# On Singularities Of Extremal Periodic Strings

F. Franek (joint work with A. Deza)

Department of Computing and Software
McMaster University, Hamilton, Ontario, Canada

Workshop on Challenges in Combinatorics on Words
at the Fields Institute, April 22-26, 2013

# Outline



McMaster University

- In1998 by *Fraenkel* and *Simpson* showed that the number of distinct squares in a string of length $n$ is at most $2n$ and hypothesized that the bound should be $n$.
  In 2005 *Ilie* provided a simpler proof and in 2007 presented an asymptotic upper bound of $2n - \Theta(\log n)$.
- In 1999 *Kolpakov* and *Kucherov* proved that the maximum number of runs in a string is linear in the string's length and conjectured that it is in fact bounded by the length.
  Many additional authors (*Rytter, Smyth, Simpson, Puglisi, Crochemore, Ilie, Kusano, Matsubara, Ishino, Bannai, Shinohara, FF*) contributed to improving the lower and upper bounds to the current asymptotic

  $$0.944565n \leq \rho(n) \leq 1.029n$$

We consider the role played by the size of the alphabet of the string in both problems and investigate the functions $\sigma_d(n)$ and $\rho_d(n)$, i.e. the maximum number of distinct primitively rooted squares, respectively runs, over all strings of length $n$ containing exactly $d$ distinct symbols. We revisit earlier results and conjectures and express them in terms of singularities of the two functions where a pair $(d, n)$ is a *singularity* if $\sigma_d(n) - \sigma_{d-1}(n-2) \geq 2$, or $\rho_d(n) - \rho_{d-1}(n-2) \geq 2$ respectively.

| $d$ \ $n-d$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | **2** | 2 | 3 | 3 | 4 | 5 | 6 | 7 | 7 | 8 | 9 | 10 | 11 | 12 |
| 3 | 2 | **3** | 3 | 4 | 4 | 5 | 6 | 7 | 8 | 8 | 9 | 10 | 11 | 12 |
| 4 | 2 | 3 | **4** | 4 | 5 | 5 | 6 | 7 | 8 | 9 | 9 | 10 | 11 | 12 |
| 5 | 2 | 3 | 4 | **5** | 5 | 6 | 6 | 7 | 8 | 9 | 10 | 10 | 11 | 12 |
| 6 | 2 | 3 | 4 | 5 | **6** | 6 | 7 | 7 | 8 | 9 | 10 | 11 | 11 | 12 |
| 7 | 2 | 3 | 4 | 5 | 6 | **7** | 7 | 8 | 8 | 9 | 10 | 11 | 12 | 12 |
| 8 | 2 | 3 | 4 | 5 | 6 | 7 | **8** | 8 | 9 | 9 | 10 | 11 | 12 | 13 |
| 9 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | **9** | 9 | 10 | 10 | 11 | 12 | 13 |
| 10 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | **10** | 10 | 11 | 11 | 12 | 13 |
| 11 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | **11** | 11 | 12 | 12 | 13 |
| 12 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | **12** | 12 | 13 | 13 |
| 13 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | **13** | 13 | 14 |
| 14 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | **14** | 14 |
| 15 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | **15** |

Table 1.1: $(d, n-d)$ table for $\sigma_d(n)$ with $2 \le d \le 15$ and $2 \le n-d \le 15$

http://optlab.mcmaster.ca/~jiangm5/research/square.html

McMaster University

$n-d$

| | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2 | **2** | 2 | 3 | 4 | 5 | 5 | 6 | 7 | 8 | ***8*** | 10 | 10 | 11 | 12 |
| | 3 | 2 | **3** | 3 | 4 | 5 | 6 | 6 | 7 | 8 | 9 | ***10*** | 11 | 11 | 12 |
| | 4 | 2 | 3 | **4** | 4 | 5 | 6 | 7 | 7 | 8 | 9 | 10 | 11 | 12 | 12 |
| | 5 | 2 | 3 | 4 | **5** | 5 | 6 | 7 | 8 | 8 | 9 | 10 | 11 | 12 | 13 |
| | 6 | 2 | 3 | 4 | 5 | **6** | 6 | 7 | 8 | 9 | 9 | 10 | 11 | 12 | 13 |
| | 7 | 2 | 3 | 4 | 5 | 6 | **7** | 7 | 8 | 9 | 9 | 10 | 11 | 112 | 13 |
| | 8 | 2 | 3 | 4 | 5 | 6 | 7 | **8** | 8 | 9 | 10 | 11 | 11 | 12 | 13 |
| $d$ | 9 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | **9** | 9 | 10 | 11 | 12 | 12 | 13 |
| | 10 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | **10** | 10 | 11 | 12 | 13 | 13 |
| | 11 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | **11** | 11 | 12 | 13 | 13 |
| | 12 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | **12** | 12 | 13 | 14 |
| | 13 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | **13** | 13 | 14 |
| | 14 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | **14** | 14 |
| | 15 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | **15** |

Table 1.2: $(d, n-d)$ table for $\rho_d(n)$ with $2 \leq d \leq 15$ and $2 \leq n-d \leq 15$

`http://optlab.mcmaster.ca/~bakerar2/research/runmax/index.html`

McMaster University

## Proposition

**($s_1$)** $0 \leq \sigma_d(n+1) - \sigma_d(n) \leq 2$ *for* $n \geq d \geq 2$

**($s_2$)** $\sigma_d(n) \leq \sigma_{d+1}(n+1)$ *for* $n \geq d \geq 2$

**($s_3$)** $\sigma_d(n) < \sigma_{d+1}(n+2)$ *for* $n \geq d \geq 2$

**($s_4$)** $\sigma_d(n) = \sigma_{d+1}(n+1)$ *for* $2d \geq n \geq d \geq 2$

**($s_5$)** $\sigma_d(n) \geq n-d$, $\sigma_d(2d+1) \geq d$ *and* $\sigma_d(2d+2) \geq d+1$ *for* $2d \geq n \geq d \geq 2$

**($s_6$)** $\sigma_{d-1}(2d-1) = \sigma_{d-2}(2d-2)$ *and* $0 \leq \sigma_d(2d) - \sigma_{d-1}(2d-1) \leq 1$ *for* $d \geq 4$

**($s_7$)** $1 \leq \sigma_{d+1}(2d+2) - \sigma_d(2d) \leq 2$ *for* $d \geq 2$.

McMaster University

## Corollary

**($c_1$)** $\sigma_2(n) \leq 2n - 51$ *for* $n \geq 41$

**($c_2$)** $\sigma(n) \leq 2n - 19$ *for* $n \geq 30$.

## Conjecture

*For any* $n \geq d \geq 2$, $\sigma_d(n) \leq n - d$

## Theorem

*Let* $(d, 2d)$ *be the first singularity on the main diagonal, i.e. the least* $d$ *such that* $\sigma_d(2d) - \sigma_{d-1}(d-2) \geq 2$. *Then any square-maximal* $(d, 2d)$*-string does not contain a pair but must contain at least* $\lceil \frac{2d}{3} \rceil$ *singletons.*

## Theorem

**($e_1$)** *no* $(d, 2d)$ *singularity* $\iff \{\sigma_d(n) \le n{-}d$ *for* $n \ge d \ge 2\}$

**($e_2$)** $\{\sigma_d(n) \le n{-}d$ *for* $n \ge d \ge 2\} \iff$
$\{\sigma_d(4d) \le 3d$ *for* $d \ge 2\}$

**($e_3$)** $\{\sigma_d(n) \le n{-}d$ *for* $n \ge d \ge 2\} \iff$
$\{\sigma_d(2d{+}1){-}\sigma_d(2d) \le 1$ *for* $d \ge 2\}$

**($e_4$)** *no* $(d, 2d{+}1)$ *singularity* $\implies$
$\{$*no* $(d, 2d)$ *singularity and* $\sigma_d(n) \le n{-}d{-}1$ *for* $n > 2d \ge 4\}$

**($e_5$)** $\{\sigma_d(2d) = \sigma_d(2d{+}1)$ *for* $d \ge 2\} \implies$
$\{$*no* $(d, 2d)$ *singularity and* $\sigma_d(n) \le n{-}d{-}1$ *for* $n > 2d \ge 4\}$

**($e_6$)** $\{\sigma_d(2d) = \sigma_d(2d{+}1)$ *for* $d \ge 2\} \implies$
$\{$*square-maximal* $(d, 2d)$*-strings are, up to relabelling, unique and equal to* $a_1a_1a_2a_2a_2 \ldots a_da_d\}$

## Proposition

($r_1$) $\rho_d(n) \leq \rho_{d+1}(n{+}1)$ *for* $n \geq d \geq 2$

($r_2$) $\rho_d(n) \leq \rho_d(n{+}1)$ *for* $n \geq d \geq 2$

($r_3$) $\rho_d(n) < \rho_{d+1}(n{+}2)$ *for* $n \geq d \geq 2$

($r_4$) $\rho_d(n) = \rho_{d+1}(n{+}1)$ *for* $2d \geq n \geq d \geq 2$

($r_5$) $\rho_d(n) \geq n{-}d$, $\rho_d(2d{+}1) \geq d$ *and*
$\rho_d(2d{+}2) \geq d{+}1$ *for* $2d \geq n \geq d \geq 2$

($r_6$) $\rho_{d-1}(2d{-}1) = \rho_{d-2}(2d{-}2) = \rho_{d-3}(2d{-}3)$ *and*
$0 \leq \rho_d(2d){-}\rho_{d-1}(2d{-}1) \leq 1$ *for* $d \geq 5$

McMaster University

## Proposition

*Let $(d, 2d)$ be the first singularity on the main diagonal, i.e. the least $d$ such that $\rho_d(2d) - \rho_{d-1}(2d-2) \geq 2$. Then any run-maximal $(d, 2d)$-string does not contain a symbol occurring exactly $2, 3, \ldots, 7$ or $8$ times, and must contains at least $\lceil \frac{7d}{8} \rceil$ singletons.*

## Conjecture

*For any $n \geq d \geq 2$, $\rho_d(n) \leq n - d$*

## Theorem

**($e_1$)** *no* $(d,2d)$ *singularity* $\iff \{\rho_d(n) \le n{-}d$ *for* $n \ge d \ge 2\}$

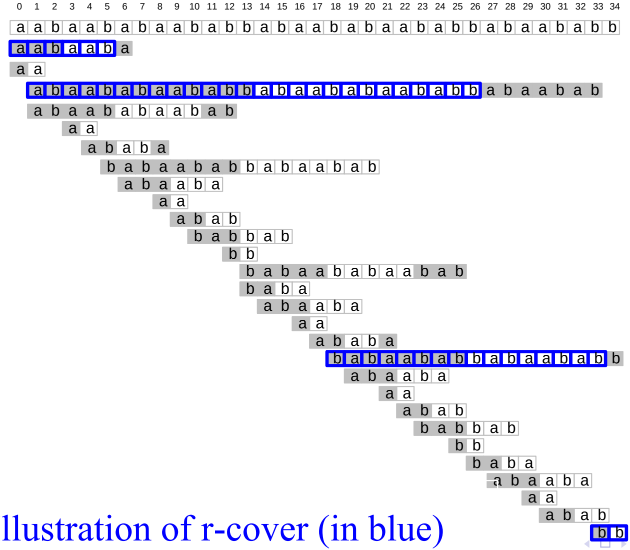
**($e_2$)** $\{\rho_d(n) \le n{-}d$ *for* $n \ge d \ge 2\} \iff$
$\{\rho_d(9d) \le 8d$ *for* $d \ge 2\}$

**($e_3$)** $\{\rho_d(n) \le n{-}d$ *for* $n \ge d \ge 2\} \iff$
$\{\rho_d(2d{+}1){-}\rho_d(2d) \le 1$ *for* $d \ge 2\}$

**($e_4$)** *no* $(d,2d{+}1)$ *singularity* $\Longrightarrow$
$\{$*no* $(d,2d)$ *singularity and* $\rho_d(n) \le n{-}d{-}1$ *for* $n > 2d \ge 4\}$

**($e_5$)** $\{\rho_d(2d) = \rho_d(2d{+}1)$ *for* $d \ge 2\} \Longrightarrow$
$\{$*no* $(d,2d)$ *singularity and* $\rho_d(n) \le n{-}d{-}1$ *for* $n > 2d \ge 4\}$

**($e_6$)** $\{\rho_d(2d) = \rho_d(2d{+}1)$ *for* $d \ge 2\} \Longrightarrow \{$*square-maximal* $(d,2d)$*-strings are, up to relabelling, unique and equal to* $a_1a_1a_2a_2a_2\dots a_da_d\}$




# Illustration of r-cover (in blue)

d = 2, n = 10 Covered: 154 not Covered: 357

---

d = 2, n = 15 Covered: 4074 not Covered: 12,309

---

d = 2, n = 20 Covered: 109,437 not Covered: 414,850

---

d = 3, n = 10 Covered: 183 not Covered: 9,147

---

d = 3, n = 15 Covered: 21,681 not Covered: 2,353,420

---

d = 3, n = 20 Covered: 1,908,923 not Covered: 578,697,523

The values for $\sigma_d(n)$ and $\rho_d(n)$ computed to date:

- For $\sigma_d(n)$ function, so far we have found two singularities: $(3, 35)$ as $\sigma_3(35) = 25$ and $\sigma_2(33) = 23$, and $(3, 36)$ as $\sigma_3(36) = 26$ and $\sigma_2(34) = 24$.
- $\sigma_3(33) = 24 \geq \sigma_2(33) = 23$ (colorblueno binary string of length 33 attains the maximum)
- For $\rho_d(n)$ function, so far we have found three singularities: $(3, 15)$ as $\rho_3(15) = 10$ and $\rho_2(13) = 8$, $(3, 43)$ as $\rho_2(41) = 33$ and $\rho_3(43) = 35$, and $(4, 44)$, as $\rho_3(42) = 33$ and $\rho_4(44) \geq \rho_3(43) = 35$

McMaster University

# *THANK YOU*