# On the appearance of seeds in words

Manolis Christodoulakis
Department of Electrical and Computer Engineering,
University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus
christodoulakis.manolis@ucy.ac.cy

Michalis Christou
Department of Informatics, King's College London,
Strand, London WC2R 2LS, UK
michalis.christou@kcl.ac.uk

Maxime Crochemore
Department of Informatics, King's College London,
Strand, London WC2R 2LS, UK
Université Paris-Est, France
Maxime.Crochemore@kcl.ac.uk

Costas S. Iliopoulos
Department of Informatics, King's College London,
Strand, London WC2R 2LS, UK
Curtin University, Digital Ecosystems & Business Intelligence
Institute, Center for stringology & Applications, Australia
csi@dcs.kcl.ac.uk

**Abstract**

A seed of a word $x$ is a cover of a superword of $x$. In this paper we study the frequency of appearance of seeds in words. We give bounds for the average number of seeds in a word and we investigate the maximum number of distinct seeds that can appear in a word. More precisely, we prove that a word has $O(n)$ seeds on average and that the maximum number of distinct seeds in a word is between $\frac{1}{6}n^2 + o(n^2)$ and $\frac{1}{4}n^2 + o(n^2)$, and we reveal some properties of an extremal word for the last case.

JCMCC **95** (2015), pp. 147-160

# 1 Introduction

Words, also called strings, appear in many areas of Mathematics and Computer Science as well as in several interdisciplinary areas such as pattern matching, data compression and bioinformatics (see [19, 20]). Periodicity in words is a fundamental key to the understanding of their structure. Some fundamental periodicities in a word include the runs and powers occurring in it, such as squares and cubes. Apart from algorithmic interest, in the last years a lot of research has been done on bounds on the maximal number of distinct periodicities in a word. These bounds are essential elements of the analysis of some algorithms on words.

The "runs" conjecture, proposed by Kolpakov and Kucherov [16], states that the number of maximal periodicities (runs) in a word of length $n$, is at most $n$. The first upper bound given was $5n$ [24], which was improved to $3.48n$ [23], to $3.44n$ [25], to $1.6n$ [6], to $1.52n$ for binary words [11] and finally to $1.048n$ [7]. Regarding the lower bound a first estimate of $0.927n$ was given in [10] and improved further to $0.944542n$ [22], to $0.94457567n$ [21] and eventually to $0.944575712n$ [26]. The exact bounds are still unknown. The maximal number of cubic runs in a word was found to be between $0.5n$ and $0.406n$ [8].

Regarding the maximal number of squares in a word, Fraenkel and Simpson showed that it is at most $2n$ [9], a result proved later in a simpler way by Ilie [13] and improved to $2n - \Theta(logn)$ [14]. The same number for partial words with one hole was found to be at most $\frac{7n}{2}$ [1]. Kucherov et al. showed that a binary word must contain at least $0.55080n$ square occurrences [18]. The maximal number of cubes in a word has been shown to be between $\frac{n}{2}$ and $\frac{4n}{5}$ [17]. In a more general scenario, Crochemore et al. [4, 5] proved a $\Theta(n \log n)$ bound on the maximal number of occurrences of primitively rooted $k$th powers in a word.

The concept of quasiperiodicity is a generalization of the notion of periodicity. In a periodic repetition the occurrences of the single periods do not overlap. In contrast, the quasiperiods of a quasiperiodic word may overlap. We call a border $u$ of a non-empty word $x$ a cover of $x$, if every letter of $x$ is within some occurrence of $u$ in $x$. Seeds are regularities of words strongly related to the notion of cover, as a seed is a cover of a superword of the word.

There is not much known about bounds on the number of quasiperiodicities in words. Some scientific work has been done on that direction, mainly related to Fibonacci words, e.g. identification of all covers of a circular Fibonacci word [2, 15], identification of all maximal quasiperiodicities in Fibonacci words [12] and identification of all covers and seeds of a

Fibonacci word [2].

In this paper we study the frequency of appearance of seeds in words. Using combinatorial properties of words we prove that a word has $O(n)$ seeds on average. Furthermore we show that the maximum number of distinct seeds in a word is between $\frac{1}{6}n^2 + o(n^2)$ and $\frac{1}{4}n^2 + o(n^2)$ and we reveal some properties for the structure of an extremal word for the last case. It is important to note that we restrict seeds to be factors of the given word.

## 2 Definitions

Throughout this paper we consider a word $x$ of length $|x| = n$, $n \geq 0$, on a fixed alphabet $\Sigma = \{a_1, a_2, \ldots, a_\sigma\}$ ($\sigma = |\Sigma|$). Whenever $x$ is a non-empty word it is represented as $x[1 \mathinner{..} n]$. A word $w$ is a *factor* of $x$ if $x = uwv$ for two words $u$ and $v$. It is a *prefix* of $x$ if $u$ is empty and a *suffix* of $x$ if $v$ is empty. A *proper* factor of $x$ is a factor which is not equal to $x$ itself; *proper* prefixes, suffixes and borders are defined similarly.

A word $u$ is a *border* of $x$ if $u$ is both a proper prefix and a suffix of $x$. *The border* of $x$, denoted by *border*$(x)$, is the length of the longest border of $x$. A non-empty word $u$ is a *period* of $x$ if $x$ is a prefix of $u^k$ for some positive integer $k$ ($x^k$ is the concatenation of $k$ copies of $x$), or equivalently if $x$ is a prefix of $ux$. The length of $u$ is also called a period (or an integer period) of $x$. *The period* of $x$, denoted by *period*$(x)$, is the length of the shortest period of $x$. The exponent of $x$ is the ratio $|x|/$*period*$(x)$.

A word $w$ is a *square* of $x$ if it is a factor of $x$ and $w = yy$ for some non empty word $y$. A word $w$ is a *cube* of $x$ if it is a factor of $x$ and $w = yyy$ for some non empty word $y$. More generally a word $w$ is an *$r$-power* of $x$ if it is a factor of $x$ and $w = y^r$, for some non empty word $y$, and $r \in \{2, 3, 4, \ldots\}$. A run is a maximal (non-extendable) occurrence of a repetition of rational exponent at least two. That means, the factor $x[i \mathinner{..} j]$ is a run if it has the following three properties:

- $x[i \mathinner{..} j]$ has period $p$ and $j - i + 1 \geq 2p$
- $x[i-1] \neq x[i+p-1]$ (if $x[i-1]$ is defined), $x[j+1] \neq x[j-p+1]$ (if $x[j+1]$ is defined)
- $x[i..i+p-1]$ is primitive, that is, it is not a proper integer power (2 or larger) of another word.

For two words $u = u[1 \mathinner{..} m]$ and $v = v[1 \mathinner{..} n]$ where a suffix of $u$ equals a prefix of $v$, $u[m-\ell+1\,..\,m] = v[1\,..\,\ell]$ for some $1 \le \ell \le m$, the word $u[1\,..\,m]v[\ell+1\,..\,n] = u[1\,..\,m-\ell]v[1\,..\,n]$ is called a *superposition* of $u$ and $v$ with an *overlap* of length $\ell$.

A word $w$ is a *quasiperiodic square* of $x$ if it is a factor of $x$ and $w = yv = uy$, where $y$, $v$ and $u$ are non empty words and $|y| > |v|$. In this case, the factor $y$ is called an *overlapping factor* of $x$.

A word $y$ of length $m$ is a *cover* of $x$ if both $m < n$ and there exists a set of positions $P \subseteq \{1,\ldots,n-m+1\}$ that satisfies both $x[i\,..\,i+m-1] = y$ for all $i \in P$ and $\bigcup_{i\in P}\{i,\ldots,i+m-1\} = \{1,\ldots,n\}$. A word $v$ is a *seed* of $x$, if it is a cover of a superword of $x$, where a superword of $x$ is a word of form $uxv$ and $u, v$ are possibly empty words. A *left seed* of a word $x$ is a prefix of $x$ that is a cover of a superword of $x$ of the form $xv$, where $v$ is a possibly empty word. Similarly, a *right seed* of a word $x$ is a suffix of $x$ that is a cover of a superword of $x$ of the form $vx$, where $v$ is a possibly empty word.

Figure 1: *aba* is the shortest seed of *abaababaabaab*

The following example shows all left seeds, right seeds, covers and seeds of the word $F_6$ = **abaababaabaab** and Figure 1 illustrates that *aba* is the shortest seed of $F_6$.

**Example 1.**

| | |
|---|---|
| Covers of $F_6$ | abaab, abaababaabaab |
| Left seeds of $F_6$ | aba, abaab, abaaba, abaababa, abaababaa, abaababaab, abaababaaba, abaababaabaa, abaababaabaab |
| Right seeds of $F_6$ | abaab, abaabaab, babaabaab, ababaabaab, aababaabaab, baababaabaab, abaababaabaab |
| Seeds of $F_6$ | aba, abaab, abaaba, abaababa, abaababaa, abaababaab, abaababaaba, abaababaabaa, abaababaabaab |

# 3 Average case

In this section we study the behaviour of the average number of seeds in a word of $\Sigma^n$, the set of words of length $n$. This number is given by the expected value of the number of seeds when we consider a word $x = x[1\,..\,n]$ with all letters of $x$ drawn independently from $\Sigma = \{a_1, a_2, \ldots, a_\sigma\}$ with equal probability $\frac{1}{\sigma}$ for each letter.

We show that a word of length $n$ has $O(n)$ seeds on average by using some combinatorial properties and series relations. However, before that, we need to prove the following basic lemma:

**Lemma 1.** *A seed of length $\ell$ which appears only once in $x$ implies the appearance of either a periodic prefix or a periodic suffix of $x$ of length at least $\frac{n+\ell}{2}$ and period at most $\ell$.*

*Proof.* Let $s = x[i\,..\,i+\ell-1] \neq x$ be a seed of $x$ occurring only once in $x$. Therefore, $x = usv$ with $0 \leq |u| < \ell$ and $0 \leq |v| < \ell$, where not both $u$ and $v$ are empty strings.

Let's assume first that both $|v| > 0$ and $|u| > 0$. Then a prefix of $s$ must occur after position $i$, say at $i+j$ $(1 \leq j \leq \ell)$, and consequently $x$ can be written as $x = u(s[1\,..\,j])^{\frac{n-|u|}{j}}$. Hence the suffix of $x$ of length $n-|u|$ is periodic with period at most $\ell$. Similarly, a suffix of $s$ must end before position $i+\ell-1$, yielding a periodic prefix of $x$ of length $n-|v|$ and period at most $\ell$. As $|u|+|v| = n-\ell$ either $n-|u|$ or $n-|v|$ is at least $\frac{n+\ell}{2}$.

If only $|u| = 0$ or only $|v| = 0$ then $x$ has period at most $\ell$ and our requirements are met. □

In a similar manner, we can prove the following lemma.

**Lemma 2.** *A seed $y = x[i\,..\,i+\ell-1]$ of length $\ell$, where $1 \leq i \leq n-\ell+1$ and $1 \leq \ell \leq n-1$, which appears at least twice in $x$ implies the appearance of a square of form $x[i\,..\,i+2\ell-1] = yy$ or $x[i-\ell\,..\,i+\ell-1] = yy$ or the appearance of a quasiperiodic square of form $x[i\,..\,i+\ell+k-1] = yv$ or $x[i-k\,..\,i+\ell-1] = vy$, where $1 \leq k \leq \ell-1$ and $v$ is a substring of $x$ such that $|v| = k$ , in $x$.*

*Proof.* As there exist at least two occurrences of $y$ in $x$ and $y$ is a cover of a superstring of $x$ then there should be a $y$ starting in position $j \in \{i+1, \ldots, i+\ell\}$ or ending in position $j \in \{i-1, \ldots, i+\ell-2\}$. Those two occurrences form the required square (if they are consecutive) or quasiperiodic square (if they overlap). □

The following lemma regarding the polylogarithm function will also be required to prove further results. Basically the lemma evaluates the polylogarithm function at $-1$.

**Lemma 3.** *For $z \in \mathbb{C}$ such that $|z| < 1$, $Li_{-1}(z) = \sum_{k=1}^{+\infty} kz^k = \frac{z}{(1-z)^2}$.*

Using the above lemmas, we are now able to prove the main result of this section.

**Theorem 4.** *On average a word of length $n$ has $O(n)$ seeds.*

*Proof.* Let $x$ be a word of length $n$, with its letters drawn independently from $\Sigma = \{a_1, a_2, \ldots, a_\sigma\}$ with a constant probability distribution $(\frac{1}{\sigma}, \frac{1}{\sigma}, \ldots, \frac{1}{\sigma})$.

In what follows we are using "$E$" for the expectation, i.e. the weighted average of the possible values of a variable, and "$P$" for the probability of an event.

$$\begin{aligned} E(\text{number of seeds in } x) &= E(\#\text{ of seeds appearing only once in } x) \\ &\quad + E(\#\text{ of seeds appearing at least twice in } x) \end{aligned}$$

First we find the expected value for the number of seeds appearing only once in length $n$ words (case 1) and then the expected value for the number of seeds appearing more than once (case 2).

**Case 1:**

As in Lemma 1 we get:

$$\begin{aligned} &E(\text{number of Case 1 seeds}) \\ &= \sum_{\ell=1}^{n} \sum_{i=1}^{n-\ell+1} P(x[i\,..\,i+\ell-1] \text{ is a Case 1 seed}) \\ &\le \sum_{\ell=1}^{n} \sum_{i=1}^{n-\ell+1} P(x \text{ has a periodic prefix or a periodic suffix of length} \\ &\text{at least } \frac{n+\ell}{2} \text{ and period at most } \ell) \\ &\le 2\sum_{\ell=1}^{n} \sum_{i=1}^{n} P(x \text{ has a periodic prefix of length at least } \frac{n+\ell}{2} \text{ and} \\ &\text{period at most } \ell) \end{aligned}$$

$$\leq 2n\sum_{\ell=1}^{n}\sum_{k=1}^{\ell}\frac{\sigma^k}{\sigma^{\frac{n+\ell}{2}}} \leq \frac{2n}{\sigma^{\frac{n}{2}}}\sum_{\ell=1}^{n}\frac{1}{\sigma^{\frac{\ell}{2}}}\frac{\sigma(\sigma^\ell-1)}{\sigma-1} \leq \frac{2\sigma n}{(\sigma-1)\sigma^{\frac{n}{2}}}\sum_{\ell=1}^{n}(\sigma^{\frac{\ell}{2}}-\sigma-\frac{\ell}{2})$$

$$\leq \frac{2\sigma n}{(\sigma-1)\sigma^{\frac{n}{2}}}\left(\frac{\sigma^{\frac{1}{2}}(\sigma^{\frac{n}{2}}-1)}{\sigma^{\frac{1}{2}}-1}-\frac{\sigma^{-\frac{1}{2}}(\sigma^{-\frac{n}{2}}-1)}{\sigma^{-\frac{1}{2}}-1}\right)$$

$$= \frac{2\sigma^{\frac{3}{2}}}{(\sigma-1)(\sqrt{\sigma}-1)}n+o(n)$$

**Case 2:**

As in Lemma 2 we get:

$$E(\text{number of Case 2 seeds}) = \sum_{\ell=1}^{n}\sum_{i=1}^{n-\ell+1} P(x[i\mathinner{.\,.}j] \text{ is a Case 2 seed})$$

$$\leq \sum_{\ell=1}^{n}\sum_{i=1}^{n-\ell+1}(\sum_{k=1}^{\ell} P(x[i\mathinner{.\,.}i+\ell-1+k] \text{ is a quasiperiodic square})$$

$$+ P(x[i\mathinner{.\,.}i+2\ell-1] \text{ is a square}) + \sum_{k=1}^{\ell} P(x[i-k\mathinner{.\,.}i+\ell-1] \text{ is a}$$

$$\text{quasiperiodic square}) + P(x[i-\ell\mathinner{.\,.}i+\ell-1] \text{ is a square}))$$

$$\leq 2\sum_{\ell=1}^{n}\sum_{i=1}^{n-\ell+1}\sum_{k=1}^{\ell-1}\frac{\sigma^k}{\sigma^{\ell+k}} + 2\sum_{\ell=1}^{n-\ell+1}\sum_{i=1}^{n}\frac{\sigma^\ell}{\sigma^{2\ell}} = 2\sum_{\ell=1}^{n}\sum_{i=1}^{n-\ell+1}\sum_{k=1}^{\ell}\frac{\sigma^k}{\sigma^{\ell+k}}$$

$$= 2\sum_{\ell=1}^{n}\sum_{i=1}^{n-\ell+1}\sum_{k=1}^{\ell}\frac{1}{\sigma^\ell} = 2\sum_{\ell=1}^{n}\frac{n-\ell+1}{\sigma^\ell} = 2(n+1)\sum_{\ell=1}^{n}\frac{1}{\sigma^\ell} - 2\sum_{\ell=1}^{n}\frac{\ell}{\sigma^\ell}$$

$$= 2(n+1)\frac{(1-\sigma^{-n})}{\sigma-1} - 2\sum_{\ell=1}^{n}\frac{\ell}{\sigma^\ell}$$

Lemma 3 suggests that the last series are bounded by constant terms, thus proving the theorem. □

# 4 Maximum number of distinct seeds in a word

It is easy to see that there exist words in which every factor is a seed, e.g. $a^n$. In this section we are investigating how many distinct seeds can appear in a word. We denote the maximum number of distinct seeds in a word of length $n$ by $Seeds(n)$.

## 4.1 Lower Bound

Fibonacci words provide us with a first lower bound.

**Lemma 5.** *There exists an infinite family of words for which:*

$$Seeds(n) \geq \frac{\phi^2+1}{2\phi^6}n^2 + o(n^2),$$

*where $\phi$ is the golden ratio.*

*Proof.* In [2, 3] it was shown that for Fibonacci words it holds that:

$$\lim_{n\to+\infty} \frac{\text{Seeds}(F_n)}{|F_n|^2} = \frac{\phi^2+1}{2\phi^6} = 0.100813061875578\ldots \qquad \square$$

Next we show that periodic words are quite rich in word regularities. In particular, words that are squares have more seeds than Fibonacci words.

**Lemma 6.** $Seeds(n) \geq \frac{n^2}{8} + o(n^2)$.

*Proof.* When $n$ is even we consider the word $(a_1a_2\ldots a_{\frac{n}{2}})^2$. Obviously every factor of length greater than $\frac{n}{2}$ is a seed of the word. There are $\frac{n}{2}$ such factors starting from the first position in the word, $\frac{n}{2}-1$ such factors starting from the second position of the word and so on. Overall:

$$\sum_{i=1}^{\frac{n}{2}} i = \frac{\frac{n}{2}(\frac{n}{2}+1)}{2} = \frac{n^2}{8} + \frac{n}{4}$$

Similarly, when $n$ is odd we consider the word $(a_1a_2\ldots a_{\frac{n-1}{2}})^2 a_{\frac{n+1}{2}}$ which yields $\frac{n^2}{8} + \frac{n-1}{2}$ distinct seeds. $\square$

In the following lemma, we prove that among all words having similar structure to the squares considered in the proof of the previous lemma (i.e. words of form $(a_1a_2\ldots a_{|p|})^c = p^c$, where $c|p| = n$ and $1 \leq |p| < n$), cubes or words close to being cubes (i.e. words in which $|p| = \frac{n}{3} + o(n)$) achieve maximum number of distinct seeds.

**Lemma 7.** $Seeds(n) \geq \frac{1}{6}n^2 + o(n^2)$.

*Proof.* We consider the word $(a_1a_2\ldots a_{|p|})^c = p^c$, where $c|p| = n$ and $1 \leq |p| < n$. Obviously every factor of length greater than $|p|$ or equal to $|p|$ is a seed of the word. There are $n - |p| + 1$ such factors starting from the first position in the word, $n - |p|$ such factors starting from the second position of the word, etc. Due to factors repeating, we only consider factors starting from the first period of the word. We distinguish two cases, according to $|p|$.

For $\frac{n}{2} \leq |p| < n$ we get the following number of distinct seeds:

$$\sum_{i=1}^{n-|p|+1} i = \frac{(n-|p|+1)(n-|p|+2)}{2} = \frac{(n-\frac{n}{c}+1)(n-\frac{n}{c}+2)}{2}$$
$$= \frac{(1-\frac{1}{c})^2}{2} n^2 + o(n^2)$$

Having assumed that $\frac{n}{2} \leq |p| < n$, it follows that $1 < c \leq 2$ and the above expression maximizes for $c = 2$, giving $\frac{1}{8}n^2 + o(n^2)$ different seeds.

For $1 \leq |p| < \frac{n}{2}$ we get the following number of distinct seeds:

$$\sum_{i=n-2|p|+2}^{n-|p|+1} i = \frac{(n-|p|+1)(n-|p|+2)}{2} - \frac{(n-2|p|+1)(n-2|p|+2)}{2}$$
$$= \frac{(n-\frac{n}{c}+1)(n-\frac{n}{c}+2)}{2} - \frac{(n-2\frac{n}{c}+1)(n-2\frac{n}{c}+2)}{2}$$
$$= \frac{(1-\frac{1}{c})^2 - (1-\frac{2}{c})^2}{2} n^2 + o(n^2) = \frac{\frac{2}{c} - \frac{3}{c^2}}{2} n^2 + o(n^2)$$

As $2 < c \leq n$ the above expression maximizes for $c = 3$, giving $\frac{1}{6}n^2 + o(n^2)$ different seeds. Even when $c = 3$ cannot be achieved, choosing $p = \lfloor \frac{n}{3} \rfloor$ gives the required bound. □

## 4.2 Upper Bound

A first upper bound is given using the restriction that a seed must be a factor of the word.

**Lemma 8.** *The number of distinct seeds of a non-empty word $x$ is at most $\frac{n(n+1)}{2}$, where $|x| = n$.*

*Proof.* The number of non-empty factors of $x$ is $n + \binom{n}{2} = \frac{n(n+1)}{2}$. □

In the following lemma we prove a better upper bound using more combinatorial properties.

**Lemma 9.** *$Seeds(n) \leq \frac{1}{4}n^2 + o(n^2)$.*

*Proof.* We first consider $n$ to be even. A seed of length $\ell$, $n/2+1 \leq \ell \leq n$, appears at least once in $x$ (as it is a factor of $x$). Therefore, $x$ has at most $n - \ell + 1$ seeds of length $\ell$. Overall:

$$\sum_{\ell=n/2+1}^{n} (n - \ell + 1) < \sum_{i=1}^{n/2} i = \frac{1}{8}n^2 + o(n^2)$$

A seed of length $\ell$, $1 \leq \ell \leq n/2$, has at least one starting position in $x[1\,..\,\ell]$ (as it covers $x$). Therefore, $x$ has at most $\ell$ different seeds of length $\ell$. Overall:

$$\sum_{\ell=1}^{n/2} \ell = \frac{1}{8}n^2 + o(n^2)$$

By summing the two expressions we get $\frac{1}{4}n^2 + o(n^2)$ as an upper bound. The proof for odd $n$ is similar. □

## 4.3 Structure of the extremal word

The bounds revealed in the previous subsections help us to get some information for a word that maximizes $Seeds(n)$. In particular, we are able to restrict the length of repeating factors and the period and exponent of a run in such a word.

**Theorem 10.** *A word $x$ of length $n$ having the maximum number of seeds $Seeds(n)$ contains no repeating factors of length $cn + o(n)$, where*

$$\sqrt{\frac{2}{3}} < c \leq 1.$$

*Proof.* For large $n$, a repeating factor of length $cn$ would mean more than $\frac{1}{3}n^2$ repeats (the subwords of the factor) and hence less than $\frac{1}{2}n^2 - \frac{1}{3}n^2 = \frac{1}{6}n^2$ candidate seeds, contradicting Lemma 7. □

**Theorem 11.** *There are no runs of period $p$ and exponent $c$ in a word achieving the upper bound for which $\frac{1}{2}c^2p^2 - \frac{1}{2}cp + \frac{3}{2}p^2 > \frac{1}{3}n^2$.*

*Proof.* A run of period $p$ and exponent $c$ in $x$ gives at least

$$\frac{1}{2}(cp)(cp+1) - \sum_{i=cp-2p+2}^{cp-p+1} i$$

repeating factors. The dominating terms of this expression are

$$\frac{1}{2}c^2p^2 - \frac{1}{2}cp + \frac{3}{2}p^2.$$

Therefore, as of Theorem 7, for large $n$ we have:

$$\frac{1}{2}n^2 - \left(\frac{1}{2}c^2p^2 - \frac{1}{2}cp + \frac{3}{2}p^2\right) \geq \frac{1}{6}n^2$$

□

# 5 Conclusion

In this paper we have studied the frequency of appearance of seeds in words. We have given some bounds for the average number of seeds in a word and we have investigated the maximum number of distinct seeds that can appear in a word. More precisely, we have proved that a word has $O(n)$ seeds on average. We have also shown that the maximum number of distinct seeds in a word is between $\frac{1}{6}n^2 + o(n^2)$ and $\frac{1}{4}n^2 + o(n^2)$ and we have revealed some properties achieved by an extremal word for the last case. It is important to note that we have restricted seeds to be factors of the given word. We conclude by showing the existing best bounds for the maximum number of distinct regularities in a word in the following table:

| **Regularity** | **Maximum number in a word** |
|---|---|
| Runs | Between $1.048n$ [7] and $0.944575712n$ [26] |
| Cubic runs | Between $0.5n$ and $0.406n$ [8] |
| Squares | Between $n - o(n)$ [9] and $2n - \Theta(logn)$ [14] |
| Cubes | Between $\frac{n}{2}$ and $\frac{4n}{5}$ [17] |
| Seeds | Between $\frac{n^2}{6} + o(n^2)$ and $\frac{n^2}{4} + o(n^2)$ |

# References


[1] F. Blanchet-Sadri and R. Mercas. A note on the number of squares in a partial word with one hole. *RAIRO- Theoretical Informatics and Applications*, 43(4):767–774, 2009.

[2] M. Christou, M. Crochemore, and C. S. Iliopoulos. Quasiperiodicities in Fibonacci strings. In *Local Proceedings of International Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM)*, 2012.

[3] M. Christou, M. Crochemore, and C. S. Iliopoulos. Quasiperiodicities in Fibonacci strings. *Ars Combinatoria*, 2012. (accepted).

[4] M. Crochemore, S. Fazekas, C. Iliopoulos, and I. Jayasekera. Bounds on powers in strings. In M. Ito and M. Toyama, editors, *Developments in Language Theory*, volume 5257 of *Lecture Notes in Computer Science*, pages 206–215. Springer, 2008.

[5] M. Crochemore, S. Fazekas, C. Iliopoulos, and I. Jayasekera. Number of occurrences of powers in strings. *International Journal of Foundations of Computer Science*, 21(4):535–547, 2010.

[6] M. Crochemore, L. Ilie, and L. Tinta. Towards a solution to the runs conjecture. In P. Ferragina and G. M. Landau, editors, *Combinatorial Pattern Matching*, volume 5029 of *Lecture Notes in Computer Science*, pages 290–302. Springer, 2008.

[7] M. Crochemore, L. Ilie, and L. Tinta. The "runs" conjecture. *Theoretical Computer Science*, 412(27):2931–2941, 2011.

[8] M. Crochemore, C. Iliopoulos, M. Kubica, J. Radoszewski, W. Rytter, and T. Waleń. On the maximal number of cubic runs in a string. In A. H. Dediu, H. Fernau, and C. Martín-Vide, editors, *Language and Automata Theory and Applications*, volume 6031 of *Lecture Notes in Computer Science*, 2010.

[9] A. Fraenkel and J. Simpson. How many squares can a string contain? *Journal of Combinatorial Theory, Series A*, 82(1):112–120, 1998.

[10] F. Franek, Q. Yang, and J. Holub. An asymptotic lower bound for the maximal number of runs in a string. *International Journal of Foundations of Computer Science*, 19(1):195–203, 2008.

[11] M. Giraud. Not so many runs in strings. In C. Martín-Vide, F. Otto, and H. Fernau, editors, *Language and Automata Theory and Applications*, volume 5196 of *Lecture Notes in Computer Science*. Springer, 2008.

[12] R. Groult and G. Richomme. Optimality of some algorithms to detect quasiperiodicities. *Theoretical Computer Science*, 411(34-36):3110–3122, 2010.

[13] 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):163–164, 2005.

[14] L. Ilie. A note on the number of squares in a word. *Theoretical Computer Science*, 380(3):373–376, 2007.

[15] C. Iliopoulos, D. Moore, and W. Smyth. The covers of a circular Fibonacci string. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 26:227–236, 1998.

[16] R. Kolpakov and G. Kucherov. Finding maximal repetitions in a word in linear time. In *Symposium on Foundations of Computer Science-FOCS*, volume 99, pages 596–604, 1999.

[17] M. Kubica, J. Radoszewski, W. Rytter, and T. Waleń. On the maximal number of cubic subwords in a string. In J. Fiala, J. Kratochvíl, and M. Miller, editors, *International Workshop on Combinatorial Algorithms*, volume 5874 of *Lecture Notes in Computer Science*. Springer, 2009.

[18] G. Kucherov, P. Ochem, and M. Rao. How many square occurrences must a binary sequence contain? *Journal of Combinatorics*, 10(1):12, 2003.

[19] M. Lothaire, editor. *Algebraic Combinatorics on Words*. Cambridge University Press, 2001.

[20] M. Lothaire, editor. *Applied Combinatorics on Words*. Cambridge University Press, 2005.

[21] W. Matsubara, K. Kusano, H. Bannai, and A. Shinohara. A series of run-rich strings. In A. H. Dediu, A.-M. Ionescu, and C. Martín-Vide, editors, *Language and Automata Theory and Applications*, volume 5457 of *Lecture Notes in Computer Science*. Springer, 2009.

[22] W. Matsubara, K. Kusano, A. Ishino, H. Bannai, and A. Shinohara. New lower bounds for the maximum number of runs in a string. In *Prague Stringology Conference*, volume 2008, pages 140–145, 2008.

[23] S. Puglisi, J. Simpson, and W. Smyth. How many runs can a string contain? *Theoretical Computer Science*, 401(1-3):165–171, 2008.

[24] W. Rytter. The number of runs in a string: Improved analysis of the linear upper bound. In B. Durand and W. Thomas, editors, *STACS*, volume 3884 of *Lecture Notes in Computer Science*. Springer, 2006.

[25] W. Rytter. The number of runs in a string. *Information and Computation*, 205(9):1459–1469, 2007.

[26] J. Simpson. Modified padovan words and the maximum number of runs in a word. *Australasian Journal of Combinatorics*, 46:129–145, 2010.