# A series of run-rich strings

Wataru Matsubara[1], Kazuhiko Kusano[1], Hideo Bannai[2], and Ayumi Shinohara[1]

[1] Graduate School of Information Science, Tohoku University,
Aramaki aza Aoba 6-6-05, Aoba-ku, Sendai 980-8579, Japan
{matsubara@shino., kusano@shino., ayumi@ }ecei.tohoku.ac.jp
[2] Department of Informatics, Kyushu University,
744 Motooka, Nishiku, Fukuoka 819-0395 Japan.
bannai@i.kyushu-u.ac.jp

**Abstract.** We present a new series of run-rich strings, and give a new lower bound 0.94457567 of the maximum number of runs in a string. We also introduce the general conjecture about a asymptotic behavior of the numbers of runs in the strings defined by any recurrence formula, and show the lower bound can be improved further to 0.94457571235.

## 1 Introduction

Repetitions in strings is an important element in the analysis and processing of strings. It was shown in [1] that when considering *maximal repetitions*, or *runs*, the maximum number of runs $\rho(n)$ in any string of length $n$ is $O(n)$, leading to a linear time algorithm for computing all the runs in a string. Although they were not able to give bounds for the constant factor, there have been several works to this end [2–8]. The currently known best upper bound[3] is $\rho(n) \leq 1.029n$, obtained by calculations based on the proof technique of [5, 8]. The technique bounds the number of runs for each string by considering runs in two parts: runs with long periods, and runs with short periods. The former is more sparse and easier to bound while the latter is bounded by an exhaustive calculation concerning how runs of different periods can overlap in an interval of some length.

On the other hand, an asymptotic lower bound on $\rho(n)$ was first presented in [9], where it is shown that for any $\varepsilon > 0$, there exists an integer $N > 0$ such that for any $n > N$, $\rho(n) \geq (\alpha - \varepsilon)n$, where $\alpha = \frac{3}{1+\sqrt{5}} \approx 0.927$. Although it was conjectured in [10] that this bound is optimal, a new bound was shown in [11], improving the lower bound to $\alpha = 174719/184973 \approx 0.944565$. The bound was obtained by considering the runs of an infinite series of strings $w, w^2, w^3, \ldots$, based on a run-rich string $w$. To the best of our knowledge, the current best lower bound is $\alpha = 27578248/29196442 \approx 0.9445756438404378$ achieved by a run-rich string discovered by Simon Puglisi and Jamie Simpson[4].

In this paper, we design a new series of run-rich strings defined by a simple recurrence formula, and improve the bound further to 0.94457567. We give a

[3] Presented on the website http://www.csd.uwo.ca/faculty/ilie/runs.html
[4] personal communication

conjecture for the exact number of runs contained in each string of the series, and show that the series improves the bound further to $\alpha \approx 0.94457571235$.

## 2 Preliminaries

Let $\Sigma$ be a finite set of symbols, called an *alphabet*. Strings $x$, $y$ and $z$ are said to be a *prefix*, *substring*, and *suffix* of the string $w = xyz$, respectively.

The length of a string $w$ is denoted by $|w|$. The $i$-th symbol of a string $w$ is denoted by $w[i]$ for $1 \le i \le |w|$, and the substring of $w$ that begins at position $i$ and ends at position $j$ is denoted by $w[i : j]$ for $1 \le i \le j \le |w|$. A string $w$ has period $p$ if $w[i] = w[i+p]$ for $1 \le i \le |w| - p$. A string $w$ is called *primitive* if $w$ cannot be written as $u^k$, where $k$ is a positive integer, $k \ge 2$.

A string $u$ is a *run* if it is periodic with (minimum) period $p \le |u|/2$. A substring $u = w[i : j]$ of $w$ is a *run in w* if it is a run of period $p$ and neither $w[i-1 : j]$ nor $w[i : j+1]$ is a run of period $p$, that means the run is maximal. We denote the run $u = w[i : j]$ in $w$ by the triple $\langle i, j-i+1, p\rangle$ consisting of the begin position $i$, the length $|u|$, and the minimum period $p$ of $u$. For a string $w$, we denote by $run(w)$ the number of runs in $w$.

For example, the string aabaabaaaacaacac contains the following 7 runs: $\langle 1,2,1\rangle = \mathtt{a}^2$, $\langle 4,2,1\rangle = \mathtt{a}^2$, $\langle 7,4,1\rangle = \mathtt{a}^4$, $\langle 12,2,1\rangle = \mathtt{a}^2$, $\langle 13,4,2\rangle = (\mathtt{ac})^2$, $\langle 1,8,3\rangle = (\mathtt{aab})^{\frac{8}{3}}$, and $\langle 9,7,3\rangle = (\mathtt{aac})^{\frac{7}{3}}$. Thus $run(\mathtt{aabaabaaaacaacac}) = 7$.

We are interested in the behavior of the *maxrun function* defined for all $n > 0$ by

$$\rho(n) = \max\{run(w) \mid w \text{ is a string of length } n\}.$$

Franěk, Simpson and Smyth [10] showed a beautiful construction of a series of strings which contain many runs, and later Franěk and Qian Yang [9] formally proved a family of true asymptotic lower bounds arbitrarily close to $\frac{3}{1+\sqrt{5}}n$ as follows.

**Theorem 1 ([9]).** *For any $\varepsilon > 0$ there exists a positive integer $N$ so that $\rho(n) \ge \left(\frac{3}{1+\sqrt{5}} - \varepsilon\right) n$ for any $n \ge N$.*

## 3 A New Series of Run-Rich Strings

In this section, we show a construction of a series of *run-rich* binary strings, and we give new lower bound of the number of runs in string. The series $\{t_n\}$ of strings is defined by

$$
\begin{aligned}
t_0 &= \texttt{0110101101001011010},\\
t_1 &= \texttt{0110101101001},\\
t_2 &= \texttt{01101011010010110101011010},\\
t_k &= \begin{cases} t_{k-1}t_{k-2} & (\text{if } k \mod 3 = 0),\\ t_{k-1}t_{k-4} & (\text{otherwise}),\end{cases}\\
&\quad \text{for any integer } k > 2.
\end{aligned} \tag{1}
$$

Table 1 shows the length of $\{t_n\}$ and the number of runs in $\{t_n\}$ for $i = 0, 1, \ldots, 44$. We actually counted the number of runs by implementing the linear-time algorithm proposed by Kolpakov and Kucherov [1] combined with the space-effiecient algorithm to compute Lempel Ziv Factorization proposed by Crochemore et al. [12]. In our PC with 18GB RAM, $t_{44}$ was the longest possible string to be handled. As we can see, these strings contain many runs and the ratio $run(t_n)/|t_n|$ in the third column is monotonically increasing as $n$ grows. We are interested in its limit value, and we will try to estimate it in Section 4.

Using this result in Table 1, we improve the bound. $\{t_n\}$ contains enough runs, but we can improve the bound further by considering the string $t_n^k$. First, we give a previous result about the number of runs in an infinite string obtained by concatenating the same string infinite many times.

**Theorem 2 ([11]).** *For any string $w$ and any $\varepsilon > 0$, there exists a positive integer $N$ such that for any $n \geq N$,*

$$\frac{\rho(n)}{n} > \frac{run(w^3) - run(w^2)}{|w|} - \varepsilon.$$

From Theorem 2, we show a new lower bound.

**Theorem 3.** *For any $\varepsilon > 0$ there exists a positive integer $N$ so that $\rho(n) > (\alpha - \varepsilon)\, n$ for any $n \geq N$, where $\alpha = \frac{48396453}{51236184} \approx 0.94457567$.*

*Proof.* From Table 1, we have $|t_{41}| = 51236184$, $run(t_{41}) = 48396417$, $run(t_{41}^2) = 96792871$, and $run(t_{41}^3) = 145189324$. Therefore from Theorem 2, we have

$$\frac{\rho(n)}{n} > \frac{145189324 - 96792871}{51236184} - \varepsilon.$$

□

Needless to say this bound is not optimal. If we can calculate $run(t_n)$ for larger $n$, we would be able to obtain better bounds.

## 4 Analysis of Asymptotic Behavior

In this section, we analyze the asymptotic behavior of the number of runs in $\{t_n\}$. We conjecture that $\lim_{n\to\infty} run(t_n)/|t_n| \approx 0.94457571235$.

**Table 1.** The length of $\{t_n\}$ and number of runs in $\{t_n\}$

| $n$ | $\lvert t_n\rvert$ | $run(t_n)$ | $run(t_n)/\lvert t_n\rvert$ | $run(t_n^2)$ | $run(t_n^3)$ | $run(t_n^k)/k\lvert t_n\rvert$ |
|---|---|---|---|---|---|---|
| 0 | 19 | 13 | 0.6842105263 | 29 | 44 | 0.7894736842 |
| 1 | 13 | 7 | 0.5384615385 | 19 | 30 | 0.8461538462 |
| 2 | 24 | 17 | 0.7083333333 | 39 | 60 | 0.8750000000 |
| 3 | 37 | 28 | 0.7567567568 | 62 | 95 | 0.8918918919 |
| 4 | 56 | 47 | 0.8392857143 | 99 | 150 | 0.9107142857 |
| 5 | 69 | 56 | 0.8115942029 | 120 | 183 | 0.9130434783 |
| 6 | 125 | 110 | 0.8800000000 | 227 | 343 | 0.9280000000 |
| 7 | 162 | 143 | 0.8827160494 | 295 | 446 | 0.9320987654 |
| 8 | 218 | 197 | 0.9036697248 | 402 | 606 | 0.9357798165 |
| 9 | 380 | 346 | 0.9105263158 | 704 | 1061 | 0.9394736842 |
| 10 | 505 | 467 | 0.9247524752 | 943 | 1418 | 0.9405940594 |
| 11 | 667 | 617 | 0.9250374813 | 1246 | 1874 | 0.9415292354 |
| 12 | 1172 | 1094 | 0.9334470990 | 2200 | 3305 | 0.9428327645 |
| 13 | 1552 | 1451 | 0.9349226804 | 2916 | 4380 | 0.9432989691 |
| 14 | 2057 | 1930 | 0.9382596014 | 3872 | 5813 | 0.9436071949 |
| 15 | 3609 | 3391 | 0.9395954558 | 6799 | 10206 | 0.9440288168 |
| 16 | 4781 | 4501 | 0.9414348463 | 9016 | 13530 | 0.9441539427 |
| 17 | 6333 | 5964 | 0.9417337755 | 11945 | 17925 | 0.9442602242 |
| 18 | 11114 | 10480 | 0.9429548317 | 20977 | 31473 | 0.9443944574 |
| 19 | 14723 | 13887 | 0.9432180941 | 27793 | 41698 | 0.9444406711 |
| 20 | 19504 | 18405 | 0.9436525841 | 36827 | 55248 | 0.9444729286 |
| 21 | 34227 | 32307 | 0.9439039355 | 64636 | 96964 | 0.9445174862 |
| 22 | 45341 | 42808 | 0.9441344479 | 85635 | 128461 | 0.9445314395 |
| 23 | 60064 | 56712 | 0.9441928609 | 113446 | 170179 | 0.9445424880 |
| 24 | 105405 | 99540 | 0.9443574783 | 199102 | 298663 | 0.9445567098 |
| 25 | 139632 | 131868 | 0.9443966999 | 263760 | 395651 | 0.9445614186 |
| 26 | 184973 | 174698 | 0.9444513524 | 349418 | 524137 | 0.9445648824 |
| 27 | 324605 | 306586 | 0.9444894564 | 613199 | 919811 | 0.9445695538 |
| 28 | 430010 | 406152 | 0.9445175694 | 812328 | 1218503 | 0.9445710565 |
| 29 | 569642 | 538042 | 0.9445265623 | 1076111 | 1614179 | 0.9445722050 |
| 30 | 999652 | 944219 | 0.9445477026 | 1888465 | 2832710 | 0.9445737117 |
| 31 | 1324257 | 1250831 | 0.9445530588 | 2501691 | 3752550 | 0.9445742027 |
| 32 | 1754267 | 1657010 | 0.9445597506 | 3314047 | 4971083 | 0.9445745716 |
| 33 | 3078524 | 2907866 | 0.9445649928 | 5815764 | 8723661 | 0.9445750626 |
| 34 | 4078176 | 3852116 | 0.9445683560 | 7704261 | 11556405 | 0.9445752219 |
| 35 | 5402433 | 5102974 | 0.9445696041 | 10205980 | 15308985 | 0.9445753423 |
| 36 | 9480609 | 8955120 | 0.9445722316 | 17910272 | 26865423 | 0.9445755014 |
| 37 | 12559133 | 11863017 | 0.9445729255 | 23726068 | 35589118 | 0.9445755531 |
| 38 | 16637309 | 15715165 | 0.9445737288 | 31430362 | 47145558 | 0.9445755921 |
| 39 | 29196442 | 27578212 | 0.9445744108 | 55156461 | 82734709 | 0.9445756438 |
| 40 | 38677051 | 36533368 | 0.9445748074 | 73066770 | 109600171 | 0.9445756606 |
| 41 | 51236184 | 48396417 | 0.9445749707 | 96792871 | 145189324 | 0.9445756733 |
| 42 | 89913235 | 84929820 | 0.9445752897 | N/A | N/A | N/A |
| 43 | 119109677 | 112508068 | 0.9445753765 | N/A | N/A | N/A |
| 44 | 157786728 | 149041473 | 0.9445754715 | N/A | N/A | N/A |

To make the analysis easier, we classify the strings of $\{t_n\}$ into the following three forms and we focus attention on $\{a_n\}$.

$$\begin{aligned} a_n &= t_{3m} = b_{n-1}c_{n-1}, \\ b_n &= t_{3m+1} = a_n a_{n-1}, \\ c_n &= t_{3m+2} = b_n b_{n-1}. \end{aligned}$$

By definition, we can get the closed form of $\{a_n\}$ as follows:

$$\begin{aligned} a_n &= b_{n-1}c_{n-1} \\ &= b_{n-1}b_{n-2}b_{n-1} \\ &= a_{n-1}a_{n-2}a_{n-2}a_{n-3}a_{n-1}a_{n-2}. \end{aligned}$$

So we will analyze the length of $\{a_{2n}\}$ in Section 4.1, and the number of runs in Section 4.2.

### 4.1 Length

At first we give the generating function of $|a_{2n}| = |t_{6n}|$.

**Lemma 1.** *Let $L_A(z)$ be the generating function of $|a_{2n}|$. $L_A(z)$ can be represented as follows:*

$$L_A(z) = \frac{-17z^2 + 65z - 19}{z^3 - 5z^2 + 10z - 1}.$$

*Proof.*

$$\begin{aligned} |a_k| &= |a_{k-1}a_{k-2}a_{k-2}a_{k-3}a_{k-1}a_{k-2}| \\ &= 2|a_{k-1}| + 3|a_{k-2}| + |a_{k-3}|. \end{aligned}$$

Let $g_n = |a_n|$,

$$\begin{aligned} g_0 &= |a_0| = 19, \\ g_1 &= |a_1| = 37, \\ g_2 &= |a_2| = 125, \\ g_n &= 2g_{n-1} + 3g_{n-2} + g_{n-3} \ (\ n > 2\ ). \end{aligned}$$

Therefore, we have general term of $g_n$ as follows:

$$g_n = 2g_{n-1} + 3g_{n-2} + g_{n-3} + 19_{[n=0]} - 1_{[n=1]} - 6_{[n=2]},$$

where the expression $m_{[n=k]}$ means the function such that the result is $m$ if $n = k$, and 0 if $n \neq k$.

Let $L(z)$ be the generating function of $g_n$. We have

$$\begin{aligned}
L(z) &= 2\sum_n g_{n-1}z^n + 3\sum_n g_{n-2}z^n + \sum_n g_{n-3}z^n \\
&\quad + \sum_n (19_{[n=0]} - 1_{[n=1]} - 6_{[n=2]})z^n \\
&= 2zL(z) + 3z^2L(z) + z^3L(z) + 19 - z - 6z^2 \\
&= \frac{6z^2+z-19}{z^3+3z^2+2z-1}.
\end{aligned}$$

By definition, $|a_{2n}| = |t_{6n}| = |t_{3(2n)}| = g_{2n}$,

$$\begin{aligned}
\sum_n g_{2n}z^{2n} &= \frac{1}{2}\left(L(z) + L(-z)\right) \\
&= \frac{1}{2}\left(\frac{6z^2+z-19}{z^3+3z^2+2z-1} + \frac{6z^2-z-19}{-z^3+3z^2-2z-1}\right) \\
&= \frac{1}{2}\left(\frac{-17z^4+65z^2-19}{z^6-5z^4+10z^2-1}\right).
\end{aligned}$$

Therefore, the generating function of $|a_{2n}|$ is as follows:

$$L_A(z) = \sum_n |a_{2n}|z^n = \sum_n g_{2n}z^n = \frac{1}{2}\left(\frac{-17z^2+65z-19}{z^3-5z^2+10z-1}\right).$$

□

To solve this generating function, we use the following theorem. If $A(z)$ is a power series $\sum_{k\geq 0} a_k z^k$, we will find it convenient to write $[z^n]A(z) = a_n$.

**Theorem 4 (Rational Expansion Theorem for Distinct Roots [13]).** *If $R(z) = P(z)/Q(z)$, where $Q(z) = q_0(1-\rho_1 z)\ldots(1-\rho_\ell z)$ and the numbers $(\rho_1, \ldots, \rho_\ell)$ are distinct, and if $P(z)$ is a polynomial of degree less than $\ell$, then*

$$[z^n]R(z) = a_1\rho_1^n + \cdots + a_\ell\rho_\ell^n, \text{ where } a_k = \frac{-\rho_k P(1/\rho_k)}{Q'(1/\rho_k)}.$$

Using this theorem, we will show the general term of $|a_n|$. Let $Q(z) = z^3 - 5z^2 + 10z - 1$ and $Q^R(z) = -z^3 + 10z^2 - 5z + 1$. Therefore $Q^R(z)$ is the "reflected" polynomial of $Q(z)$. Let $(\alpha, \beta, \gamma)$ be the roots of $Q^R(z)$. Therefore $Q^R(z) = (z-\alpha)(z-\beta)(z-\gamma)$, and $Q(z) = (1-\alpha z)(1-\beta z)(1-\gamma z)$. By Theorem 4, we have the general term of $|a_n|$ as follows.

**Theorem 5.** *$|a_n| = f(\alpha)\alpha^n + f(\beta)\beta^n + f(\gamma)\gamma^n$ for $f(x) = \frac{x\left(19x^2-65x+17\right)}{10x^2-10x+3}$, where $(\alpha, \beta, \gamma)$ are the roots of the equation $-z^3 + 10z^2 - 5z + 1 = 0$. The values* *of $\alpha$, $\beta$, and $\gamma$ are as follows:*

$$\alpha = \frac{10}{3} + \frac{1}{3}\sqrt[3]{\frac{1577}{2} - \frac{21\sqrt{69}}{2}} + \frac{1}{3}\sqrt[3]{\frac{1}{2}\left(1577 + 21\sqrt{69}\right)},$$

$$\beta = \frac{10}{3} - \frac{1}{6}\left(1 - i\sqrt{3}\right)\sqrt[3]{\frac{1577}{2} - \frac{21\sqrt{69}}{2}} - \frac{1}{6}\left(1 + i\sqrt{3}\right)\sqrt[3]{\frac{1}{2}\left(1577 + 21\sqrt{69}\right)},$$

$$\gamma = \frac{10}{3} - \frac{1}{6}\left(1 + i\sqrt{3}\right)\sqrt[3]{\frac{1577}{2} - \frac{21\sqrt{69}}{2}} - \frac{1}{6}\left(1 - i\sqrt{3}\right)\sqrt[3]{\frac{1}{2}\left(1577 + 21\sqrt{69}\right)}.$$

### 4.2 Number of Runs

Instead of trying to count the numbers of runs in the strings defined by the recurrence (1) only, we take a general approach here. We address ourselves to find general formulae which express the numbers of runs in strings defined by some recurrence, or equivalently, by some morphism.

Let $m$, $k$, $\gamma_1, \gamma_2 \ldots \gamma_k$ be integers such that $1 \le \gamma_j \le m$ for any $1 \le j \le k$, and $s_0, s_1, \ldots, s_{m-1} \in \Sigma^+$ be any nonempty strings. We consider a series of strings $\{s_n\}$ defined by the recurrence formula

$$s_n = s_{n-\gamma_1} s_{n-\gamma_2} \ldots s_{n-\gamma_k} \quad (n \ge m).$$

We pay our attentions to the quantity $\Delta_n = run(s_n) - \sum_{i=1}^{k} run(s_{n-\gamma_i})$. It is the difference between the number of newly created runs and the number of merged runs by the concatenation. Let $p$ be the least common multiple of the two integers $\gamma_1$ and $\gamma_k$. We observe that $\{\Delta_n\}$ is a mixture of $p$ arithmetic progressions with the same common difference, except initial several terms. More formally, we have the following conjecture.

**Lemma 2 (Conjecture).** *There exist integers $A$ and $n_0$ such that $\Delta_n = \Delta_{n-p} + A$ for any $n \ge n_0$.*

For example, in $\{a_n\}$ we have $\Delta_n = \Delta_{n-2} + 25$ for $n \ge 5$ (see Table 2). Unfortunately, we have not succeeded to give a formal proof to the conjecture at this time of writing. However, we have verified it for various instances and encountered no counter examples. Based on the conjecture, we have the following corollary, which is very useful to calculate the generating function.

**Corollary 1 (Based on conjecture Lemma 2).** *There exist integers $A, B_0, \ldots, B_{p-1}$, and $n_0$ such that*

$$run(s_{pn+i}) = \sum_{j=1}^{k}(run(s_{pn+i-\gamma_j})) + An + B_i,$$

*for any $n \ge n_0$.*

**Table 2.** The length of $\{a_n\}$ and number of runs in $\{a_n\}$.

| $n$ | $\lvert a_n \rvert$ | $run(a_n)$ | $\varDelta_n$ | $\varDelta_n - \varDelta_{n-2}$ |
|---|---|---|---|---|
| 0 | 19 | 13 | | |
| 1 | 37 | 28 | | |
| 2 | 125 | 110 | | |
| 3 | 380 | 346 | 29 | |
| 4 | 1172 | 1094 | 44 | |
| 5 | 3609 | 3391 | 55 | 26 |
| 6 | 11114 | 10480 | 70 | 26 |
| 7 | 34227 | 32307 | 80 | 25 |
| 8 | 105405 | 99540 | 95 | 25 |
| 9 | 324605 | 306586 | 105 | 25 |
| 10 | 999652 | 944219 | 120 | 25 |
| 11 | 3078524 | 2907866 | 130 | 25 |
| 12 | 9480609 | 8955120 | 145 | 25 |
| 13 | 29196442 | 27578212 | 155 | 25 |
| 14 | 89913235 | 84929820 | 170 | 25 |

Note that $a_n = a_{n-1}a_{n-2}a_{n-2}a_{n-3}a_{n-1}a_{n-2}$, from Corollary 1, we have $p = 2$ and the recurrence formula of $run(a_n)$ for large $n$ as follows:

$$\begin{aligned} run(a_{2n}) &= 2run(a_{2n-1}) + 3run(a_{2n-2}) + run(a_{2n-3}) + 25n - 5, \\ run(a_{2n+1}) &= 2run(a_{2n}) + 3run(a_{2n-1}) + run(a_{2n-2}) + 25n + 5. \end{aligned}$$

Let us consider the progression $\{r_n\}$ defined by

$$\begin{aligned} r_0 &= 15, \\ r_1 &= 27, \\ r_2 &= 110, \\ r_{2k} &= 2r_{2k-1} + 3r_{2k-2} + r_{2k-3} + 25k - 5 \quad (k \geq 2), \\ r_{2k+1} &= 2r_{2k} + 3r_{2k-1} + r_{2k-2} + 25k + 5 \quad (k \geq 1). \end{aligned}$$

We can see that $run(a_n) = r_n$ for any $n \geq 2$.

To analyze the asymptotic behavior of $run(a_n)$, we give the general term of $\{r_n\}$.

Let $X(z), Y(z)$ be the generating functions of $\{r_{2n}\}$ and $\{r_{2n+1}\}$:

$$\begin{aligned} X(z) &= \sum r_{2n} z^n, \\ Y(z) &= \sum r_{2n+1} z^n. \end{aligned}$$

Then,

$$
\begin{aligned}
X(z) &= 2zY(z) + 3zX(z) + z^2Y(z) + \tfrac{25z}{(1-z)^2} - \tfrac{5z}{1-z} - 15 + 9z, \\
Y(z) &= \ 2X(z) + 3zY(z) + zX(z) + \tfrac{25z}{(1-z)^2} + \tfrac{5z}{1-z} \ -3.
\end{aligned}
$$

To solve above simultaneous equations, we have

$$
X(z) = -\frac{19z^4 - 103z^2 + 164z^2 - 70z + 15}{(z-1)^2(z^3 - 5z + 10z - 1)}.
$$

Let $\alpha, \beta, \gamma$ are the roots of equation $-z^3 + 10z^2 - 5z + 1 = 0$. We have the general term $r_{2n}$ from $X(z)$ as follows:

$$
r_{2n} = g(\alpha)\alpha^n + g(\beta)\beta^n + g(\gamma)\gamma^n + O(n)
$$

where,

$$
g(x) = \frac{x\left(15x^4 - 70x^3 + 164x^2 - 103x + 19\right)}{12x^4 - 52x^3 + 6x^2 - 28x + 5}.
$$

Therefore we have the lower bounds of the maximal number of runs.

**Theorem 6 (Based on Conjecture Lemma 2).**

$$
\frac{\rho(n)}{n} \geq 0.94457571235.
$$

*Proof.*

$$
\begin{aligned}
\rho(n) &\geq \lim_{n\to\infty} \frac{run(a_{2n})}{|a_{2n}|} = \lim_{n\to\infty} \frac{r_{2n}}{|a_{2n}|} \\
&= \lim_{n\to\infty} \frac{g(\alpha)\alpha^n + g(\beta)\beta^n + g(\gamma)\gamma^n + O(n)}{f(\alpha)\alpha^n + f(\beta)\beta^n + f(\gamma)\gamma^n} \\
&= \frac{g(\alpha)\alpha^n}{f(\alpha)\alpha^n} \qquad (|\alpha| > |\beta| = |\gamma|) \\
&= \frac{\frac{\alpha\left(15\alpha^4 - 70\alpha^3 + 164\alpha^2 - 103\alpha + 19\right)}{12\alpha^4 - 52\alpha^3 + 6\alpha^2 - 28\alpha + 5}}{\frac{\alpha(19\alpha^2 - 65\alpha + 17)}{10\alpha^2 - 10\alpha + 3}} \\
&= \frac{\left(3 - 10\alpha + 10\alpha^2\right)\left(99 - 488\alpha + 889\alpha^2\right)}{\left(17 - 65\alpha + 19\alpha^2\right)\left(73 - 356\alpha + 683\alpha^2\right)} \\
&= \frac{7714 - 109145\sqrt[3]{\frac{2}{-27669823 + 9298929\sqrt{69}}} + \sqrt[3]{\frac{-27669823 + 9298929\sqrt{69}}{2}}}{8079} \\
&\approx 0.94457571235.
\end{aligned}
$$

□

## 5 Conclusion

In this paper, we showed a new series $\{t_n\}$ of run-rich strings defined by a simple recurrence formula, and we succeeded to improve the lower bound to 0.94457567 of the maximum number of runs in a string by using concrete string $t_{41}$. If we count the number of runs in a more longer strings $t_n^2$ and $t_n^3$ for $n > 41$, the bound can be improved further. Moreover, we gave a conjecture about the numbers of runs in the strings defined by any recurrence formula. Based on the conjecture, we evaluated the value $\lim_{n\to\infty} run(t_n)/|t_n|$ accurately, which yields the best lower bound so far. We are trying to give a proof of the conjecture.

Recently, Baturo et al. [6] derived an explicit formula for the number of runs in any standard Sturmian words. Moreover, they showed how to compute the number of runs in a standard Sturmian words in linear time with respect to the size of its compressed representation, that is, the recurrences describing the string. We are interested in extending it to a general strings described by any recurrences for further research.

## References


1. Kolpakov, R., Kucherov, G.: Finding maximal repetitions in a word in linear time. In: Proc. 40th Annual Symposium on Foundations of Computer Science (FOCS'99). (1999) 596–604
2. Rytter, W.: The number of runs in a string: Improved analysis of the linear upper bound. In: Proc. 23rd Annual Symposium on Theoretical Aspects of Computer Science (STACS 2006). Volume 3884 of LNCS. (2006) 184–195
3. Rytter, W.: The number of runs in a string. Inf. Comput. **205**(9) (2007) 1459–1469
4. Puglisi, S.J., Simpson, J., Smyth, W.F.: How many runs can a string contain? Theoretical Computer Science **401**(1–3) (2008) 165–171
5. Crochemore, M., Ilie, L.: Maximal repetitions in strings. J. Comput. Syst. Sci. **74** (2008) 796–807
6. Baturo, P., Piatkowski, M., Rytter, W.: The number of runs in Sturmian words. In: Proc. CIAA 2008. (2008) 252–261
7. Giraud, M.: Not so many runs in strings. In: Proc. LATA 2008. (2008) 245–252
8. Crochemore, M., Ilie, L., Tinta, L.: Towards a solution to the "runs" conjecture. In: Proc. CPM 2008. Volume 5029 of LNCS. (2008) 290–302
9. Franěk, F., Yang, Q.: An asymptotic lower bound for the maximal-number-of-runs function. In: Proc. Prague Stringology Conference (PSC'06). (2006) 3–8
10. Franěk, F., Simpson, R., Smyth, W.: The maximum number of runs in a string. In: Proc. 14th Australasian Workshop on Combinatorial Algorithms (AWOCA2003). (2003) 26–35
11. Matsubara, W., Kusano, K., Ishino, A., Bannai, H., Shinohara, A.: New lower bounds for the maximum number of runs in a string. In: Proc. The Prague Stringology Conference 2008. (2008) 140–145
12. Crochemore, M., Ilie, L., Smyth, W.F.: A simple algorithm for computing the lempel ziv factorization. In: Proc. DCC 2008. (2008) 482–488
13. Graham, R., Knuth, D., Patashnik, O.: Concrete mathematics Second Edition. Addison-Wesley (1995)