Quasiperiodicities in Fibonacci strings

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

Maxime Crochemore
King's College London, London WC2R 2LS, UK
Université Paris-Est, France
maxime.crochemore@kcl.ac.uk

Costas Iliopoulos
King's College London, London WC2R 2LS, UK
Digital Ecosystems & Business Intelligence Institute
Curtin University, GPO Box U1987, Perth WA 6845, Australia
csi@dcs.kcl.ac.uk

Abstract

We consider the problem of finding quasiperiodicities in Fibonacci strings. A factor u of a string y is a *cover* of y if every letter of y falls within some occurrence of u in y. A string v is a *seed* of y, if it is a cover of a superstring of y. A left seed of a string y is a prefix of y that it is a cover of a superstring of y. Similarly a right seed of a string y is a suffix of y that it is a cover of a superstring of y. In this paper, we present some interesting results regarding quasiperiodicities in Fibonacci strings, we identify all covers, left/right seeds and seeds of a Fibonacci string and all covers of a circular Fibonacci string.

Introduction

The notion of periodicity in strings is well studied in many fields like combinatorics on strings, pattern matching, data compression and automata theory (see [16, 17]), because it is of paramount importance in several applications, not to talk about its theoretical aspects.

The concept of quasiperiodicity is a generalization of the notion of periodicity, and was defined by Apostolico and Ehrenfeucht in [1]. In a periodic repetition the occurrences of the single periods do not overlap. In contrast, the quasiperiods of a quasiperiodic string may overlap. We call a factor u of a nonempty string y a cover of y, if every letter of y is within some occurrence of u in y. Note that we consider the *aligned covers*, where the cover u of y needs to be a border (i.e. a prefix and a suffix) of y. *Seeds* are regularities of strings strongly related to the notion of cover, as a seed is a cover of a superstring of the string. They were first defined and studied by Iliopoulos, Moore and Park [13]. A left seed of a string y, firstly defined in [4], is a prefix of y that is a cover of a superstring of y. Similarly a right seed of a string y, also firstly defined in [4], is a suffix of y that is a cover of a superstring of y.

A fundamental problem is to find all covers of a string. A linear time algorithm was given by Moore and Smyth [18], Li and Smyth [15] (this algorithm gives also all the covers for every prefix of the string) and an $O(\log(\log(|y|)))$ work optimal parallel algorithm was given later by Iliopoulos and Park [12]. The corresponding problem on seeds is harder, the fastest and only algorithm was by Iliopoulos, Moore and Park [13], running in $O(|y| \log |y|)$, until recently Kociumaka et al gave a linear time algorithm [14].

Fibonacci strings are important in many concepts [2] and are often cited as a worst case example for many string algorithms. Over the years much scientific work has been done on them, e.g. locating all factors of a Fibonacci string [6], characterizing all squares of a Fibonacci string [9, 10], identifying all covers of a circular Fibonacci string [11], identifying all borders of a Fibonacci string [7], finding palindromes of a Fibonacci string [8], etc. Some research has also been extended to Tribonacci strings [19, 20, 21].

In this paper we are presenting results from our previous work on Fibonacci strings [5]. We identify all left/right seeds, covers and seeds of a Fibonacci string as well as all covers of a circular Fibonacci string, using a different approach than that of Iliopoulos, Moore and Smyth [11]. We then extend our previous results by giving comments on the number of quasiperiodicities in Fibonacci strings. It is important to note that we restrict to those quasiperiodicities that are factors of the considered strings.

The rest of the paper is structured as follows. In Section 1 we present the basic definitions used throughout the paper. In Section 2 we prove some properties of seeds, covers, periods and borders used later for finding quasiperiodicities in Fibonacci strings. We are then able to identify quasiperiodicities in Fibonacci strings (Section 3) and circular Fibonacci strings (Section 4). In Section 5 we give further comments on the number of distinct seeds in Fibonacci strings and finally we give some future proposals and a brief conclusion in Section 6.

1 Definitions and Problems

Throughout this paper we consider a string y of length $|y| = n$, $n > 0$, on a fixed alphabet. It is represented as $y[1\,..\,n]$. A string w is a *factor* of y if $y = uwv$ for two strings u and v. It is a *prefix* of y if u is empty and a *suffix* of y if v is empty. We denote the longest common prefix of two strings x and y as $LCP(x, y)$. A string u is a *border* of y if u is both a prefix and a suffix of y. *The border* of y, denoted by *border*(y), is the length of the longest border of y. A string u is a *period* of y if y is a prefix of u^k for some positive integer k, or equivalently if y is a prefix of uy. *The period* of y, denoted by *period*(y), is the length of the shortest period of y. For a string $u = u[1\,..\,m]$ such that u and v share a common part $u[m-\ell+1\,..\,m] = v[1\,..\,\ell]$ for some $1 \le \ell \le m$, the string $u[1\,..\,m]v[\ell+1\,..\,n] = u[1\,..\,m-\ell]v[1\,..\,n]$ is called a *superposition* of u and v with an *overlap* of length ℓ. A string x of length m is a *cover* of y if both $m < n$ and there exists a set of positions $P \subseteq \{1, \ldots, n-m+1\}$ that satisfies both $y[i\,..\,i+m-1] = x$ for all $i \in P$ and $\bigcup_{i \in P}\{i, \ldots, i+m-1\} = \{1, \ldots, n\}$. A string v is a *seed* of y, if it is a cover of a superstring of y, where a superstring of y is a string of form uyv and u, v are possibly empty strings. A left seed of a string y is a prefix of y that is a cover of a superstring of y of the form yv, where v is a possibly empty string. Similarly a right seed of a string y is a suffix of y that is a cover of a superstring of y of the form vy, where v is a possibly empty string.
We define a (finite) Fibonacci string F_k , $k \in \{0, 1, 2, \ldots\}$, as follows:
$F_0 = b$, $F_1 = a$, $F_n = F_{n-1}F_{n-2}$ $n \in \{2, 3, 4, \ldots\}$
A (finite) circular Fibonacci string $C(F_k)$, $k \in \{0, 1, 2, \ldots\}$, is made by concatenating the first letter of F_k to its last letter. As before a factor u of $C(F_k)$ is a cover of $C(F_k)$ if every letter of $C(F_k)$ falls within an occurrence of u within $C(F_k)$.

The following example shows all left seeds, right seeds, covers and seeds of the string F_6 = abaababaabaab.

All covers of F_6: abaab, abaababaabaab
All left seeds of F_6: aba, abaab, abaaba, abaababa, abaababaa, abaababaab, abaababaaba, abaababaabaa, abaababaabaab
All right seeds of F_6: abaab, abaabaab, babaabaab, ababaabaab, aababaabaab, baababaabaab, abaababaabaab
All seeds of F_6: aba, abaab, abaaba, abaababa, abaababaa, abaababaab, abaababaaba, abaababaabaa, abaababaabaab

F_0 b
F_1 a
F_2 ab
F_3 aba
F_4 abaab
F_5 abaababa
F_6 abaababaabaab
F_7 abaababaabaababaababa

Figure 1: The first eight Fibonacci strings

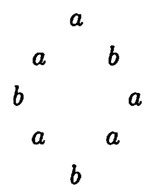

Figure 2: $C(F_5)$

2 Properties

In this section, we prove and also quote some properties for the covers, the left/right seeds and the seeds of a given string as well as some facts on Fibonacci strings that will prove useful later on.

Lemma 1. [4] A string z is a left seed of y iff it is a cover of a prefix of y whose length is at least the period of y.

Proof. Direct: Suppose a string z is a cover of a prefix of y, say uv, larger or equal to period(y), where $|u|$ = period(y) and v is a possibly nonempty string. Let k the smallest integer such that y a prefix of u^k. Then z is a cover of $u^k v = ywv$, for some string w, possibly empty. Therefore z is a left seed of y.
Reverse: Let z be a left seed of y.

- if $|z| \leq border(y)$. Then a suffix v of z (possibly empty) is a prefix of the border (consider the left seed that covers $y[period(y)]$). Then z is a cover of uv, where u is the period of y.

- if $|z| > border(y)$. Let z not a cover of a prefix of y larger or equal to $|period(y)|$. Let v a border of y such that $|v| = border(y)$. Then v is a factor of z, such that $z = uvw$, where u and w are nonempty strings (consider the left seed that covers $y[period(y)]$). This gives uv a longest border for y, which is a contradiction.

□

Lemma 2. [3] A string z is a right seed of y iff it is a cover of a suffix of y whose length is at least the period of y.

Proof. Direct consequence of Lemma 1. □

The following property of borders, which we give without proof, is quoted in almost every publication regarding periodicity.

Lemma 3. Let u be a border of x and let $z \neq u$ be a factor of x such that $|z| \leq |u|$. Then z is a border of x if and only if z is a border of u.

Lemma 4. [18] Let u be a proper cover of x and let $z \neq u$ be a factor of x such that $|z| \leq |u|$. Then z is a cover of x if and only if z is a cover of u.

Proof. Clearly if z is a cover of u and u is a cover of x the z is a cover of x. Suppose now that both z and u are covers of x. Then z is a border of x and hence of u ($|z| \leq |u|$); thus z must also be a cover of u. □

Lemma 5. [7]

$$\text{All borders of } F_n \text{ are: } \begin{cases} \{\}, & n \in \{0,1,2\} \\ \{F_{n-2}, F_{n-4}, F_{n-6}, \dots F_1\}, & n = 2k+1, k \geq 1 \\ \{F_{n-2}, F_{n-4}, F_{n-6}, \dots F_2\}, & n = 2k, k \geq 2 \end{cases} \tag{1}$$

Lemma 6. [10] $F_k = P_k\delta_k$, where $P_k = F_{k-2}F_{k-3}\dots F_1$, $k \geq 2$ and $\delta_k = ab$ if k is even or $\delta_k = ba$ otherwise.

Proof. Easily proved by induction. □

It is sometimes useful to consider the expansion of a Fibonacci string as a concatenation of two Fibonacci factors. We define the (F_m, F_{m-1})-expansion of F_n, where $n \in \{2,3,\dots\}$ and $m \in \{1,2,\dots,n-1\}$, as follows:

- Expand F_n using the recurrence formula as $F_{n-1}F_{n-2}$.
- Expand F_{n-1} using the recurrence formula as $F_{n-2}F_{n-3}$.
- Keep expanding as above until F_{m+1} is expanded.

Lemma 7. The (F_m, F_{m-1})-expansion of F_n, where $n \in \{2,3,\dots\}$ and $m \in \{1,2,\dots,n-1\}$ is unique.

Proof. Easily proved by induction. □

Lemma 8. The starting positions of the occurrences of F_m in F_n are the starting positions of the factors considered in the (F_m, F_{m-1})-expansion of F_n, where $n \in \{2,3,\dots\}$ and $m \in \{1,2,\dots,n-1\}$ except from the last F_{m-1}, if it is a border of F_n.

Proof. Using the recurrence relation we can get the (F_m, F_{m-1})-expansion of F_n as shown before:
$F_n = F_m F_{m-1} F_m F_m F_{m-1} F_m F_{m-1} F_m \dots$
We can now observe many occurrences of F_m in F_n. Any other occurrence should have one of the following forms (note that there are no consecutive F_{m-1} in the above expansion):

- xy, where x is a nonempty suffix of F_m and y a nonempty prefix of F_{m-1}. Then both x and y are borders of F_m. It holds that $|x|+|y| = |F_m| = |F_{m-1}| + |F_{m-2}|$, but $|x| \leq |F_{m-2}|$, $|y| \leq |F_{m-2}|$ and so there exists no such occurrence of F_m in F_n.

- xy, where x is a nonempty suffix of F_{m-1} and y a nonempty prefix of F_m. Then y is also a border of F_m and so belongs to $\{F_{m-2}, F_{m-4}, \dots F_1\}$, if n is odd, or to $\{F_{m-2}, F_{m-4} \dots F_2\}$, otherwise. But $|x|+|y| = |F_m|$ and $0 < |x| \leq |F_{m-1}|$ so in either case the only solution is $x = F_{m-1}$ and $y = F_{m-2}$ giving the occurrences of F_m at the starting positions of F_{m-1} in the above expansion.

- $xF_{m-1}y$, where x is a nonempty suffix of F_m and y a nonempty prefix of F_m. Then both x and y are borders of F_m. It holds that $|x|+|y| = |F_{m-2}|$, but as both x and y are nonempty $|x| \leq |F_{m-4}|$, $|y| \leq |F_{m-4}|$ and so there exists no such occurrence of F_m in F_n.

- xy, where x is a nonempty suffix of F_m and y a nonempty prefix of F_m(note that there is no such occurrence in the F_{n-2}, F_{n-1} expansion). Then both x and y are borders of F_m. It holds that $|x| + |y| = |F_m|$, but as both x and y are nonempty $|x| \leq |F_{m-2}|$, $|y| \leq |F_{m-2}|$ and so there exists no such occurrence of F_m in F_n.

□

Lemma 9. For every integer $n \geq 5$, $F_n[1 \,..\, |F_{n-1}| - 1]$ is not a left seed of F_n.

Proof. Using the recurrence relation we can expand F_n , $n \geq 5$, in the following two ways:
$F_n = F_{n-2}F_{n-3}F_{n-2} = F_{n-2}F_{n-2}F_{n-5}F_{n-4}$
Then one can see that $x = F_n[1 \,..\, |F_{n-1}| - 1] = F_{n-2}P_{n-3}\delta_{n-3}[1]$ (Lemma 6). Using Lemma 8 we can see that by expanding x from the prefix and suffix positions of F_{n-2} we cover F_n except $F_n[|F_{n-1}|]$. Expanding F_{n-2} from its middle occurrence yields the factor $y = F_{n-2}F_{n-5}P_{n-4}\delta_{n-4}[1] = F_{n-2}P_{n-3}\delta_{n-4}[1]$. It is easy to see that x and y differ at their last letter and hence the above result follows.

□

Lemma 10. For every integer $n \geq 5$, xF_{n-4}, where x is a suffix of F_{n-3} and $0 < |x| < F_{n-3}$, is not a right seed of F_n.

Proof. Using the recurrence relation we can expand F_n, $n \geq 5$, in the following way:
$F_n = F_{n-4}F_{n-5}F_{n-4}F_{n-4}F_{n-5}F_{n-4}F_{n-5}F_{n-4}$
Then any right seed of the form xF_{n-4}, $0 < |x| < |F_{n-3}|$, has x as a suffix of $F_{n-4}F_{n-5}$. Clearly the 3 occurrences of F_{n-4} at the starting positions of F_{n-5} (Lemma 8) cannot be expanded to their left to give right seeds as an F_{n-4} is to their left, which has a different ending than that of F_{n-5} (Lemma 6). Then $F_n[|F_{n-4}| + |F_{n-5}| + |F_{n-4}| + 1]$ cannot be covered by expanding the other 5 occurrences of F_{n-4} in F_n. □

3 Quasiperiodicities in Fibonacci strings

In this section we identify quasiperiodicities in Fibonacci strings (left seeds, right seeds, seeds, covers).

Identifying all covers of a Fibonacci string is made possible by identifying the longest cover of the string and then applying Lemma 4 as shown in the theorem below.

Theorem 1.

$$\text{All covers of } F_n \text{ are: } \begin{cases} F_n, & n \in \{0,1,2,3,4\} \\ \{F_n, F_{n-2}, F_{n-4}, F_{n-6}, \ldots F_3\}, & n = 2k+1, k \geq 2 \\ \{F_n, F_{n-2}, F_{n-4}, F_{n-6}, \ldots F_4\}, & n = 2k, k \geq 3 \end{cases} \tag{2}$$

Proof. It is easy to see that the theorem holds for $n \in \{0, 1, 2, 3, 4\}$. Using the recurrence relation we can expand F_n , $n \geq 5$, in the following two ways:
$F_n = F_{n-2}F_{n-3}F_{n-2} = F_{n-2}F_{n-2}F_{n-5}F_{n-4}$
It is now obvious that F_{n-2} is a cover of F_n. By Lemma 5 F_{n-2} is also the longest border of F_n and therefore the second longest cover of F_n (after F_n). Similarly F_{n-4} is the longest cover of F_{n-2}, F_{n-6} is the longest cover of F_{n-4}, etc. Hence by following Lemma 4 we get the above result. □

Corollary 1.

$$\text{The number of covers of } F_n \text{ is: } \begin{cases} 1, & n \in \{0,1,2,3,4\} \\ \frac{n-1}{2}, & n = 2k+1, k \geq 2 \\ \frac{n}{2} - 1, & n = 2k, k \geq 3 \end{cases} \tag{3}$$

Identifying left seeds of a Fibonacci string F_n is made possible for large n by characterizing each possible left seed as a factor of the form $F_m x$, where $m \in \{3, \ldots n-1\}$ and x a possibly empty prefix of F_{m-1}. We then use the (F_m, F_{m-1})-expansion of F_n along with Lemma 9 and the following result follows.

Theorem 2. All left seeds of F_n are:

- F_n, if $n \in \{0, 1, 2\}$
- $\{ab, aba\}$, if $n = 3$
- $\{F_{n-1}x\colon x$ a possibly empty prefix of $F_{n-2}\} \cup_{m=3}^{n-2} \{F_m x\colon x$ a possibly empty prefix of $F_{m-1}[1 \mathinner{..} |F_{m-1}| - 2]\}$, if $n \geq 4$

Proof. It is easy to see that the theorem holds for $n \in \{0, 1, 2, 3, 4\}$. For even $n \geq 5$ by Theorem 1 $\{F_n, F_{n-2}, F_{n-4}, \ldots F_4\}$ are covers of F_n and therefore left seeds of F_n. Again by Theorem 1 $\{F_{n-1}, F_{n-3}, F_{n-5}, \ldots F_3\}$ are all covers of F_{n-1} which is the period of F_n and hence by Lemma 1 $\{F_n, F_{n-1}, F_{n-2}, \ldots F_3\}$ are left seeds of F_n. By making similar observations for odd $n \geq 5$ we get that $\{F_n, F_{n-1}, F_{n-2}, \ldots F_3\}$ are all left seeds of F_n in either case. Only a and ab might be shorter left seeds but they are rejected as they are not left seeds of F_4 and so they are not left seeds of any longer Fibonacci string (F_4 is a prefix of every other F_n, $n \geq 5$). Therefore the remaining left seeds are of the form $F_m x$, where $m \in \{3, 4, \ldots, n-1\}$ and $0 < |x| < |F_{m-1}|$. Using the recurrence relation we can get the (F_m, F_{m-1})-expansion of F_n for any $m \in \{3, 4, \ldots, n-1\}$ as shown before:
$F_n = F_m F_{m-1} F_m F_m F_{m-1} F_m F_{m-1} F_m \ldots$
We then try to expand the seed from each F_m, F_{m-1} in the above expansion as of Lemma 8 (note that there are no consecutive F_{m-1} in the above expansion).
$F_m F_{m-1} = F_m P_{m-1} \delta_{m-1}$
$F_m F_m = F_m F_{m-1} F_{m-2} = F_m P_{m-1} \delta_{m-1} F_{m-2}$
$F_{m-1} F_m = F_{m-1} F_{m-2} F_{m-3} F_{m-2} = F_m F_{m-3} P_{m-2} \delta_{m-2} = F_m P_{m-1} \delta_{m-2}$
It is now obvious that any $F_m x$, where $m \in \{3, 4, \ldots, n-1\}$ and $0 < |x| \leq |F_{m-1}| - 2$ is a left seed of F_n. $F_{n-1} F_{n-2}[1 \mathinner{..} |F_{n-2}| - 1]$ is the only other left seed as it covers the period of F_n (Lemma 1). That there are no left seeds of form $F_m x$, where $m \in \{3, 4, \ldots, n-2\}$ and $|x| = |F_{m-1}| - 1$ follows from Lemma 9.

□

Corollary 2.

$$\text{The number of left seeds of } F_n \text{ is: } \begin{cases} 1, & n \in \{0, 1, 2\} \\ 2, & n \in \{3\} \\ |F_n| - n + 3, & n \geq 4 \end{cases} \qquad (4)$$

Identifying right seeds of a Fibonacci string F_n is made possible for large n by characterizing each possible right seed as a factor of the form xF_m, where $m \in \{3, 5, \ldots n-2\}$ if n is odd or $m \in \{4, 6, \ldots n-2\}$ if n is even, and x is a possibly empty suffix of F_{m+1}. We then use the (F_m, F_{m-1})-expansion of F_n along with Lemma 10 and the following result follows.

Theorem 3. All right seeds of F_n are:

- F_n, if $n \in \{0, 1, 2\}$
- $\{F_n, F_{n-2}, F_{n-4}, F_{n-6}, \ldots F_3\} \cup \{xF_{n-3}F_{n-2}$: x a possibly empty suffix of $F_{n-2}\}$, if $n = 2k+1$, $k \geq 1$
- $\{F_n, F_{n-2}, F_{n-4}, F_{n-6}, \ldots F_4\} \cup \{xF_{n-3}F_{n-2}$: x a possibly empty suffix of $F_{n-2}\}$, if $n = 2k$, $k \geq 2$

Proof. It is easy to see that the theorem holds for $n \in \{0, 1, 2, 3, 4\}$. For even $n \geq 5$, by Theorem 1 $\{F_n, F_{n-2}, F_{n-4}, \ldots F_4\}$ are covers of F_n and therefore right seeds of F_n. Only $\{baab, aab, ab, b\}$ might be shorter right seeds but they are rejected as they are not right seeds of F_6 and so they are not right seeds of any F_n, where n even and $n \geq 5$ (F_6 is a suffix of every other F_n, n even and $n \geq 5$). Similarly for odd $n \geq 5$ $\{F_n, F_{n-2}, F_{n-4}, \ldots F_3\}$ are right seeds of F_n and F_3 is its shortest right seed. Therefore the remaining right seeds are of the form xF_m, where $0 < |x| < |F_{m+1}|$ and $m \in \{4, 6, \ldots, n-2\}$, if n is even, or $m \in \{3, 5, \ldots, n-2\}$, otherwise.
The only other right seeds are of the form $xF_{n-3}F_{n-2}$,where x is a suffix of F_{n-2} and $0 \leq |x| < F_n$, as it is easy to see that they cover the period of F_n (Lemma 2).
The fact that there are no right seeds of form xF_{n-2}, where $0 < |x| < |F_{n-3}|$, follows from Lemma 8. Clearly the middle occurrence of F_{n-2} cannot be expanded to the left as an F_{n-2} is to its left, which has a different ending than that of F_{n-3} at the left of the last F_{n-2}. Then $F_n[|F_{n-2}|+1]$ cannot be covered by expanding the other 2 occurrences of F_{n-2} in F_n.
The fact that there are no right seeds of the form xF_m, where $0 < |x| < |F_{m+1}|$ and $m \in \{4, 6, \ldots, n-4\}$, n is even, or $m \in \{3, 5, \ldots, n-4\}$, otherwise, follows from Lemma 10. □

Corollary 3.

$$\text{The number of right seeds of } F_n \text{ is: } \begin{cases} 1, & n \in \{0,1,2\} \\ |F_{n-2}| + \frac{n-1}{2}, & n = 2k+1, k \geq 1 \\ |F_{n-2}| + \frac{n}{2} - 1, & n = 2k, k \geq 2 \end{cases} \tag{5}$$

Identifying all seeds of a Fibonacci string F_n is made possible for large n by characterizing each possible seed as a factor of the form xF_my, where $m \in \{3, 4, \dots n-1\}$ and x, y follow some restrictions such that F_m is the longest Fibonacci factor in the seed and no occurrence of F_m in the seed starts from a position in x. We then use the (F_m, F_{m-1})-expansion of F_n along with Lemma 8 and the result below follows.

Theorem 4. All seeds of F_n are:

- all left/right seeds of F_n, if $n \in \{0, 1, 2, 3\}$
- all left/right seeds of F_n and baa, if $n = 4$
- all left/right seeds of F_n,
  strings of form $\{xF_my$: x a suffix of F_m,y a prefix of F_{m-1}, $0 < |x| < |F_m|$, $0 < |y| < |F_{m-1}| - 1$, $|x| + |y| \geq F_{m-1}$ and $m \in \{3, \dots, n-3\}\}$,
  strings of form $\{xF_{m-1}F_my$: x a suffix of F_m,y a prefix of F_{m-1}, $|x| + |y| \geq F_m$ and $m \in \{3, \dots, n-3\}\}$,
  strings of form $\{xF_{n-2}y$: x a suffix of F_{n-2},y a prefix of $F_{n-5}F_{n-4}$, $0 < |x| < |F_{n-2}|$, $0 < |y| \leq |F_{n-3}|$ and $|x| + |y| \geq |F_{n-3}|\}$, if $n \geq 5$

Proof. It is easy to see that the theorem holds for $n \in \{0, 1, 2, 3, 4\}$. For $n \geq 5$ it is obvious that all left seeds of F_n and all right seeds of F_n are also seeds of F_n.
Therefore the remaining seeds are of the form xF_my, such that F_m is the leftmost occurrence of the longest Fibonacci string present in the seed, $m \in \{3, 4, \dots, n-2\}$, $|x| > 0$ and $|y| > 0$.
For $m = n-2$ the expansion of $F_n = F_{n-2}F_{n-3}F_{n-2} = F_{n-2}F_{n-2}F_{n-5}F_{n-4}$ is very small so we consider it separately. By expanding the middle occurrence of F_{n-2} we get the seed $xF_{n-2}y$, where $0 < |x| < |F_{n-2}|$, $0 < |y| \leq |F_{n-3}|$ and $|x| + |y| \geq |F_{n-3}|$. As of Lemma 8 the remaining seeds of form xF_my, such that F_m is the leftmost occurrence of the longest Fibonacci string present in the seed, $m \in \{3, 4, \dots, n-3\}$, $|x| > 0$ and $|y| > 0$, have their leftmost F_m factor occurring in the start position of either an F_m or an F_{m-1} in the (F_m, F_{m-1})-expansion of $F_n = F_mF_{m-1}F_mF_mF_{m-1}F_mF_{m-1}F_m \dots$
We consider the following cases (note that there are no consecutive F_{m-1} in the above expansion):

- A seed of form xF_my, such that F_m has a F_{m-1} to its left in the (F_m, F_{m-1})-expansion of F_n and $0 < |x| < F_{m-1}$ (otherwise there exist a new leftmost occurrence of F_m in the seed). The occurrences of F_m that we are considering have starting positions only from a F_m in the expansion of F_n, then y can be up to $F_{m-1}[1 \,..\, |F_{m-1}| - 1]$ (otherwise a F_{m+1} is created). But such a seed fails to cover $F_n[|F_mF_{m-1}F_mF_{m-1}| - 1]$.

- A seed of form xF_my, such that F_m has a F_m to its left in the (F_m, F_{m-1})-expansion of F_n and $0 < |x| < F_m$ (otherwise a F_{m+1} is created). If the occurrences of F_m that we are considering have starting positions both from a F_m and a F_{m-1} in the expansion of F_n, then y can be up to $F_{m-1}[1..|F_{m-1}| - 2]$ (otherwise the factors differ). Furthermore $|x| + |y| \geq |F_{m-1}|$,such as to cover $F_n[|F_mF_{m-1}| + 1..|F_mF_{m-1}F_m|]$. Such a seed covers F_n as $F_{m+2} = F_mF_{m-1}F_m = F_mF_mP_{m-1}\delta_{m-2}$ is a left seed of F_n (Theorem 2) composing F_n with concatenations of overlap 0 (factors are joined by considering the seed that its leftmost F_m starts from the next F_{m+2}) or F_m (factors are joined as $|x|+|y| \geq F_{m-1}$). If the occurrences of F_m that we are considering have starting positions only from a F_m in the expansion of F_n, then y can be up to $F_{m-1}[1..|F_{m-1}|-1]$ (otherwise a F_{m+1} is created). But such a seed fails to cover $F_n[|F_mF_{m-1}|]$. If the occurrences of F_m that we are considering have starting positions only from a F_{m-1} in the expansion of F_n, then $|y|$ can be up to $2|F_{m-1}|-1]$ (otherwise a F_{m+1} is created). Furthermore $|x|+|y| \geq |F_m|+|F_{m-1}| = |F_{m+1}|$,such as to cover $F_n[|F_mF_{m-1}| + 1..|F_mF_{m-1}F_mF_m|]$. Such a seed covers F_n as $F_{m+2} = F_mF_{m-1}F_m = F_mF_mP_{m-1}\delta_{m-2}$ is a left seed of F_n (Theorem 2) composing F_n with concatenations of overlap 0 (factors are joined as $|x| + |y| \geq |F_{m+1}|$) or F_m (factors are joined as $|x| + |y| \geq |F_{m+1}| > |F_{m-1}|$).

□

Corollary 4. The number of seeds of F_n is $\Omega(|F_n|^2)$.

4 Quasiperiodicities in circular Fibonacci strings

Finding all covers of a circular Fibonacci string is now obvious, we just need to check the seeds of the relevant Fibonacci string. Those which are covers of a superstring of form xF_ny, where x is a possibly empty suffix of F_n and y is a possibly empty prefix of F_n are covers of $C(F_n)$.

Theorem 5. All covers of $C(F_n)$ are:

- F_n, if $n \in \{0, 1, 2, 3\}$
- F_n and F_{n-1}, if $n = 4$
- F_n, strings of form $\{F_mx$: x a possibly empty prefix of $F_{m-1}[1..|F_{m-1}| - 2]$ and $m \in \{3, \ldots, n-1\}\}$,
strings of form $\{xF_my$: x a suffix of F_m, y a prefix of F_{m-1}, $0 < |x| <$

$|F_m|$, $0 < |y| < |F_{m-1}| - 1$, $|x| + |y| \geq F_{m-1}$ and $m \in \{3, \ldots, n-2\}\}$, strings of form $\{xF_{m-1}F_m y$: x a suffix of F_m, y a prefix of F_{m-1}, $|x| + |y| \geq F_m$ and $m \in \{3, \ldots, n-3\}\}$, if $n \geq 5$

Proof. It is easy to see that the theorem holds for $n \in \{0, 1, 2, 3, 4\}$. For larger n the covers of $C(F_n)$ are at most the seeds of F_n. A seed is a cover of $C(F_n)$ iff it covers a superstring of F_n of form $xF_n y$, where x is a possibly empty suffix of F_n and y is a possibly empty prefix of F_n. We consider the following cases:

- Left seeds of form $F_n[1\,..\,|F_k| + i]$, where $i \in \{0, 1, \ldots, |F_{k-1}| - 2\}$ and $k \in \{3, 4, \ldots n-1\}$, are covers of $F_n F_k[1\,..\,i]$,if F_k is a cover of F_n, or covers of $F_n F_k[1\,..\,|F_{k-2}| + i]$ otherwise, and hence covers of $C(F_n)$ in both cases. Clearly F_n is also a cover of $C(F_n)$. $F_n[1\,..\,|F_n| - 1]$ is not a cover of $C(F_n)$ as it fails to cover a prefix of $F_n F_n$ longer than $|F_n| - 1$ (consider the F_{n-1}, F_{n-2} expansion of F_n along with Lemma 8).

- The only right seeds of F_n that are covers of $C(F_n)$ are the covers of F_n (included above). Right seeds of form $xF_{n-3}Fn-2$, where x is a possibly empty suffix of F_{n-2} and $0 \leq |x| < |F_{n-2}|$, fail to cover a suffix of $F_n F_n = Fn - 2F_{n-3}Fn - 2$ longer than $|xF_{n-3}F_{n-2}|$ (consider the F_{n-2}, F_{n-3} expansion of F_n along with Lemma 8), and so they are not covers of $C(F_n)$.

- Seeds of form $xF_m y$ where x a suffix of F_m and y a prefix of F_{m-1}, $0 < |x| < |F_m|$, $0 < |y| < |F_{m-1}| - 1$, $|x| + |y| \geq F_m$ and $m \in \{3, 4, \ldots, n-3\}$ are covers of $xF_n F_m y$,if F_m is a cover of F_n, or covers of $xF_{m-1}F_n F_{m-2}y$ otherwise, and hence covers of $C(F_n)$ in both cases.

- Seeds of form $xF_{m-1}F_m y$ where x a suffix of F_m and y a prefix of F_{m-1}, $0 < |x| < |F_m|$, $0 < |y| < |F_{m-1}|$, $|x| + |y| \geq F_m$ and $m \in \{3, 4, \ldots, n-3\}$ are covers of $xF_{m-1}F_m F_n y$,if F_m is a cover of F_n, or covers of $xF_{m-1}F_n F_m y$ otherwise, and hence covers of $C(F_n)$ in both cases.

- Seeds of form $xF_{n-2}y$ where x a suffix of F_{n-2} and y a prefix of $F_{n-5}F_{n-4}$ such that $0 < |x| < |F_{n-2}|$, $0 < |y| < |F_{n-3}| - 1$ and $|x| + |y| \geq |F_{n-3}|$ are covers of $xF_n F_{n-2}y$ and hence covers of $C(F_n)$. When $y = F_{n-5}F_{n-4}$ or $F_{n-5}F_{n-4}[1\,..\,|F_{n-4}| - 1]$ the seed fails to cover $xF_n F_{n-2}y$, the first F_{n-2} of F_n cannot be expanded further to the right. Trying to force an overlap of $xF_{n-2}y$ to the left of $F_n F_{n-2}y$ gives the superstrings $xF_{n-3}F_n F_{n-2}y$ and $xF_{n-2}F_{n-5}F_n F_{n-2}y$ (consider the occurrences of F_{n-4} in F_n), which are not made of suffixes of F_n, as clearly F_{n-3} and $Fn - 5$ are not borders of F_n (Lemma 5), and so they are not covers of $C(F_n)$.

□

Corollary 5. The number of covers of $C(F_n)$ is $\Omega(|C(F_n)|^2)$.

5 Bounds on the number of seeds of a string

It is easy to see that the number of seeds a string x is bounded above by $\frac{|x|^2}{2}$.

Theorem 6. The number of distinct seeds of a nonempty string x is at most $\frac{n(n+1)}{2}$, where $|x| = n$.

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

We have seen that the number of seeds a Fibonacci string F_n is $\Omega(|F_n|^2)$ (Theorem 4). The following theorem proves that as $n \to +\infty$ the ratio of the number of distinct seeds of F_n to the square of its length converges to $\frac{\phi^2+1}{2\phi^6} = 0.100813061875578\ldots$, i.e. still not very close to $\frac{1}{2}$.

Theorem 7. $\lim_{n\to+\infty} \frac{\text{Seeds}(F_n)}{|F_n|^2} = \frac{\phi^2+1}{2\phi^6}$

Proof. Summing all the seeds of Theorem 4 and considering only the quadratic terms we get:

$$\lim_{n\to+\infty} \frac{\text{Seeds}(F_n)}{|F_n|^2}$$
$$= \lim_{n\to+\infty} \sum_{m=0}^{n-3} \frac{|F_m|^2 + 2|F_m||F_{m-1}|}{2|F_n|^2}$$
$$= \lim_{n\to+\infty} \sum_{m=0}^{n-3} \frac{2|F_m|^2 + |F_{m+1}|^2 - |F_{m-2}|^2}{4|F_n|^2}$$
$$= \lim_{n\to+\infty} \frac{2|F_{n-3}||F_{n-2}| + |F_{n-2}||F_{n-1}| - |F_{n-5}||F_{n-4}|}{4|F_n|^2}$$
(as $\sum_{i=0}^{n} |F_i|^2 = |F_n||F_{n+1}|$)
$$= \lim_{n\to+\infty} \frac{1}{4|F_n|^2}(2|F_{n-3}||F_{n-2}| + |F_{n-2}|(|F_{n-2}| + |F_{n-3}|)$$
$$-(2|F_{n-3}| - |F_{n-2}|)(|F_{n-2}| - |F_{n-3}|))$$
$$= \lim_{n\to+\infty} \frac{|F_{n-2}|^2 + |F_{n-3}|^2}{2|F_n|^2}$$
$$= \frac{\phi^{-4}+\phi^{-6}}{2} \text{ (as } \lim_{n\to+\infty} \frac{F_n}{F_{n-1}} = \phi)$$

$= \frac{\phi^2+1}{2\phi^6}$
$= 0.100813061875578\ldots$ □

6 Conclusion and Future Work

In this paper we have presented our preliminary results on Fibonacci strings, we have identified all left seeds, right seeds, seeds and covers of every Fibonacci string as well as all covers of a circular Fibonacci string under the restriction that these quasiperiodicities are also factors of the given Fibonacci string. We were then able to give lower bounds on the number of distinct seeds of a string. Beyond their obvious theoretical interest, those results might prove useful in testing algorithms that find quasiperiodicities in strings and giving worst case examples on them or in extending the above work in general Sturmian strings (the infinite Fibonacci string, a string which has every Fibonacci string as a prefix, is Sturmian).

References

[1] A. Apostolico and A. Ehrenfeucht. Efficient detection of quasiperiodicities in strings. *Theor. Comput. Sci.*, 119(2):247–265, 1993.

[2] J. Berstel. Fibonacci words-a survey. *The book of L*, pages 13–27.

[3] M. Christou, M. Crochemore, O. Guth, C. Iliopoulos, and S. Pissis. On the right-seed array of a string. *Computing and Combinatorics*, pages 492–502, 2011.

[4] M. Christou, M. Crochemore, C. Iliopoulos, M. Kubica, S. Pissis, J. Radoszewski, W. Rytter, B. Szreder, and T. Waleń. Efficient seeds computation revisited. In *Combinatorial Pattern Matching*, pages 350–363. Springer, 2011.

[5] M. Christou, M. Crochemore, and C. S. Iliopoulos. Quasiperiodicities in Fibonacci strings. In *Local Proceedings of the* 38^{th} *International Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM 2012)*, 2012.

[6] W. Chuan and H. Ho. Locating factors of the infinite Fibonacci word. *Theoretical computer science*, 349(3):429–442, 2005.

[7] L. Cummings, D. Moore, and J. Karhumaki. Borders of Fibonacci strings. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 20:81–88, 1996.

[8] X. Droubay. Palindromes in the Fibonacci word. *Information Processing Letters*, 55(4):217–221, 1995.

[9] A. Fraenkel and J. Simpson. The exact number of squares in Fibonacci words. *Theoretical Computer Science*, 218(1):95–106, 1999.

[10] C. Iliopoulos, D. Moore, and W. Smyth. A characterization of the squares in a Fibonacci string. *Theoretical Computer Science*, 172(1-2):281–291, 1997.

[11] 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.

[12] C. Iliopoulos and K. Park. A work-time optimal algorithm for computing all string covers. *Theoretical Computer Science*, 164(1-2):299–310, 1996.

[13] C. S. Iliopoulos, D. W. G. Moore, and K. Park. Covering a string. *Algorithmica*, 16:289–297, Sept. 1996.

[14] T. Kociumaka, M. Kubica, J. Radoszewski, W. Rytter, and T. Waleń. A linear time algorithm for seeds computation. In *Proceedings of the 23^{rd} Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1095–1112. SIAM, 2012.

[15] Y. Li and W. Smyth. Computing the cover array in linear time. *Algorithmica*, 32(1):95–106, 2002.

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

[17] M. Lothaire, editor. *Appplied Combinatorics on Words*. Cambridge University Press, 2005.

[18] D. Moore and W. Smyth. An optimal algorithm to compute all the covers of a string. *Information Processing Letters*, 50(5):239–246, 1994.

[19] G. Richomme, K. Saari, and L. Zamboni. Balance and abelian complexity of the tribonacci word. *Advances in Applied Mathematics*, 45(2):212–231, 2010.

[20] S. Rosema and R. Tijdeman. The tribonacci substitution. *Integers: Electronic Journal of Combinatorial Number Theory*, 5(3):A13, 2005.

[21] B. Tan and Z. Wen. Some properties of the tribonacci sequence. *European Journal of Combinatorics*, 28(6):1703–1719, 2007.