

Lower Bounds for the Number of Repetitions in $2D$ Strings

Paweł Gawrychowski

Samah Ghazawi

Gad M. Landau

SPIRE 4-6.10.2021

2

Lower Bounds for the Number of Repetitions in 2D Strings

Motivation

Combinatorial properties
of repetitions in 2D strings



3 Repetitions in 1D strings

The number of distinct squares in a string of length n

➤ Squares

*aab**ab**aaaa**ab**aaaba*

➤ A string of length n contains at most ??? distinct squares:

- $2n$ - Fraenkel and Simpson 1998
- $2n - \Theta(\log n)$ - Ilie 2007
- $11/6n$ - Deza, Franek, and Thierry 2015
- $1.5n$ - Thierry 2020 (preprint)

conjectured to be n

- The irrelevance of the alphabet size in solving the square conjecture - Manea and Seki 2015
- For each value of n , there exists a string of length n containing at least $n - o(n)$ distinct squares - Fraenkel and Simpson 1998

4 Repetitions in 1D strings

The number of runs in a string of length n

➤ Runs

$bcabcabcabcdd|ef|ef|ef|ef|eg$

➤ A string of length n contains at most ??? runs:

- $O(n)$ - Kolpakov and Kucherov 1999
- $5n$ - Rytter 2007
- $3.48n$ - Puglisi, Simpson, and Smyth 2008
- $1.6n$ - Crochemore and Ilie 2008
- $1.52n$ - Giraud 2008
- $1.029n$ - Crochemore, Ilie, and Tinta 2011
- n - Bannai, I, Inenaga, Nakashima, Takeda, and Tsuruta 2017

5 Repetitions in 1D strings

The number of runs in a string of length n

➤ Runs

$bc|abc|abc|abc|dd|e|f|e|f|e|f|e|g$

➤ A **binary** string of length n contains at most **???** runs:

- $n - 3$ - Bannai, I, Inenaga, Nakashima, Takeda, and Tsuruta **2014**
- $0.957n$ - Fisher, Holub, I, and Lewenstein **2015**
- $0.9482n$ - Holub **2017**

6 Repetitions in 1D strings

The number of runs in a string of length n

➤ Runs

$bc|abc|abc|abc|dd|ef|ef|ef|ef|g$

➤ For each value of n , there exists a string of length n containing at least ??? runs:

➤ $\frac{3}{1+\sqrt{5}}n \sim 0.927n$ - Franek and Yang 2006

➤ $0.944542n$ – Kusano, Mtsubara, Ishino, Bannai, Shinohara 2008

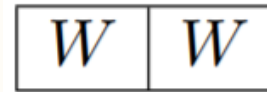
➤ $0.944565n$ – Mtsubara, Kusano, Ishino, Bannai, Shinohara 2008

➤ $0.944575712n$ – Simpstone 2010

7 Repetitions in 2D strings

The number of distinct tandems in a string of size $n \times n$

➤ Tandems



➤ Apostolico and Brimkov 2000

Primitive



$$\Theta(n^2 \log^2 n)$$

➤ Charalampopoulos, Radoszewski, Rytter, Waleń, and Zuba 2020

$$\Theta(n^3) \quad \text{Alphabet size} = n$$

a	b	a	b	b
a	a	a	a	a
a	b	a	b	c
c	d	c	d	c
e	f	e	f	c
a	a	a	a	d

8 Repetitions in 2D strings

The number of distinct quartics in a string of size $n \times n$

➤ Quartics

W	W
W	W

a	b	a	a	b
a	a	a	a	a
a	b	a	b	c
a	a	a	a	c
a	b	a	b	c
a	a	a	a	d

➤ Apostolico and Brimkov 2000

Primitive

W

\neq

W'	W'
W'	W'

$\Theta(n^2 \log n)$

➤ Charalampopoulos, Radoszewski, Rytter, Waleń, and Zuba 2020

$O(n^2 \log^2 n)$

9 Repetitions in 2D strings

The number of runs in a string of size $n \times n$

➤ Runs

W	...	W	W'
...
W	...	W	W'
W''	...	W''	W'''

➤ Amir, Landau, Marcus, and Sokol 2018

$O(n^3)$ and $\Omega(n^2)$

b	b	b	b	b	b
a	b	a	b	a	b
a	a	a	a	a	b
a	b	a	b	a	b
a	a	a	a	a	b
a	b	a	b	a	b
c	c	c	c	a	b

➤ Charalampopoulos, Radoszewski, Rytter, Waleń, and Zuba 2020

$O(n^2 \log^2 n)$

10 Our results

➤ Distinct Tandems

$$\Omega(n^3)$$

➤ Distinct Quartics

$$\Omega(n^2 \log n)$$

➤ Runs

$$\Omega(n^2 \log n)$$

Binary alphabet

W	W
-----	-----

W	W
W	W

W	...	W	W'
...
W	...	W	W'
W''	...	W''	W'''

Theorem: There exists an infinite family of $n \times n$ binary 2D strings containing $\Omega(n^2 \log n)$ runs.

11 Runs

$$A_1, A_2, \dots, A_l$$

$$A_i \text{ of size } N_i \times N_i$$

$$A_i = f(A_{i-1})$$

$$N_1 = 8$$

$$R_1 = 1$$

A_1

0	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1
0	0	0	0	0	0	1	0

12 Runs

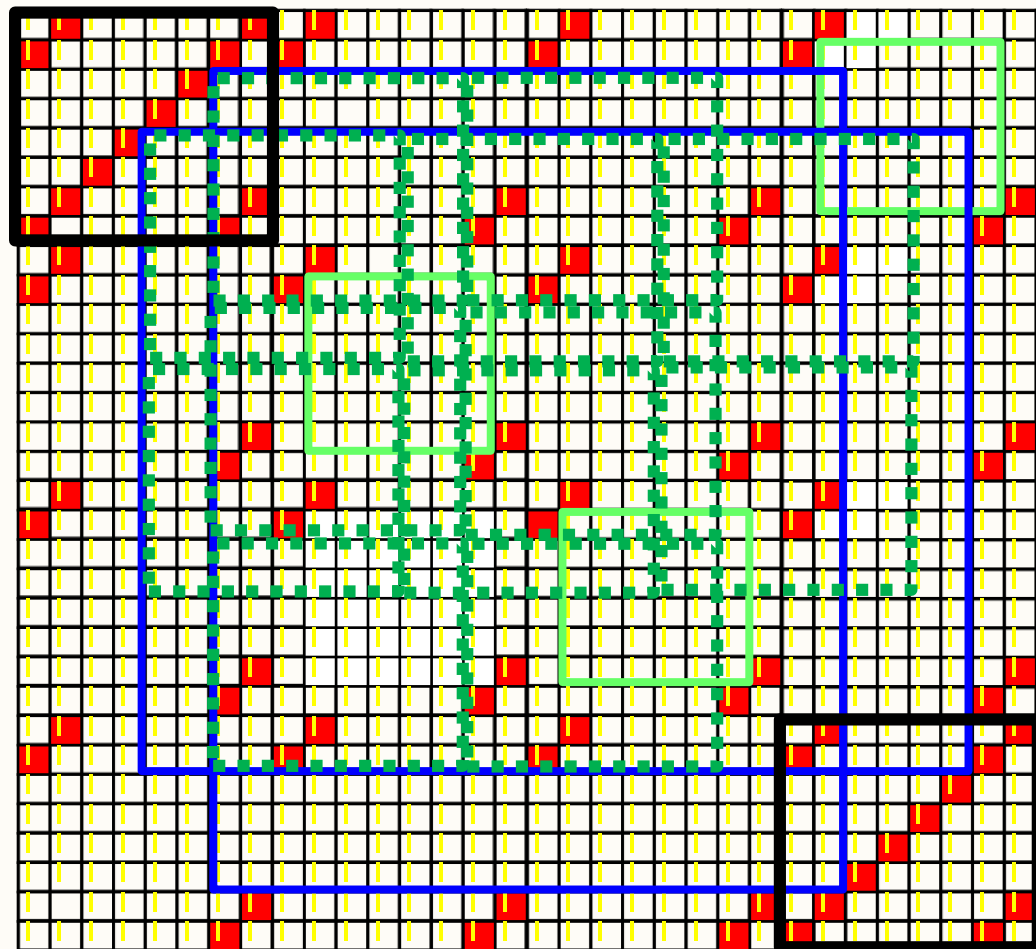
A_2

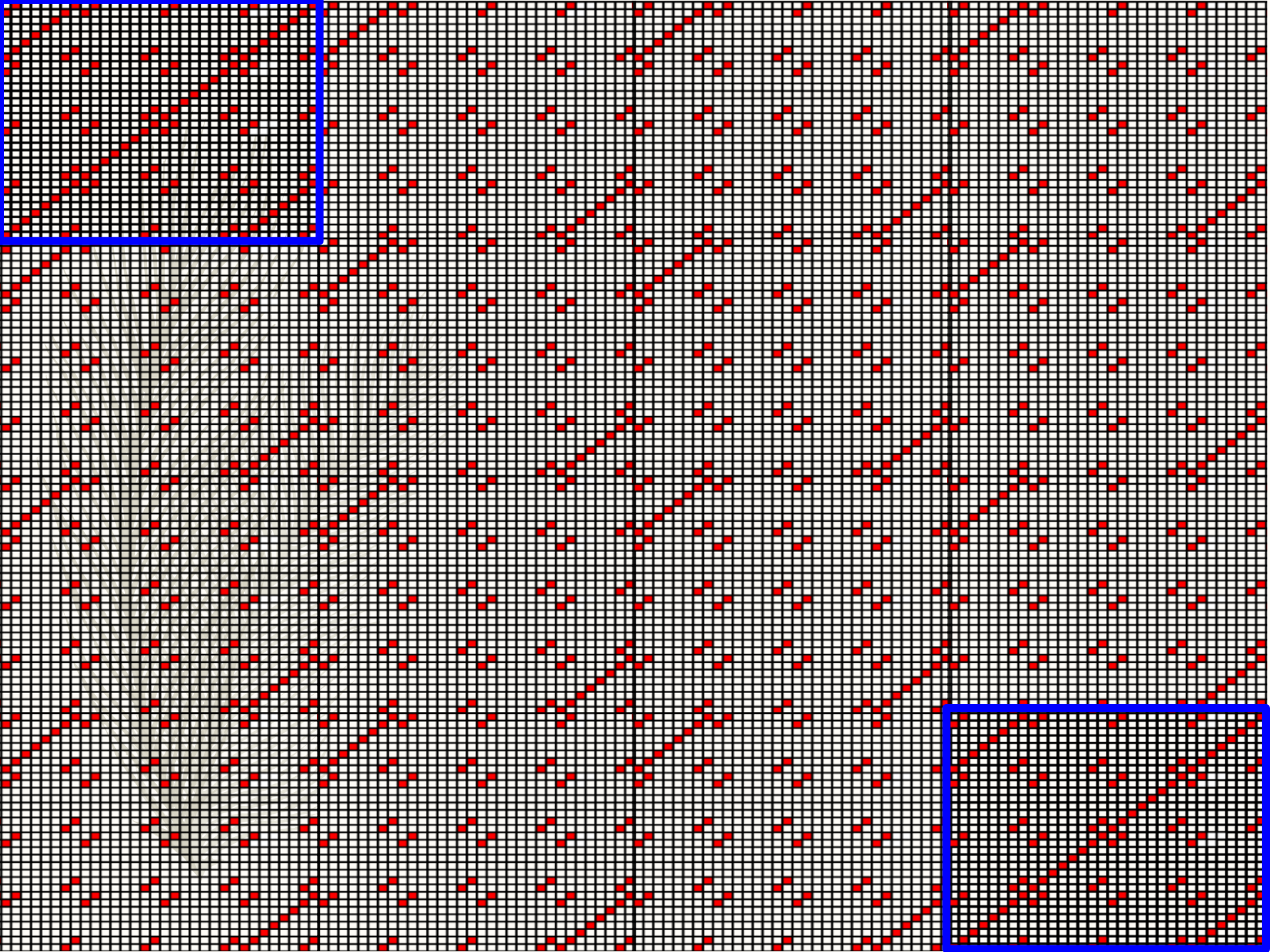
$$N_2 = 4 \cdot N_1 = 4 \cdot 8 = 32$$

$$14 \cdot R_1 = 14$$

$$R_2 = (N_1 - 1)^2 = 7^2 = 49$$

$$T_2 = R_2 + 14 \cdot R_1 = 63$$






$$14 \cdot R_2$$

$$R_3 = (N_2 - 1)^2 = 31^2$$

$$T_3 = R_3 + 14 \cdot R_2 + 216 \cdot R_1$$

$$14 \cdot 14 + 2 \cdot 2 \cdot 5$$

15 Runs

$$A_i$$

Columns/rows: $N_i = 2 \cdot 4^i$

New runs: $R_i = (N_{i-1} - 1)^2 = (2 \cdot 4^{i-1} - 1)^2$

X_j = the number of copies of A_j in A_i for $1 \leq j \leq i$

$X_i = 1, X_{i-1} = 14, X_{i-2} = 216$

Total runs: $T_i = \sum_{j=1}^i X_j \cdot R_j = \sum_{j=1}^i X_j \cdot (2 \cdot 4^{j-1} - 1)^2$

16 Runs

$$A_l$$

A_l is of size $n \times n$ where:

$$n = 2 \cdot 4^l$$

$$\forall_j X_j \geq \frac{5}{6} \cdot 16^{l-j}$$

$$T_l = \sum_{j=1}^l X_j \cdot (2 \cdot 4^{j-1} - 1)^2 \geq \sum_{j=1}^l \frac{5}{6} \cdot 16^{l-j} (2 \cdot 4^{j-1} - 1)^2$$

Theorem: There exists an infinite family of $n \times n$ **binary** 2D strings containing $\Omega(n^2 \log n)$ runs.

17 Summary

1D strings:

- $n - o(n) \leq \text{\#Squares} \leq n$ - conjectured to be n
- $0.944575712n \leq \text{\#Runs} \leq n$

2D strings:

- $n^3 \leq \text{\#Tandems} \leq n^3$
- $n^2 \log n \leq \text{\#Quartics} \leq n^2 \log^2 n$
- $n^2 \log n \leq \text{\#Runs} \leq n^2 \log^2 n$

18 Future directions

- Reducing the gap for quartics and/or runs.
- Repetitions in **3D** strings.
- Different **notions** of repetitions.





Thank you
