

# On the Number of Distinct Squares

## Abstract

Frantisek Franek

Department of Computing and Software  
McMaster University, Hamilton, Ontario, Canada  
[franek@mcmaster.ca](mailto:franek@mcmaster.ca)

**Abstract.** Counting the number of types of squares rather than their occurrences, we consider the problem of bounding the maximum number of distinct squares in a string. Fraenkel and Simpson showed in 1998 that a string of length  $n$  contains at most  $2n$  distinct squares and indicated that all evidence pointed to  $n$  being a natural universal upper bound. Ilie simplified the proof of Fraenkel-Simpson's key lemma in 2005 and presented in 2007 an asymptotic upper bound of  $2n\Theta(\log n)$ . We show that a string of length  $n$  contains at most  $\lfloor 11n/6 \rfloor$  distinct squares for any  $n$ . This new universal upper bound is obtained by investigating the combinatorial structure of FS-double squares (named so in honour of Fraenkel and Simpson's pioneering work on the problem), i.e. two rightmost-occurring squares that start at the same position, and showing that a string of length  $n$  contains at most  $\lfloor 5n/6 \rfloor$  FS-double squares. We will also discuss a much more general approach to double-squares, i.e. two squares starting at the same position and satisfying certain size conditions. A complete, so-called canonical factorization of double-squares that was motivated by the work on the number of distinct squares is presented in a separate contributed talk at this conference. The work on the problem of the number of distinct squares is a joint effort with Antoine Deza and Adrien Thierry.

*At the time of the presentation of this talk, the slides of the talk are also available at  
<http://www.cas.mcmaster.ca/~franek/PSC2014/invited-talk-slides.pdf>*

*This work was supported by the Natural Sciences and Engineering Research Council of Canada*

## References

1. M. CROCHMORE AND W. RYTTER: *Squares, cubes, and time-space efficient string searching*. Algorithmica, 13:405–425, 1995.
2. A. DEZA AND F. FRANEK: *A d-step approach to the maximum number of distinct squares and runs in strings*. Discrete Applied Mathematics, 163:268–274, 2014.
3. A. DEZA, F. FRANEK, AND M. JIANG: *A computational framework for determining square-maximal strings*. In J. Holub and J. Žďárek, editors, Proceedings of the Prague Stringology Conference 2012, pp. 111–119, Czech Technical University in Prague, Czech Republic, 2012.
4. A. S. FRAENKEL AND J. SIMPSON: *How many squares can a string contain?* Journal of Combinatorial Theory, Series A, 82(1):112–120, 1998.
5. F. FRANEK, R. C. G. FULLER, J. SIMPSON, AND W. F. SMYTH: *More results on overlapping squares*. Journal of Discrete Algorithms, 17:2–8, 2012.
6. L. ILIE: *A simple proof that a word of length  $n$  has at most  $2n$  distinct squares*. Journal of Combinatorial Theory, Series A, 112(1):163–163, 2005.
7. L. ILIE: *A note on the number of squares in a word*. Theoretical Computer Science, 380(3):373–376, 2007.
8. E. KOPYLOVA AND W. F. SMYTH: *The three squares lemma revisited*. Journal of Discrete Algorithms, 11:3–14, 2012.

9. M. KUBICA, J. RADOSZEWSKI, W. RYTTER, AND T. WALEŃ: *On the maximum number of cubic subwords in a word.* European Journal of Combinatorics, 34:27–37, 2013.
10. N. H. LAM: *On the number of squares in a string.* AdvOL-Report 2013/2, McMaster University 2013.
11. M. J. LIU: *Combinatorial optimization approaches to discrete problems.* Ph.D. thesis, Department of Computing and Software, McMaster University 2013.