately for each type. First, type one.

PROPOSITION 6.8. *Let (u, U) be a double square of type at least 3 and (v, V) be a U-size double square of type 1, period y and residue y_1. If $|y_1| < |x|$ then $|y_1| + |\mathrm{gcp}\,(y, \bar{y})| < |x|$ and if $|y_1| \geq |x|$ then $|y_1| = |x^k|$ for some positive integer k and $|y_1| + |\mathrm{gcp}\,(y, \bar{y})| + |y_1| < |v| + |x|$.*

Proof. As before, we denote $u = ab$, $v = bs$ and $x^{q+p}x_1 = sb$; both of b, s are no shorter than $|x|$.

First, suppose $|y_1| < |x|$ but $|y_1| + |\mathrm{gcp}\,(y_1y_2, y_2y_1)| \geq |x|$. Let d be the prefix of y_2y_1 such that $|y_1d| = |x|$. We see then that y_1d is a prefix of y_1y_2 and a factor of x^* and at the same time as the prefix e of length $|x|$ of b is the prefix of y, we get $y_1d = e$. Due to Synchronization Remark, we have $|y| = |x^l x_1|$ for $l > 0$ which shows that $|y_1| = |v| - |y|$ is a (positive) multiple of $|x|$, hence $|y_1| \geq |x|$ despite assumption. Thus $|y_1| + |\mathrm{gcp}\,(y_1y_2, y_2y_1)| < |x|$.

Now consider the case $|y_1| \geq |x|$. We set $s = x^i s''$, $i > 0$, where s'' is a proper prefix of x and $b = t'x^j x_1$, $j \geq 0$, where t' is a proper suffix of x. We have two issues concerning y_1: $|y_1| \geq |s|$ and $|y_1| < |x|$, but we observe that in both cases, by the assumption $|y_1| \geq |x|, |b| \geq |x|$ and $|s| \geq |x|$, the Synchronization Remark yields $|y| = |x^l x_1|$ for some $l > 0$ and $|y_1| = |x^k|$. Now, first issue.

(S1) $|y_1| \geq |s|$. Then y is a prefix of b and s is a suffix of y_1. Moreover, if we write $y_1 = t''x^i s''$, we must see at once that $|t''| < |x|$ (if not y would be a power of a conjugate of x, which is obviously impossible). Next, as y_1 is a prefix of b, a factor of x^* and $i > 0$, we get $t' = t''$, that is $yt'' = t''x^j x_1$, in particular $j = l$. Thus

$$y_1 = (t''s'')^k, x = s''t''$$

while, for some word x'_1,

$$y = (t''s'')^l x'_1, x'_1 t'' = t''x_1.$$

We must have

$$|y_1| + |\mathrm{gcp}\,(y_1y_2, y_2y_1)| < |t''x^j x_1 s''|$$

otherwise $t''x^j x_1 s''$, of length as much as $|y| + |x|$, is a prefix of the longer $y_1y_1y_2$, that is, a prefix of

$$y_1 y = (t''s'')^k (t''s'')^l x'_1,$$

which is a prefix of

$$(t''s'')^k (t''s'')^l x'_1 t'' = (t''s'')^k (t''s'')^l t'' x_1,$$

which is a prefix of

$$(t''s'')^k (t''s'')^l t'' x = (t''s'')^k (t''s'')^l t'' s'' t''$$

that is, a prefix of $(t''s'')^*$. Now note that t'' is a suffix of $t''x^j x_1$ and the synchronization forces $t''x^j x^1 \in (t''s'')^* t''$ that means $|x_1| = 0$ modulo $|x|$. This obvious contradiction proves the inequality. Thus

$$|y_1| + |\mathrm{gcp}\,(y_1y_2, y_2y_1)| + |y_1| < |t''x^j x_1 s''| + |t''x^i s''| = |t''x^j x_1 x^i s''| + |t''s''| = |v| + |x|.$$

Next, the second issue.

(S2) $|y_1| \leq |s|$. Then y_1 is a suffix of s and we have also

$$y_1 = (t''s'')^k, x = s''t''.$$

Hence $y = t'x^jx_1x^ls''$, $l \geq 0$. But the fact that y_1 is a prefix of y and $|b| \geq |x|$ imply that $t'' = t'$ and we have at last

$$y = t''x^jx_1x^ls'', x^ls''y_1 = s.$$

Now, it should be

$$|y_1| + |\text{gcp}\,(y_1y_2, y_2y_1)| < |t''x^jx_1| + |x|$$

because in the opposite case $t''x^jx_1x$ is a prefix of $y_1\text{gcp}\,(y_1y_2, y_2y_1)$, which is a prefix of $y_1y_1y_2 = (t''s'')^kt''x^jx_1x^ls''$. Since $|y_1| \geq |x|$, $|y_1| + |t''x^jx_1| \geq |t''x^jx_1x|$, that is, $t''x^jx_1x$ is a prefix of

$$(t''s'')^kt''x^jx_1 = t''x^{k+j}x_1.$$

By synchronization, we obtain $t''x^jx_1 \in t''x^*$ that imply again $|x_1| = 0$ modulo $|x|$, a contradiction. Finally,

$$|y_1| + |\text{gcp}\,(y_1y_2, y_2y_1)| + |y_1| = |y| - |x^ls''| + |x| + |y_1| = |v| + |x| - |x^ls''| < |v| + |x|$$

that concludes the proof.

We have an immediate corollary, due to Proposition 2.9 for the case of type 1.

Corollary 6.9. *Let (u, U) be a double square and (v, V) be a U-size double square of type 1 under the same hypothesis as in Proposition 6.8. If $|y_1| < |x|$ then $2\eta(v, V) \leq |x|$ and if $|y_1| \geq |x|$ then $3\eta(v, V) \leq |v| + |x|$.*

We investigate a little further the case when the residue of U-size double square is a multiple of the period of the underlying (u, U), the case (S2) in the proof above, which as we see, leads to the worst bound ever. However, in the meantime, these squares cause some favourable (for us) effects, so not everything (bad) is too bad.

Lemma 6.10. *Let (u, U) be a double square and (v, V) be a U-size double square of type 1 under the same hypothesis and notation as in Proposition 6.8 with $u = ab$,$v = bs$, $x^{q+p}x_1 = sb$. If $|y| \leq |b| : b = yt''$ and $y = (t''s'')^lx'_1$ then*

$$aVV = ux^{q+p}x_1x^lx_1sy.$$

If, alternatively, $|y| > |b| : y = bx^ls''$, $b = t''x^jx_1$, then

$$aVV = ux^{q+p}x_1x^{l+1+j}x_1x^{q+p}x_1x^ls''.$$

Note that both sy and x^ls'' are prefixes of $x^{q+p}x_1$. The use of the Lemma 6.10 is in the fact that some factor of $x_2x_1x^{q+p}x_1x$ is repeated. More concretely, if $|y| \leq |b|$, the factor x_2x_1sy occurs at the position $|u| - |x| + 1$ of u, or the same, of aVV but it occurs later as a suffix of aVV. If $|y| > |b$, analogously, the factor $x_2 x_1 x^{q+p} x_1 x^l s''$ is repeated in the same manner.

Proof. It is just a bit of computation.

Case (S1) $|y| \leq |b|$. This is the same case (S1) in the previous proposition, from there we have

$$y_1 = (t''s'')^k, x = s''t''$$

and

$$s = (s''t'')^{k-1}s''$$

and

$$y = (t''s'')^l x_1', x_1't'' = t''x_1.$$

We compute

$$\begin{aligned} aVV &= avvy_2y_1y_1y_2 = avyy_1y_2y_1y_1y_2 = usyyy_1y \\ &= u(s''t'')^{k-1}s''(t''s'')^l x_1'(t''s'')^l x_1'(t''s'')^k(t''s'')^l x_1' \\ &= u(s''t'')^{k-1+l}s''t''x_1(s''t'')^l x_1(s''t'')^{k-1+l}s''x_1' \\ &= ux^{k+l}x_1x^lx_1x^{k-1+l}s''x_1' \end{aligned}$$

as $q + p = k + l$. This equals to

$$ux^{q+p}x_1x^lx_1x^{q+p-1}s''x_1' = ux^{q+p}x_1x^lx_1sy$$

as $s''x_1't'' = s''t''x_1$ and $x^{q+p}x_1 = syt''$. Note that here $l > 0$, as $|y| > |x|$.

Case (S2) $|y| \geq |b|$. Likewise, from Case (S2) of the previous proposition, we have

$$y_1 = (t''s'')^k, x = s''t''$$

but now

$$s = x^l s'' y_1$$

and

$$y = t''x^jx_1x^ls''.$$

We compute

$$\begin{aligned} aVV &= avvy_2y_1y_1y_2 = ux^{q+p}x_1sy_2y_1y_1y_2 = ux^{q+p}x_1x^ls''y_1y_2y_1y_1y_2 \\ &= ux^{q+p}x_1x^ls''yy_1y = ux^{q+p}x_1x^lx^ls''t''x^jx_1x^ls''(t''s'')^kt''x^jx_1x^ls'' \\ &= ux^{q+p}x_1(s''t'')^{l+1+j}x_1x^l(s''t'')^{l+1+k+j}x_1x^ls'' \\ &= ux^{q+p}x_1x^{l+1+j}x_1x^{q+p}x_1x^ls'' \end{aligned}$$

as desired to prove.

In the next proposition we show that, if $p > 2$, the U-size double squares of type 2 are fairly restrictive.

PROPOSITION 6.11 *Let (u, U) be a double square of type $(p, q), p \geq 3$, period x and residue x_1. Then for every U-size double square of type 2, period y and residue y_1 it should be $|y| = |x^l x_1|$, $|y_1| = |x_2|$ and moreover, $|y_1| + |\text{gcp}\,(y, \bar{y})| < |x|$.*

Proof. Remind that always $|y| > |x|$. By the usual notation, we set

$$u = ab, v = bs, x^{q+p}x_1 = sb$$

where $|b| \geq |x|, |s| \geq |x|$. We also write

$$s = t''x^j x_1, b = x^i s'', \quad j \geq 0, i > 0$$

where t'' is a proper prefix of x and s'' proper prefix of x. We handle two possibilities. We should have $j > 0$, otherwise, $|x^i s''| > |x^{q+p-1}| \geq |x^3|$ that readily leads to a contradiction $x = \bar{x}$.

Case 1. $|y| \leq |b|$.

Then y is a prefix of b. Always $|t''x^{j-1}| < |y|$, otherwise y is a nontrivial power of a conjugate of x, hence not primitive. Further, we divide the treatment into two subcases.

Subcase 1.1. $|t''x^{j-1}| < |y| \leq |t''x^j|$. We write $y = t''x^{j-1}s'$, $x = s't'$ for some words t', s'. We must have $|t'x_1| < |x|$, if not, by Synchronization Remark, y, with $|y| > |x|$, is not primitive then.

If $j = 1$, from $|t''| + |x| + |x^i| + |s''| = |x^{q+p}|$ and $|t''| + |s''| < 2|x|$ it follows that $(i + j) \geq q + p - 1 \geq 3$. Together with $yy_1 = t'x_1x^is''$ being a prefix of $v = t''xx_1s$, that implies $t'' = t'x_1$ which in turn implies that x_1 is a suffix of x and also x is a prefix of x_1x. Both these facts mean that $x_1x = xx_1$ showing x is not primitive which is impossible.

Thus $j > 1$. Analogously, noting that $i > 0$, we also have $t'x_1 = t''$. Therefore $|y| = |x^jx_1|$ and $|yy_1| = |t'x_1x^is''| = |x^{i+1}|$ that means $i + 1 > j$. In case $i > j$, take into account that x_1 is a suffix of x, we will obtain the same contradiction, as above. So, it remains $i = j$ and $|y_1| = |x_2|$.

Subcase 1.2. $|t''x^j| < |y| \leq |t''x^jx_1|$. Remind that $j > 0$. We write then $y = t''x^js'$, $s't' = x_1$, $|s'| > 0$ for some word t'. Because $|t'| < |x_1|$ we have $|t'x| < |t''x^jx_1|$ hence $t' = t''$. Therefore $|y| = |x^jx_1|$, hence $|yy_1| = 0$, $|y_1| = |x_2|$ both modulo $|x|$. As $t' = t''$ we see that yy_1 is a factor of x^* that indeed shows that $|y_1| < |x|$ by Periodicity Remark. Consequently, $|y_1| = |x_2|$.

Case 2. $|y| \geq |t''x^jx_1|$.

Now yy_1 is a suffix of x^is'', which implies immediately that $|y_1| < |x|$ for y to be primitive. Next, taken into account that b is a prefix of y and, as always in this situation by synchronization, $|y| = |x_1|$ modulo $|x|$. Therefore $|yy_1| = 0$ modulo $|x|$ and, again $|y_1| = |x_2|$ modulo $|x|$, so $|y_1| = |x_2|$.

Thus, in all instances, we have $|y_1| = |x_2| < |x|$. Finally, to prove that

$$|y_1| + |\text{gcp}\,(y, \bar{y})| < |x|$$

we just assume the contrary and seek a contradiction. Let d be a prefix of $\text{gcp}\,(y, \bar{y})$ of length $|x| - |y_1|$, then y_1d is a prefix of length $|x|$ of $y_1y_2y_1$, that is, of y_1y_2. Now that $|s| \geq |x|$,

$|b| \geq |x|$ we get $|yy_1| = |x_1|$ modulo $|x|$, which is in an obvious contradiction with $|y| = |x_1|$ modulo $|x|$. This completes the proof.

We have an immediate corollary, due to Proposition 2.9, for the replicating capacity of the U-size double square of type 2, which is less than its counterpart of type 1.

COROLLARY 6.12. *Let (u, U) be a double square and (v, V) be a U-size double square of type 2 under the same hypothesis as in Proposition 6.11. Then $\eta(v, V) \leq |x_1|$ and, by the way, $q + p$ is odd.*

Proof. Since $|y| = |x^l x_1|$, by $|v| = |U|$ we get $q + p = 2l + 1$. The other statement is also straightforward by definition.

After a long journey we are now able to attend to the proof itself in the next last section.

7 Proof of the Theorem 2.10

Finally we make use of the tools forged in the previous sections - no more of tiresome assertions - to properly prove the announced result, however, with the help of some little auxiliary details.

LEMMA 7.1. *Suppose that (u, U) is a double square of type 1 or 2 and (v, V) is a double square, not isomorphic to (u, U) and starting inside u with $u = ab, v = bs$. Then*

$$2\gamma(a) \leq |s|$$

where $\gamma(a)$ is the number of double squares isomorphic to (u, U) (including it) and preceding (v, V).

Proof. Let (p, q), x and x_1 be the type, the period and the residue of (u, U), respectively. Suppose that (u', U') is the latest replica of (u, U) that precedes (v, V) and we write $us' = a'u'$, for some words s', a' with $a' = s'$ and $|s'| = |a'| \leq \eta(u, U) - 1 \leq |x_1| - 1$ if $p = 1$ and $|s'| = |a| \leq |x| - 1$ if $p = 2$. It is known that (u, U) has no proper (different from itself) inner double squares (the passage before Proposition 3.4) and every outer double square is isomorphic to it, Theorem 5.1 and Theorem 5.2. So (v, V) is either a non-coperiod or a marginal coperiod double square with (u, U). If (v, V) is marginal coperiod then surely $|s| > |x|$; if (v, V) is non-coperiod then $|s| > |x_1|$, if $p = 1$ and $|s| > |x|$, if $p = 2$. Hence in any case, u' cannot end outside v and we can write $s = s's''$ for some s''. Note also that v in its turn, cannot end inside u', so $|s''| > 0$. Now relative to (u', U'), like to (u, U), as (v, V) is not isomorphic to (u, U) we have $|s''| > |x_1|$ if $p = 1$ and $|s''| > |x|$ if $p = 2$. Consequently

$$2\gamma(a) \leq 2(|a'| + 1) \leq |s'| + 1 + \eta(u, U) \leq |s'| + |x_1| + 1 \leq |s'| + |s''| = |s|$$

if $p = 1$, and

$$2\gamma(a) \leq 2(|a'| + 1) \leq |s'| + 1 + \eta(u, U) \leq |s'| + |x| + 1 \leq |s'| + |s''| = |s|$$

if $p = 2$ that is desired to prove.

In the subsequent lemma we estimate the number of U-size double squares starting inside a given interval. We will use this lemma, rather than the Lemma 7.1 directly, in the future computation.

LEMMA 7.2. *Let (u, U) be a double square of type at least 3. Suppose that (e, f, d) is a particular occurrence of a factor f of W and all U-size double squares that start inside f end outside f. Let (v_h, V_h), starting at the position h, be the latest U-size double square starting inside f such that it is not isomorphic the one immediately preceding it and (v_H, V_H), starting inside f at the position H, is the last U-size double square of all. Then*

$$H - h + 1 \leq \eta(v_h, V_h)$$

and moreover,

$$2\gamma(f) \leq |f| + (H - h + 1)$$

where $\gamma(f)$ is the number of U-size double squares starting inside f.

Proof. We denote by $i_1, i_2, \ldots, i_t$ the positions, at which the U-size double squares (u_1, U_1), (u_2, U_2), $\ldots$, (u_t, U_t)=(v_h, V_h) start, such that all double squares following (u_j, U_j) and preceding (u_{j+1}, U_{j+1}) are isomorphic to (u_j, U_j), but not to (u_{j+1}, U_{j+1}), $j = 1, 2, \ldots, t-1$, the number of which we denote by γ_j. Since all u_j are of the same length, by Lemma 7.1, we obtain,

$$2\gamma_j \leq i_{j+1} - i_j$$

$j = 1, 2, ..., t-1$. Now that there is at most $H - h + 1$ (U-size) double squares starting inside $f[h, H]$, all of them are replicas of (v_h, V_h) by assumption, hence the number of them does not exceed both $\eta(v_h, V_h)$ and $H - h + 1$, we obtain

$$2\gamma(f) = 2(\gamma_1 + \cdots + \gamma_{t-1}) + 2(H - h + 1) \leq i_{t-1} - i_1 + |f| - i_t + 1 + H - h + 1$$

$$= |f| - i_1 + 1 + H - h + 1 \leq |f| + H - h + 1.$$

Lemma is proved.

Remark. Although the Lemma 7.2 is formulated only for the U-size double squares, it is evident that the proof is also valid for the U-mean double squares of the first kind as well as of the second kind. As matter of fact, it holds for an arbitrary class of double squares of the same smaller length of type 1 or 2.

We prove the theorem by induction on the length of $|W|$. It is trivial to verify the induction basic for a few small values of $|W|$, say $|W| = 1, 2, 3, 4, 5$. Now suppose that $|W|$ is large enough. If W contains no double squares, we are done, so let (u, U) be the first double square of W (of type (p, q), period x, residue x_1). We can assume, for neat notation, that (u, U) starts at the first position of W, otherwise we are done either, by induction. That is $W = uw$ for some suffix w of W. Now for a double square (v, V) starting inside u, $u = ab$, $|a| > 0$, $v = bs$ we must have $2\gamma(a) > |s|$ for the number $\gamma(a)$ of double squares following (u, U) but strictly preceding (v, V). Indeed, if $2\gamma(a) \leq |s|$ then apply the induction hypothesis to bw with $|bw| < |W|$ with its number of double squares

$$2\gamma(bw) < -|v| + C|bw|$$

so that

$$2\gamma(a) + 2\gamma(bw) < |s| - |v| + C|bw| = |a| + |v| - |u| - |v| + C|bw|$$
$$= -|u| + |a| + C|bw| = -|u| + C|abw| = -|u| + C|W|$$

and we are finished.

For convenience we term, among the double squares starting inside u, the inner and outer coperiod, the U-mean and U-size ones, and U itself *internal* double squares and the rest *marginal* double squares. Equivalently, the double square (v, V) is marginal if and only if $u = ab$, $v = bs$ then $|s| > |x^{q+p-1}x_1|$. By the way, remind that we have defined the marginal coperiod double squares previously. Our main concern is to bound above the total number of internal double squares starting inside u then apply the induction to the suffix of W starting at the position of first marginal double square. We consider in turn when (u, U) is of type 1 or 2 and when at least 3 combined with the possibilities when there is or there is not a marginal double squares starting inside u.

7.1 Type Three or Greater

As usual denote the type of (u, U) by (p, q), period and residue by $|x|$ and $|x_1|$, respectively. First, we estimate the total number γ_p of the inner and outer double squares of (u, U) plus (u, U) itself, that is $\gamma_p + 1$.

Situation S. The last coperiod double square starting inside u is the outer one (v, V) of type (p', q') or the (u, U) itself.

We put

$$u = ab, v = bs$$

where $s = 1$ iff $v = u$. Evidently,

$$2(\gamma_p + 1) \le 2(|u| - |b| + 1) = 2|u| - 2|a| + 2 = 2|u| - 2(|v| - |s|) + 2$$
$$= (|u| + |x^q|) + (|u| - |x^q|) - ((|v| + |x^{q'}|) + (|v| - |x^{q'}|)) + 2 + 2|s|$$
$$= |x^{p-p'-q+q'}| = |x^{2(p-p')}| + 2 + 2|s|$$

as $p + q = p' + q'$ by Proposition 3.7.

Situation T. The last coperiod double square starting inside u is the inner one (v, V) of type (p', q').

Then

$$u = avt, b = vt$$

for some word t, $|t| \ge 0$. By similar computation as above

$$2\gamma_p + 2 \le |x^{2(p-p')}| + 2 - 2|t|.$$

For the U-mean and U-size double squares, we observe that they should follow all coperiod squares because, first, they end outside anyone of them, by definition and, second, they are not coperiod with them.

We further estimate the number of U-size double squares separately. As said, every U-size double square v' must start inside b, more formally, in the interval $u[|a|+1,|u|]$ of u.

Assume that we are in Situation S for the coperiod double squares, so we can write $b = ge$ and $v' = eh$. In order to avoid having a later occurrence of vv, we must have $|h| < |sv|$, or

$$|e| > |v'| - |s| - |v| = |x^{q+p}x_1| - |s| - |x^{p'}x_1| = |x^{q'}| - |s|.$$

Thus

$$|g| = |b| - |e| < (|v| - |s|) - (|x^{q'}| - |s|) = |v| - |x^{q'}| = |x^{p'-q'}x_1|.$$

Now let $\bar{g}$ be the prefix of length $|x^{p'-q'}x_1|$ of b then all of the U-size double squares must start inside $\bar{g}$ but, note that, not at the first position of $\bar{g}$, which is occupied by v. Consequently, by Lemma 7.2. above, when $\bar{g}$ plays the part of f, for the number γ_s of U-size double squares, we have

$$2\gamma_s \le |\bar{g}| - 1 + (H - h + 1) \le |x^{p'-q'}x_1| - 1 + (H - h + 1) \le |x^{p'-q'}x_1| - 1 + \eta(v_h, V_h)$$

where H, h and (v_h, V_h) have the meaning indicated in the lemma.

Now if we are in Situation T for the coperiod double squares, the computation goes the very same way, only with

$$|e| > |v'| - |v| + |t|,$$

$$|g| = |v| + |t| - |e| < |x^{p'-q'}x_1|$$

which is just the same bound for g leading to the same bound for γ_s.

Summing up, we get for the number of coperiod, the unique "special coperiod" (u, U) and U-size double squares of (u, U) the inequality

$$2\gamma_p + 2 + 2\gamma_s \le |x^{2(p-p')}| + 2 + |x^{2(p'-q')}x_1| - 1 + H - h + 1$$

$$= |x^p x_1| - |x^q| + 1 + (H - h + 1) + 2|s| \le |x^p x_1| - |x^q| + 1 + \eta(v_h, V_h) + 2|s|$$

if we are in Situation S and

$$2\gamma_p + 2 + 2\gamma_s \le |x^{2(p-p')}| + 2 + |x^{2(p'-q')}x_1| - 1 + (H - h + 1)$$

$$\le |x^{2(p-p')}| + 2 + |x^{2(p'-q')}x_1| - 1 + (H - h + 1) = |x^p x_1| - |x^q| + 1 + (H - h + 1) - 2|t|$$

$$\le |x^p x_1| - |x^q| + 1 + \eta(v_h, V_h)| - 2|t|$$

if we are in Situation T.

Finally, for the number of U-mean double squares, γ_m, we have by Proposition 6.4 that

$$\gamma_m + |s| < |x|.$$

Thus for the total number of the coperiod, including (u, U), the U-size and the U-mean double squares $\gamma_p + 1 + \gamma_s + \gamma_m$ we get

$$2\gamma_p + 2 + 2\gamma_s + 2\gamma_m \le |x^p x_1| - |x^q| + 1 + (H - h + 1) + 2|s| + 2\gamma_m$$

$$\leq |x^p x_1| - |x^q| - 1 + (H - h + 1) + 2|x|$$

$$\leq |x^p x_1| - |x^q| - 1 + \eta(v_h, V_h)| + 2|x|$$

if we are Situation S, and, for $\gamma_m < |x|$ we get

$$2\gamma_p + 2 + 2\gamma_s + 2\gamma_m \leq |x^p x_1| - |x^q| + 1 + (H - h + 1) - 2|t| + 2\gamma_m$$

$$\leq |x^p x_1| - |x^q| - 1 + (H - h + 1) - 2|t| + 2|x|$$

$$\leq |x^p x_1| - |x^q| - 1 + \eta(v_h, V_h) - 2|t| + 2|x|$$

if we are in Situation T.

In both Situation S and T, keep in mind that $H - h + 1 \leq \eta(v_h, V_h)$ and, at least, $\gamma_m < |x|$ and even $\gamma_m + |s| < |x|$ if we are in Situation S. We proceed further by dividing into subsection for easier visualization.

7.1.1 No Marginal Double Squares

We suppose now that u has no marginal double squares that means all the double ones that starting inside u are coperiod, U-mean, U-size double squares of (u, U) and (u, U) itself.

We say that a given double square is an *external* double square of (u, U) if it follows (u, U) but starts outside u. Let (u', U') be the first external double square that follows (u, U). We see then that u' is separated from u by a prefix, say, w_1 of w, formally, we write $w = w_1 u' w_2$. We denote specially the length of w_1 by $||u, u'||$.

We continue the proof under the effect of the Situation T or S. First,

Situation S. Put $\sigma = 2\gamma_p + 2 + 2\gamma_s + \gamma_m$, for which we have the bound

$$\sigma \leq |x^p x_1| - |x^q| + (H - h + 1) + 1 + 2\gamma_m + 2|s| \leq |x^p x_1| - |x^q| + (H - h + 1) + 2|x| - 1.$$

The meaning of h, H, (v_h, V_h), (v_H, V_H) is indicated in Lemma 7.2 now for the prefix $\bar{g}$ of b.

Situation S1. (v_h, V_h) has type 1, period y, residue y_1 and $|y_1| = 0$ modulo $|x|$ and $|y| = |x_1|$ modulo $|x|$; (v_H, V_H), which is isomorphic to (v_h, V_h), has also type 1, period y', $|y'| = |y|$, residue y'_1, $|y'_1| = |y_1|$.

We write

$$u = ab, v = bs$$

and

$$b = a_h b_h, \quad |a_h| = h - 1$$

$$v_h = b_h s_h$$

for some word s_h. Similarly,

$$b = a_H b_H, \quad |a_H| = H - 1$$

$$v_H = b_h s_H$$

for some word s_H.

(1) First alternative: $|y| \leq |b_h|$. We put then $b_h = yt'$. Note that $|t'| < |x|$ (Proposition 6.8, (S1)). Put further $a_H = a_h d$ for a word d, $|d| = H - h$. We have two possibilities.

(1.1) $H - h \leq |t'|$. Then $|b_h| = |y| + |t'| \geq |d| + |y'|$. Hence dy' is a prefix of b_h and $b_h = d'e$ for some word e. By Lemma 6.10, Case (S1) with therein $(v, V) = (v_H, V_H)$, the prefix of $x^{q+p}x_1$ of length $|x^{q+p}x_1| - |e|$ repeated (not necessarily adjacent), so for the first external square (u', U') of (u, U) we must have (if there is no external double squares, we put $||u, u'|| = |x^{q+p}x_1|$ and $|u'| = 0$, no worse)

$$||u, u'|| + |u'| > |x^{q+p}x_1| - |e|$$

for $u'u'$ to avoid occurring again, which implies

$$||u, u'|| + |u'| > \frac{1}{2}(|x^{q+p}x_1| - |e|).$$

Now, we set

$$\sigma - ||u, u'|| - |u'| = -|u| + r(u)$$

and estimate

$$\begin{aligned} r(u) &= |u| + \sigma - ||u, u'|| - |u'| \\ &\leq |x^p x_1| + |x^p x_1| - |x^q| + (H - h + 1) + 2|x| - 1 - \frac{1}{2}(|x^{q+p}x_1| - |e|) \\ &= \frac{3}{2}|x^p x_1| - \frac{3}{2}|x^q| + \frac{1}{2}|e| + H - h + 2|x| \\ &< \frac{3}{2}|x^p x_1| + \frac{1}{2}|x| - \frac{3}{2}|x^{q-1}| + |x| = \frac{3}{2}|x^p x_1| + \frac{3}{2}|x| - \frac{3}{2}|x^{q-1}| \end{aligned}$$

as $|x| > |d| + |e| = |t'| = H - h + |e| \geq H - h + \frac{1}{2}|e|$.

If $p > 3$ then $\frac{3}{2}|x| < \frac{3}{8}|x^p x_1|$ and, if $p = 3$, $q > 1$ then $\frac{3}{2}|x| - \frac{3}{2}|x^{q-1}| \leq 0$ so that

$$r(u) < \frac{15}{8}|u|.$$

and

$$r(u) < \frac{3}{2}|u|,$$

respectively.

If, else $p = 3, q = 1$, the argument is more delicate. We use instead the inequality

$$\sigma \leq |x^p x_1| - |x^q| + (H - h + 1) + 1 + 2\gamma_m + 2|s|.$$

As $q + p = 3 + 1 = 4$ is even, there is no U-mean double squares of second kind. For the U-mean double squares of first kind, if we write

$$ux^{q+p}x_1 = uge$$

for some word g, then they should end inside the current occurrence of e, because, by Lemma 6.10, Case (S1), with respect to (v_H, V_H) the word $\bar{x}g$ occurs again, besides the occurrence as factor of $ux^{q+p}x_1$, which is a prefix of UU. This implies that the number of U-mean double squares is less than $|e|$, i.e., $\gamma_m < |e|$. Now that t' is a common suffix of x and $\bar{x}$ (Proposition 6.8, (S1)) and s is a common prefix of x and $\bar{x}$ (Proposition 3.8), we obtain

$$H - h + |\gamma_m| + |s| < H - h + |e| + |s| = |t'| + |s| < |x|$$

or

$$1 + H - h + |\gamma_m| + |s| \leq |x|.$$

Now

$$\begin{aligned} r(u) &= |u| + \sigma - ||u, u'|| - |u'| \\ &\leq 2|x^p x_1| - |x^q| + 1 + (H - h + 1) + 2\gamma_m + 2|s| - \frac{1}{2}(|x^{q+p}x_1| - |e|) \\ &= \frac{3}{2}|x^p x_1| - \frac{3}{2}|x^q| + 1 + (H - h + 1) + \gamma_m + |s| + \frac{1}{2}|e| + \gamma_m + |s| \\ &< \frac{3}{2}|x^p x_1| - \frac{3}{2}|x| + |x| + \frac{1}{2}|e| + |x| \\ &= \frac{3}{2}|x^p x_1| + \frac{1}{2}|x| + \frac{1}{2}|x| = \frac{3}{2}|x^p x_1| + |x| \\ &< \frac{3}{2}|x^p x_1| + \frac{1}{3}|x^p x_1| = \frac{11}{6}|u| \end{aligned}$$

as $|x| < \frac{1}{3}|x^p x_1|$.

(1.2) $|t'| < H - h$. Then $|dy'| = H - h + |y'| > |t'| + |y| = |b_h|$, so we put $dy' = b_h e$ for some word e. Now for the first external double square (u', U') of (u, U) we must have

$$||u, u'|| + |u'u'| > |x^{q+p}x_1| + |e|.$$

(In case (u, U) has no external double squares, we set $||u, u'|| = |x^{q+p}x_1| + |e|$ and $|u'| = 0$, no harm) Hence

$$\begin{aligned} ||u, u'|| + |u'| &> \frac{1}{2}(|x^{q+p}x_1| + |e|) = \frac{1}{2}|x^{q+p}x_1| + \frac{1}{2}((H - h) - |t'|) \\ &\geq \frac{1}{2}|x^{q+p}x_1| + \frac{1}{2}(H - h) - \frac{1}{2}|x|. \end{aligned}$$

Now for $r(u)$

$$\begin{aligned} r(u) &= |u| + \sigma - ||u, u, || - |u'| \\ &\leq 2|x^p x_1| - |x^q| + 1 + (H - h + 1) + 2|x| - \frac{3}{2}|x^p x_1| - \frac{1}{2}|x^q| - \frac{1}{2}(H - h + 1) + \frac{1}{2}|x| \\ &\leq \frac{3}{2}|x^p x_1| - \frac{3}{2}|x^q| + \frac{1}{2}(H - h + 1) + \frac{5}{2}|x| \\ &\leq \frac{3}{2}|x^p x_1| - \frac{3}{2}|x^{q-1}| + |x| + \frac{1}{2}\eta(v_h, V_h) \\ &\leq \frac{3}{2}|x^p x_1| - \frac{3}{2}|x^{q-1}| + |x| + \frac{1}{6}(|x^{q+p}x_1| + |x|) = \frac{5}{3}|x^p x_1| - \frac{3}{2}|x^{q-1}| + \frac{7}{6}|x| \end{aligned}$$

by Proposition 6.9.

If $p > 3$, as $-\frac{3}{2}|x^{q-1}| + \frac{7}{6}|x| \leq \frac{7}{6}|x|$ and $\frac{7}{6}|x| < \frac{7}{24}|x^p x_1|$ we have

$$r(u) < \frac{5}{3}|x^p x_1| + \frac{7}{24}|x^p x_1| = \frac{47}{24}|u|.$$

If $p = 3$ and $q > 1$ then $-\frac{3}{2}|x^{q-1}| + \frac{7}{6}|x| \leq 0$ and

$$r(u) < \frac{5}{3}|u|.$$

If $p = 3$, $q = 1$, we handle analogously as it was done in (1.1) for this case. As $q + p$ is then even there are no U-mean double squares of second kind, and even more, there are no ones of the first kind either, as $\bar{x}x^{q+p}x_1 e$ occurs later than the previous evident occurrence (Lemma 6.10, (S2)), that is $\gamma_m = 0$. Now we use the inequality

$$\sigma \leq |x^p x_1| - |x^q| + 1 + (H - h + 1) + 2\gamma_m + 2|s|.$$

to estimate

$$r(u) = |u| + \sigma - ||u, u'|| - |u'| = 2|x^p x_1| - |x^q| + 1 + (H - h + 1) + 2|s| - (||u, u'|| + |u'|)$$

$$\leq 2|x^p x_1| - |x^q| + 1 + (H - h + 1) + 2|s| - \frac{1}{2}(|x^{q+p}x_1| + H - h - |t'|) + 2|s|.$$

Note that $|t'| + |s| < |x|$, so $\frac{1}{2}|t'| + 2|s| \leq 2|x| - 2$. Consequently,

$$r(u) \leq \frac{3}{2}|x^p x_1| - \frac{3}{2}|x| + \frac{1}{2}\eta(v_h, V_h) + 2|x|$$

$$\leq \frac{3}{2}|x^p x_1| - \frac{3}{2}|x| + \frac{1}{6}(|x^{q+p}x_1| + |x|) + \frac{1}{2}|x|$$

$$\leq \frac{5}{3}|x^p x_1| + \frac{1}{6}|x| + \frac{1}{6}|x| + \frac{1}{2}|x| = \frac{5}{3}|x^p x_1| + \frac{5}{6}|x|$$

$$< \frac{5}{3}|x^p x_1| + \frac{5}{18}|x^p x_1| = \frac{35}{18}|u|$$

and First Alternative is finished.

2. Second Alternative $|y| > |b_h|$. We write

$$y = b_h s', y' = b_H s'', \quad |s'| > 0,$$

so that

$$s'' = s'e, \quad |e| = |a_{H-h}| = H - h.$$

For the first external double square (u', U') of (u, U), likewise, we have

$$||u, u'|| + |u'| > \frac{1}{2}(|x^{q+p}x_1| + |s''|);$$

if (u, U) has no external squares, as before, we set $||u, u'|| = |x^{q+p}x_1| + |s''|$ and $|u'| = 0$ so the inequality is always satisfied. Now, by a customary computation

$$r(u) = |u| + \sigma - ||u, u'|| - |u'| = 2|x^p x_1| - |x^q| + 2|x| - 1 + (H - h + 1) - \frac{1}{2}(|x^{q+p}x_1| + |s'e|).$$

For $|s'e| \geq H - h + 1$, we have

$$\begin{aligned} r(u) &< \frac{3}{2}|x^p x_1| - \frac{3}{2}|x^q| + \frac{1}{2}(H - h + 1) + 2|x| \\ &\leq \frac{3}{2}|x^p x_1| - \frac{3}{2}|x^q| + \frac{1}{6}(|x^{q+p}x_1| + |x|) + 2|x| \\ &= \frac{3}{2}|x^p x_1| - \frac{4}{3}|x^q| + \frac{13}{6}|x| = \frac{5}{3}|x^p x_1| - \frac{4}{3}|x^{q-1}| + \frac{5}{6}|x| \\ &\leq \frac{5}{3}|x^p x_1| + \frac{5}{6}|x| \leq \frac{5}{3}|x^p x_1| + \frac{5}{18}|x^p x_1| = \frac{35}{18}|x^p x_1| = \frac{35}{18}|u| \end{aligned}$$

as $p \geq 3$.

Situation S2. (v_h, V_h) has type 1, period y, residue y_1 and $|y_1| < |x|$; (v_H, V_H), which is isomorphic to (v_h, V_h) has also type 1, period y', $|y'| = |y|$, residue y'_1, $|y'_1| = |y_1|$.

In this issue, by Corollary 6.9, $\eta(v_h, V_h) \leq \frac{1}{2}|x|$ and the estimation goes as usual, we use the inequality

$$\sigma \leq 2|x^p x_1| - |x^q| + \eta(v_h, V_h) + 2|x| - 1$$

and

$$||u, u'|| + |u'| > \frac{1}{2}|x^{q+p-1}x_1|$$

or a more refined

$$||u, u'|| + |u'| \geq |xx|$$

when $q + p = 4$ (prove it!), to bound above

$$r(u) = |u| + \sigma - ||u, u, || - |u'|.$$

A routine computation yields

$$r(u) \leq \frac{3}{2}|x^p x_1| - |x^q| + \frac{1}{2}|x| + 2|x| - 1 < \frac{3}{2}|x^p x_1| - |x^{q-1}| + \frac{3}{2}|x|.$$

If $p > 3$ then

$$r(u) < \frac{3}{2}|x^p x_1| + \frac{3}{2}|x| \leq \frac{3}{2}|x^p x_1| + \frac{3}{8}|x^4 x_1| \leq \frac{15}{8}|u|.$$

If $p = 3$ and $q > 1$ then

$$r(u) < \frac{3}{2}|x^p x_1| + \frac{1}{2}|x| \leq \frac{3}{2}|x^p x_1| + \frac{1}{6}|x^3 x_1| \leq \frac{5}{3}|u|.$$

If, finally, $p = 3$, $q = 1$ we use the more refined inequality above

$$r(u) \leq 2|x^p x_1| - |x^q| + \frac{5}{2}|x| - 2|x|$$

$$< 2|x^p x_1| - \frac{1}{2}|x| = |x^p x_1| + \frac{5}{2}|x| + |x_1| < |x^p x_1| + \frac{3.5}{4}|x^p x_1| = \frac{15}{8}|u|.$$

Situation S2. (v_h, V_h) has type 2, period y, residue y_1: $v = yyy_1$.

By Corollary 6.12, we have, first, that $q+p$ is odd, whence $q+p \geq 5$ and, second, $\eta(v_h, V_h) \leq |x_1|$. Now we use

$$\sigma \leq 2|x^p x_1| - |x^q| + \eta(v_h, V_h) + 2|x| - 1$$

and

$$||u, u'|| + |u'| > \frac{1}{2}|x^{q+p-1}x_1|$$

to estimate

$$r(u) = 2|x^p x_1| - |x^q| + \eta(v_h, V_h) + 2|x| - 1 - (||u, u'|| + |u'|)$$

$$< \frac{3}{2}|x^p x_1| - |x^{q-1}| + |x| + |x_1| \leq \frac{3}{2}|x^p x_1| + |x| + |x_1|.$$

If $p > 3$ then $\frac{|x|+|x_1|}{|x^4 x_1|} < \frac{2}{5}$, whence

$$r(u) < \frac{3}{2}|x^p x_1| + \frac{2}{5}|x^p x_1| = \frac{19}{10}|u|.$$

If $p = 3$, then $q > 1$ and note that $|x_1| < \frac{1}{4}|x^3 x_1|$, we obtain

$$r(u) < \frac{3}{2}|x^p x_1| - |x| + |x| + |x_1| = \frac{3}{2}|x^p x_1| + |x_1|$$

$$< \frac{3}{2}|x^p x_1| + \frac{1}{4}|x^p x_1| = \frac{7}{4}|u|.$$

Thus the Situation S2 is done, and with it the whole Situation S. We finish with the subsection 7.1.1 by handling

Situation T. It is straightforward. We use the bound

$$\sigma \leq |x^p x_1| - |x^q| + 1 + (H - h + 1) - 2|t| + 2\gamma_m$$

$$\leq |x^p x_1| - |x^q| - 1 + (H - h + 1) - 2|t| + 2|x|$$

$$\leq |x^p x_1| - |x^q| - 1 + \eta(v_h, V_h) - 2|t| + 2|x|$$

as $\gamma_m < |x|$.

The estimation proceeds just as in Situation S with the analogous cases and subcases, which correspondingly lead to the same bounds without any complications or too much care, taken into account the only fact that whenever in the Situation S

$$\gamma + |s| < |x|$$

is used for a quantity γ, the inequality

$$\gamma < |x|$$

is valid regardless of which Situation, S or T, is under way. It is guaranteed by Propositions 3.9 and 6.2, or is seen from the proof of Propositions 3.2 and 6.8.

Now we are able to complete the induction. Remind that

$$W = uw = uw_1u'w_2$$

where (u', U') is the first external double square of (u, U). If there is no such one, we take on the previous usual conventions. We have

$$2\gamma(W) = \sigma + 2\gamma(u'w_2)$$

which is, by induction

$$< \sigma - |u'| + C|u'w_2| < \sigma - |u'| - |w_1| + |w_1| + C|u'w_2|$$

$$= -|u| + |u| + \sigma - |u'| - ||u, u'|| - C|w_1u'w_2| = -|u| + r(u) + C|w_1u'w_2|.$$

Due to the estimates above for which $r(u) < C|u|$, we finally get

$$2\gamma(W) < -|u| + C|u| + C|w_1u'w_2| = -|u| + C|W|$$

what is desired to verify. Further, the next alternative

7.1.2 With Marginal Double Squares

Now we assume that (u, U) has marginal double squares and that (u', U') is the first of them. Let for the words a, b and s

$$u = ab, u' = bs$$

with $|s| > |x^{q+p-1}x_1|$. Then obviously all the double squares that follow (u, U) and strictly precede (u', U') are the internal ones, the number of which we denote by $\gamma(a)$. Let further $\bar{a}$ be the proper prefix of a such that the latest internal double square (v, V), not U-*mean* one, that precedes (u', U') starts at the position $|\bar{a}| + 1$ of u. Clearly, $\gamma(a) \le |\bar{a}| + 1 \le |a|$. Our aim here is to prove the bound

$$r(a) = |u| + 2\gamma_(a) - |u'| < C|a|$$

which ensures that it is possible to keep the induction going.

Case 1. $|u'| \le |b| + |x^{q+p}x_1|$.

We write $u' = bs$, $x^{q+p}x_1 = st$, hence $|u'| = |x^{q+p}x_1| + |b| + |t|$. Since $q + p \ge 4$, $q > 0$, $UU = ux^{q+p}x_1x^q$ we should have $|t| < |b|$. Indeed, $|t| > |b|$ implies that $t = bx^l$ for a positive integer l, so $|t| > |x|$, against the marginality of $u'u'$; if $|t| = |b|$ then $au'u' = ux^{q+p}x_1g$, but $x^{q+p}x_1g$ contains an occurrence of uu because $|g| > |x^{q+p-1}x_1| > |x^px_1| = |u|$. Now

$$r(a) = |x^px_1| + 2\gamma(a) - |x^{q+p-1}x_1| - (|b| - |t|) = 2\gamma(a) - |x^q| - (|b| - |t|).$$

Let set $T = V$ if (v, V) is an inner or outer double square of (u, U) and $T = v$ if (v, V) is a U-size double square. Then, surely, $|T| = |U| = |x^{q+p}x_1|$ and

$$\bar{a}TT = ux^{q+p}x_1\bar{a}.$$

Because $\bar{a}$ and s are prefixes of x^*, we should have

$$|t| + |\bar{a}| < |b| + |x|$$

otherwise, due to the synchronization on x, W has the prefix $ux^{q+p}x_1x^ls$, hence the prefix $ux^{q+p}x_1x^px_1$ as $|s| > |x^{q+p-1}x_1|$ which shows that W has another occurrence of uu later, a contradiction. Thus

$$|b| - |t| > |\bar{a}| - |x|.$$

1. Suppose that $|b| \geq |x|$. Then there is no U-mean square preceding (u', U') and we obtain

$$r(a) = 2\gamma(a) - |x^q| - (|b| - |t|) \leq 2(|\bar{a}| + 1) - |x^q| - (|\bar{a}| - |x| + 1)$$
$$\leq |\bar{a}| + 1 - |x^{q-1}| \leq |\bar{a}| + 1 \leq |a|.$$

2. Now suppose that $|b| < |x|$, or $|a| > |x^{p+q-1}x_1|$. Then there might be U-mean double squares preceding (u', U'). Then, proceeding as above, only taking into account the number γ_m, which is less than $|x|$, of U-mean of double squares

$$r(a) \leq |\bar{a}| + 1 - |x^{q-1}| + 2\gamma_m \leq |a| + 1 \leq |a| + 2|x| - 2.$$

If $p > 3$ then $|a| > 3|x| + |x_1|$

$$r(a) < |a| + \frac{2|x|}{|x^3x_1|}|a| \leq \frac{5}{3}|a|.$$

If $p = 3$ and $q > 1$ then

$$r(a) \leq |a| - |x^{q-1}| + 2|x| - 2 \leq |a| + |x| < |a| + \frac{1}{3}|a| = \frac{4}{3}|a|.$$

Finally, if $p = 3$ and $q = 1$ then $|a| > 2|x|$ and (u, U) has no U-mean double squares of the second kind. Since all the U-mean double squares start inside a certain suffix of u of length less than x and have the replicating capacity no more than $\frac{1}{2}|x|$ and have the same length if they are of the same kind, according to Lemma 7.2 and the subsequent remark we obtain

$$2\gamma_m \leq |x| + \frac{1}{2}|x| = \frac{3}{2}|x|.$$

So

$$r(a) \leq |\bar{a}| + 1 + 2\gamma_m \leq |a| + \frac{3}{2}|x| < |a| + \frac{1.5|x|}{2|x|}|a| \leq \frac{7}{4}|a|.$$

Case 2. $|u'| > |b| + |x^{q+p}x_1|$.

We write $u' = bx^{q+p}x_1s'$, $|s'| > 0$. This case is treated analogously as the previous case, only now

$$r(a) = 2\gamma(a) - |x^q| - (|b| + |s'|)$$

and

$$|b| + |s'| > |\bar{a}| + |x|$$

leading to the same

$$r(a) = 2\gamma(a) - |x^{q-1}| - |\bar{a}|$$

That is to say, we shall obtain the same bounds as above.

Putting together this issue, we finish the induction with $W = uw$

$$2\gamma(W) = 2\gamma(a) + 2\gamma(bw)$$

$$< -|u| + |u| + 2\gamma(a) - |u'| + C|bw| = -|u| + r(a) + C|bw|$$
$$< -|u| + C|abw| = -|u| + C|W|.$$

Here comes the last issue to handle.

7.2 Type One or Two

It is the last phase of the proof. We have observed that for all double squares (v, V) starting inside u with the relation

$$u = ab, v = bs$$

then

$$2\gamma(a) > |s|.$$

Therefore, by Lemma 7.1, all of them are isomorphic to (u, U). Consequently, if $p = 1$

$$\gamma(u) \leq \eta(u, U) \leq |x_1|$$

and, if $p = 2$

$$\gamma(u) \leq \eta(u, U) \leq |x|.$$

Further, for the first external double square (u', U') of (u, U)

$$r(u) = |u| + 2\gamma(u) - (||u, u'|| + |u'|)$$

with

$$||u, u'|| + |u'| > \frac{1}{2}|x^p x_1|.$$

In case there is no external double square for (u, U) we assume the usual convention $||u, u'|| + |u'| = \frac{1}{2}|x^{q+p}x_1|$. Thus

$$r(u) = |u| + 2\gamma(u) - (||u, u'|| + |u'|) \leq |x^p x_1| + 2|x_1| - \frac{1}{2}|xx_1|$$

$$= |x^p x_1| - \frac{1}{2}|x| + \frac{3}{2}|x_1| < |x^p x_1| + |x_1|$$

$$< |x^p x_1| + \frac{1}{2}|xx_1| = \frac{3}{2}|xx_1| = \frac{3}{2}|u|$$

if $p = 1$ and

$$r(u) = |x^p x_1| + 2|x| - \frac{1}{2}|x^2 x_1| = |x^p x_1| + 2|x| - \frac{1}{2}|x_1|$$
$$< |x^p x_1| + |x| < |x^p x_1| + \frac{1}{2}|x^2 x_1| = \frac{3}{2}|x^p x_1| = \frac{3}{2}|u|$$

if $p = 2$. Finally, with $W = uw = uw_1u'w_2$ and $|w_1| = ||u, u'||$ in mind, a routine computation yields

$$2\gamma(W) = 2\gamma(u) + 2\gamma(w) = 2\gamma(u) + 2\gamma(u'w_2)$$
$$< -|u| + r(u) + |w_1| + |u'| - |u'| + C|u'w_2| = -|u| + C|u| + |w_1| + C|u'w_2|$$
$$< -|u| + C|u| + |w_1| + C|w_1u'w_2| = -|u| + C|uw_1u'w_2| = -|u| + C|W|$$

which terminates the induction, Subsection 7.2 and the entire proof.

8 Concluding Remarks

There arise some questions from all the machinery above. We have exerted efforts only on the double square ever since, then what happens if W has no double squares at all? It is true that $\limsup \frac{\gamma(W)}{|W|} = 1$ when $|W| \to \infty$? Or there exists a constant c so that $\gamma(W) < c|W|$ for all such $|W|$? I am rather inclined to the later possibility. As far as the bounds are concerned, they are weak and bear rather a qualitative character, but there is room for improvement, and what is more important for us is, I think, the approach introduced and the tools worked out here. I may be missing a simple idea to drastically cut down the estimates or to settle the question once and for all but it is still worthy to carry out deeper investigations on the number of coperiod double squares, the U-size and U-mean double squares in connection with the Lemma 7.2.

References

[1] V. Bruyère, *On maximal codes with bounded synchronization delay,* Theoretical Computer Science **204**(1998), 11–28.

[2] N. J. Fine, H. S. Wilf, *Uniqueness Theorem for Periodic Functions,* Proceedings of the American Mathematical Society, **16**(1965), 109–114.

[3] A. S. Fraenkel, J. Simpson, *How many squares can a string contain?,* Journal of Combinatorial Theory Ser. **A 82**(1995), 112–120.

[4] L. Ilie, *A note on the number of squares in a word,* Theoretical Computer Science **380**(2007), 373–376.

[5] M. Lothaire, *Combinatorics on Words,* Addison-Wesley, Reading, MA, 1983.

[6] H. J. Shyr, *Free Monoid and Languages,* Taichung, 2001.

Nguyen Huong Lam
Hanoi Institute of Mathematics
18 Hoang Quoc Viet Road, 10307 Hanoi, Vietnam.
Email: nhlam@math.ac.vn