How many double squares can a string contain?

F. Franek, joint work with A. Deza and A. Thierry

Advanced Optimization Laboratory
Department of Computing and Software
McMaster University, Hamilton, Ontario, Canada

Department of Mathematics
University of Guelph
March, 2014
Outline

1. Motivation and background
2. Basic notions and tools
3. Double squares
4. Inversion factors
5. Rightmost double squares
6. An upper bound for the number of double squares
7. Main theorems
8. Conclusion
Motivation and background

We are dealing with finite strings over finite alphabets. There is no particular requirement about the order of the alphabet.

What is the *maximum number of distinct squares problem*?

We are counting types of squares rather than their occurrences.

```
a a b a a b a a a
```

has 6 occurrences of squares, but only 4 distinct squares, *aa*, *aabaab*, *abaaba*, and *baabaa*.

How many double squares can a string contain?
A trivial bound: the number of all occurrences of primitively rooted squares in a string of length $n$ is bounded by $O(n \log n)$ (Crochemore 1978) and the number of distinct non-primitively rooted squares is $O(n)$ (Kubica et al. 2013).

Could it be $O(n)$? And if so, what would be the constant?

Why this is not simple? In a string of length $n$, $O(\log n)$ squares can start at the same position!

```
a a b a a b a a b a a b b
```
It is easy to compute it for short strings, so why induction cannot be used?

```
a a b a a b  +  a a b a a b
```

Concatenation does both “destroys” existing types through multiple-occurrences and “creates” new types. Of course, same holds true for the reverse process - partitioning of strings.
Theorem (Fraenkel-Simpson, 1998)

There are at most $2n$ distinct squares in a string of length $n$.

Count only the rightmost occurrences. Fraenkel-Simpson showed that if there are three rightmost squares $uu$, $vv$, and $ww$ starting at the same position so that $|u| < |v| < |w|$, then $ww$ contains a farther copy of $uu$, based on Crochemore-Rytter (1995) Lemma showing that in such a case, $|w| \geq |u| + |v|$.
**Fraenkel-Simpson** hypothesized that the number of distinct squares should be bounded by $n$, i.e.

$$\sigma(n) \leq n$$

where $\sigma(n) = \max \{ s(x) : x \text{ is a string of length } n \}$.

**Fraenkel-Simpson** gave an infinite sequence of strings $\{x_n\}_{n=1}^{\infty}$ so that $|x_n| \uparrow \infty$ and

$$\frac{s(x_n)}{|x_n|} \uparrow 1$$

where $s(x) = \text{number of distinct squares in } x$. 

How many double squares can a string contain?

University of Guelph, March 2014
In 2005 Ilie provided a simpler proof of Fraenkel-Simpson’s Theorem and in 2007 presented an asymptotic upper bound of $2n – \theta(\log n)$.

In 2011 Deza-F. proposed a $d$-step approach to the problem and conjectured that $\sigma_d(n) \leq n – d$, where $\sigma_d(n) = \max \{ s(x) : x \text{ is a string of length } n \text{ with } d \text{ distinct symbols} \}$. 
Basic notions and tools

**Definition**

*non-trivial power* of a string $x$ is a concatenation of $m$ copies of $x$; $x^2$ is a *square*, $x^3$ a *cube*.

A string $x$ is *primitive* if $x \neq y^n$ for any $y$ and any $n \geq 2$.

*primitive root* of $x$ is the shortest $y$ so that $x = y^n$.

(*Note that $y$ must be primitive.*)

$x$ and $y$ are *conjugates* if $x = uv$ and $y = vu$ for some $u, v$. 

How many double squares can a string contain?

University of Guelph, March 2014
Lemma (Synchronization principle)

Given a primitive string $x$, a proper suffix $y$ of $x$, a proper prefix $z$ of $x$, and $m \geq 0$, there are exactly $m$ occurrences of $x$ in $yx^m z$.

Lemma (Common factor lemma)

For any strings $x$ and $y$, if a non-trivial power of $x$ and a non-trivial power of $y$ have a common factor of length $|x| + |y|$, then the primitive roots of $x$ and $y$ are conjugates.
**Double squares**

- *Fraenkel-Simpson*: only two rightmost squares can start at the same position. Thus, only one rightmost square or two rightmost squares may start at any position.

- Lam (2009 – unpublished) tried bounding the number of double squares and hence bound the number of distinct squares. His approach is based on a taxonomy of all possible configurations of two double squares yielding a bound of \( \frac{94}{48} n \approx 1.98n \).
A configuration of two squares

\[
\begin{array}{ccc}
U & & U \\
\hline
u & & u \\
\end{array}
\]

has been investigated in many different contexts:

- *Smyth et. al.*: with intention to find a position for amortization argument for runs conjecture.
- in computational framework by *Deza-F.-Jiang*: such configurations are used in *Liu’s* Ph.D. thesis to speed up computation of \( \sigma_d(n) \).
- *Lam*: two rightmost squares have a unique structure.
Lemma

Let $uu$ and $UU$ be two squares in a string $x$ starting at the same position with $|u| < |U|$ such that either

(a) both $uu$ and $UU$ are rightmost occurrences, or
(b) $uu$ or $UU$ is primitively rooted and $|U| < |uu|$

Then $|u| < |U| < |uu| < |UU|$ and there is a unique primitive string $u_1$, a unique proper prefix $u_2$ of $u_1$, and unique integers $e_1$ and $e_2$ satisfying $1 \leq e_2 \leq e_1$ such that $u = u_1^{e_1} u_2$ and $U = u_1^{e_1} u_2 u_1^{e_2}$; i.e. $uu$ and $UU$ form a double square.
Thus, only strings of length at least 10 may contain a double square: $|UU| = 2((u(1)+u(2))|u_1|+|u_2|) \geq 2((1+1)2+1) = 10$. 
Cyclic shift (rotation) to the right is controlled by

$$lcp(u_1, \overline{u_1})$$

while cyclic shift to the left is controlled by

$$lcs(u_1, \overline{u_1})$$

$lcp = \text{largest common prefix}$

$lcs = \text{largest common suffix}$
Motivation and background  Basic notions and tools  Double squares  Inversion factors  Rightmost double squares  An upper bound for the number of double squares  Conclusion  References

How many double squares can a string contain?

How many double squares can a string contain? University of Guelph, March 2014
$u_1 = aaabaa, u_2 = aaab, \overline{u}_2 = aa, u(1) = 2,$ and $u(2) = 1.$

How many double squares can a string contain?

University of Guelph, March 2014
Definition

For a double square $\mathcal{U}$, $\overline{v}vv\overline{v}$ where $|\overline{v}| = |\overline{u_2}|$ and $|v| = |u_2|$ is an inversion factor.

\[
\mathcal{U} = u_1 \mathcal{U}(1) u_2 u_1 \mathcal{U}(2) + \mathcal{U}(1) u_2 u_1 \mathcal{U}(2) =
\]

\[
u_1(\mathcal{U}(1)-1) u_2 \overline{u_2} u_2 u_2 \overline{u_2} u_1 \mathcal{U}(2) + \mathcal{U}(1)-2 u_2 \overline{u_2} u_2 u_2 \overline{u_2} u_1 (\mathcal{U}(2)-1)
\]

$N_1$ and $N_2$ natural inversion factors.

How many double squares can a string contain?

University of Guelph, March 2014
A cyclic shift of an inversion factor is an inversion factor, also controlled by \( lcp(u_1, \overline{u}_1) \) and \( lcs(u_1, \overline{u}_1) \).
All inversion factors are cyclic shifts of the natural ones:

**Lemma (Inversion factor lemma)**

Given a double square $\mathcal{U}$, there is an inversion factor of $\mathcal{U}$ within the string $UU$ starting at position $i \iff i \in [L_1, R_1] \cup [L_2, R_2]$. 
Inversion factor lemma for distinct squares

Theorem \textit{(Fraenkel-Simpson, Ilie)}

At most two rightmost squares can start at the same position.

Let us assume that 3 rightmost squares $uu$, $UU$, $vv$ start at the same position.

By item (c) of Inversion factor lemma, $uu$ and $UU$ form a double square $U$: $u = u_1^{(1)} u_2$ and $U = u_1^{(1)} u_2 u_1^{(2)}$.

Since the first $v$ contains an inversion factor, the second $v$ must also contain an inversion factor.

Cont. on the next slide
If the inversion factor in the second $v$ were from $[L_2, R_2]$, then $|v| = |U|$, a contradiction.
Hence $v$ must not contain an inversion factor from $[L_2, R_2]$ and so $u_1 U(1) u_2 u_1 U(1) + U(2)^{-1} u_2$ must be a prefix of $v$.
Therefore $vv$ contains another copy of $u_1 U(1) u_2 u_1 U(1) u_2 = uu$, a contradiction.
Fundamental Lemma:

**Lemma**

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

(a) $s(\mathcal{V}) < R_1(\mathcal{U})$, in which case either

(a1) $\mathcal{V}$ is an $\alpha$-mate of $\mathcal{U}$ (cyclic shift), or
(a2) $\mathcal{V}$ is a $\beta$-mate of $\mathcal{U}$ (cyclic shift of $\mathcal{U}$ to $\mathcal{V}$), or
(a3) $\mathcal{V}$ is a $\gamma$-mate of $\mathcal{U}$ (cyclic shift of $\mathcal{U}$ to $\mathcal{V}$), or
(a4) $\mathcal{V}$ is a $\delta$-mate of $\mathcal{U}$ (big tail),

or

(b) $R_1(\mathcal{U}) \leq s(\mathcal{V})$, then

(b1) $\mathcal{V}$ is a $\varepsilon$-mate of $\mathcal{U}$ (big gap).
**α-mate (cyclic shift):**

\[
\begin{array}{c}
\text{[} & \text{[} & \text{[} & \text{[} & \text{[}
\end{array}
\]
\[
\text{aaabaaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaaaabaaabaaa...
β-mate (cyclic shift of $U$ to $V$)
\( \gamma \)-mate (cyclic shift of \( U \) to \( v \))
δ-mate (big tail)

sufficiently big tail

[ ] [ ) ( ]

aabaabaabaabaabaabaabaabaabaab

[ ] [ ) ( ]
aabaabaabaabaabaabaabaabaabaabaab
\( \varepsilon \)-mate (big gap)

\[
\begin{align*}
R_1 & \quad \text{sufficiently big gap} \\
\text{aabaabaabaabaabaabaabaabaabaabaabaabaab} & \quad \text{aabaabaabaabaabaabaabaabaabaabaabaabaabaab}
\end{align*}
\]
An upper bound for the number of double squares

We show by induction a bound \( \delta(x) \leq \frac{5}{6} |x| - \frac{1}{3} |u| \), where \( uu \) is the shorter square of the leftmost double square of \( x \).

The fundamental lemma basically says that either the gap \( G(U, V) \) is "big" or the tail \( T(U, V) \) is "big" (for \( \delta \)-mate and \( \varepsilon \)-mate), or it is case of \( \alpha \)-mate, \( \beta \)-mate, or \( \gamma \)-mate.
**Lemma (Gap-Tail lemma)**

\[
\delta(x') \leq \frac{5}{6} |x'| - \frac{1}{3} |v| \quad \text{implies}
\]

\[
\delta(x) \leq \frac{5}{6} |x| - \frac{1}{3} |u| + h - \frac{1}{2} |G(U, V)| - \frac{1}{3} |T(U, V)|
\]
We deal with $\alpha$-mates, $\beta$-mates, and $\gamma$-mates separately.

It is possible as they form families, either a pure $\alpha$-family, or $\alpha+\beta$-family, or $\alpha+\beta+\gamma$-family.
Motivation and background

Basic notions and tools

Double squares

Inversion factors

Rightmost double squares

An upper bound for the number of double squares

Conclusion

References

\( U \)-family consists only of \( \alpha \)-mates

Illustration of \( \alpha \)-family with \( U(1) = U(2) \)

How many double squares can a string contain?

University of Guelph, March 2014
Illustration of $\alpha$-family with $U(1) > U(2)$

```
aaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaaaaabaaa
```
It is easy to bound the size of $\alpha$-family, as it is controlled by $lcp(u_1, \bar{u}_1)$ and $lcp(y, u_2)$ where $y$ is $x$ without $UU$: the size $\leq |u_1|$.

- Either there are no other double squares, and then it can be shown directly that the bound holds, or

- There is a $\gamma$ underneath, and we can use induction using the Gap-Tail lemma. $\gamma$ must be either $\gamma$-mate, or $\delta$-mate, or $\epsilon$-mate, and the Gap-Tail lemma can be applied to propagate the bound.
Illustration of $\alpha+\beta$-family

$\mathcal{U}$-family consists of $\alpha$-mates and $\beta$-mates

How many double squares can a string contain? University of Guelph, March 2014
It is more complicated to bound the size of a $\alpha+\beta$-family:

$$|\alpha+\beta\text{-family}| \leq \begin{cases} \left\lceil \frac{U(1)-U(2)}{2} \right\rceil |u_1| & \text{if } U(2) = 1 \\ \frac{U(1)-U(2)}{2} |u_1| & \text{if } U(2) > 1 \end{cases}$$

- Either there are no other double squares, and then it can be shown directly that the bound holds, or
- There is a $\nu$ underneath, and we can use induction using the Gap-Tail lemma. $\nu$ must be either $\delta$-mate, or $\varepsilon$-mate, and the Gap-Tail lemma can be applied to propagate the bound. (Special care needed for $\varepsilon$-mate case and super-$\varepsilon$-mate must be put in play!)
Illustration of $\alpha+\beta+\gamma$-family

How many double squares can a string contain?
It is quite complex to bound the size of a $\alpha + \beta + \gamma$-family:

$$|\alpha + \beta + \gamma\text{-family}| \leq \frac{2}{3}(u(1) + 1)|u_1|$$

- Either there are no other double squares, and then it can be shown directly that the bound holds, or

- There is a $\mathcal{V}$ underneath, and we can use induction using the Gap-Tail lemma. $\mathcal{V}$ must be either $\delta$-mate, or $\varepsilon$-mate, and the Gap-Tail lemma can be applied to propagate the bound.
Main theorems

**Theorem**

*The number of double squares in a string of length n is bounded by \( \lfloor \frac{5n}{6} \rfloor \).*

**Corollary**

*The number of distinct squares in a string of length n is bounded by \( \lfloor \frac{11n}{6} \rfloor \).*
We presented a universal upper bound of $\frac{11n}{6}$ for the maximum number of distinct squares in a string of length $n$.

A bound of $\frac{5n}{6}$ for the maximum number of double squares.

It improves the universal bound of $2n$ by Fraenkel-Simpson.

It improves the asymptotic bound of $2n - \Theta(\log n)$ by Ilie.

The combinatorics of double squares is interesting on its own and possibly can be used for some other problems.
THANK YOU
M. Crochemore and W. Rytter.
Squares, cubes, and time-space efficient string searching.

A. Deza and F. Franek.
A $d$-step approach to the maximum number of distinct squares and runs in strings.

A. Deza, F. Franek, and M Jiang.
A computational framework for determining square-maximal strings.
A.S. Fraenkel and J. Simpson.
How many squares can a string contain?

More results on overlapping squares.

L. Ilie.
A simple proof that a word of length \( n \) has at most \( 2n \) distinct squares.
L. Ilie. 
A note on the number of squares in a word. 

E. Kopylova and W.F. Smyth. 
The three squares lemma revisited. 

M. Kubica, J. Radoszewski, W. Rytter, and T. Waleń. 
On the maximum number of cubic subwords in a word. 

N. H. Lam. 
On the number of squares in a string. 
M. J. Liu.
Combinatorial optimization approaches to discrete problems.