The Exact Number of Squares in Fibonacci Words

Aviezri S. Fraenkel[1] and Jamie Simpson[2]

Abstract. All our words (sequences) are binary. A *square* is a subword of the form uu (concatenation). Two squares are *distinct* if they are of different shape, not just translates of each other. Otherwise they are *repeated*. *Fibonacci words* are defined by $f_0 = 0$, $f_1 = 1$, $f_m = f_{m-1}f_{m-2}$ for $m \geq 2$. Let $F_m = |f_m|$. Then $F_0 = 1$, $F_1 = 1$, $F_m = F_{m-1} + F_{m-2}$ $(m \geq 2)$ are the Fibonacci numbers. Let $D(n)$ and $R(n)$ be the exact number of Distinct and Repeated squares respectively in f_n. We prove: $D(n) = 2(F_{n-2} - 1)$ $(n \geq 5)$, which implies, asymptotically, $D(n) = 2(2-\varphi)F_n + o(1)$ $\quad (2(2-\varphi) \approx 0.7639)$, where φ is the golden section. We also prove: $R(n) = \frac{4}{5}nF_n - \frac{2}{5}(n+6)F_{n-1} - F_{n-2} + n + 1$ $(n \geq 3)$. This yields $R(n) = \frac{2}{5}(3-\varphi)nF_n + O(F_n) = \frac{2(3-\varphi)}{5\log_2\varphi}F_n \log_2 F_n + O(F_n)$ $\quad \left(\frac{2(3-\varphi)}{5\log_2\varphi} \approx 0.7962\right)$.

1. Introduction

We consider *binary* words (sequences) throughout. A *square* in a word is a factor (subword) of the form u^2, i.e., the concatenation of two identical subwords. Two squares are *distinct* if they are of different shape, not just translates of each other. Two squares which are not distinct, beginning at two different locations of a word, are *repeated* squares.

A word u is *primitive* if u cannot be written in the form $u = v^j$ for some integer $j \geq 2$. Otherwise u is *nonprimitive*. A square u^2 with u primitive will be called a *primitive rooted* square.

Fibonacci words are defined by $f_0 = 0$, $f_1 = 1$, $f_m = f_{m-1}f_{m-2}$ (concatenation) for all $m \geq 2$. Let $F_m = |f_m|$. Then $F_0 = 1$, $F_1 = 1$, $F_m = F_{m-1} + F_{m-2}$ $(m \geq 2)$ are the Fibonacci numbers, since $|f_m| = |f_{m-1}f_{m-2}| = |f_{m-1}| + |f_{m-2}|$. The first 10 Fibonacci words are depicted in Table 1 in §2. Fibonacci words have been studied extensively. (The query Fibonacci AND word*, run on MathSciNet on April 2, 1997, yielded 86 papers.) The *infinite Fibonacci word* is the word which contains f_m as a prefix for every $m \geq 1$. For the study of repetitions in Fibonacci words, mainly in the infinite Fibonacci word, see e.g., [**dLV1990**], [**MiP1992**], [**IMS≥1997**]. Incidentally, all squares in f_m $(m \geq 0)$ are primitive rooted (see e.g., [**CrR1994**], Ch. 2).

Denote by $D(n)$ and by $R(n)$ the exact number of Distinct and Repeated squares respectively in the Fibonacci word f_n. In §2 we prove,

[1]Department of Applied Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 76100, Israel; `http://www.wisdom.weizmann.ac.il/~fraenkel` `fraenkel@wisdom.weizmann.ac.il`

[2]School of Mathematics, Curtin University, Perth WA 6001, Australia; `http://www.cs.curtin.edu.au/~simpson` `simpson@cs.curtin.edu.au`

Theorem 1. *For $n \geq 5$, $D(n) = 2(F_{n-2} - 1)$.*

We mention in passing that in [**Sto1976**] and [**FMT1978**], the multitude of squares in Fibonacci words and generalizations thereof is viewed from a slightly different, yet related point of view.

As customary, denote the roots of $x^2 - x - 1 = 0$ by $\varphi = \frac{1+\sqrt{5}}{2}$ (the golden section) and $\overline{\varphi} = \frac{1-\sqrt{5}}{2}$. Then $\varphi + \overline{\varphi} = 1$, $\varphi - \overline{\varphi} = \sqrt{5}$, and $\varphi\overline{\varphi} = -1$, $\varphi^2 = \varphi + 1$, $\overline{\varphi}^2 = \overline{\varphi} + 1$.

Denote by $N(a)$ the nearest integer to a. Theorem 1 leads to the following asymptotic result for $D(n)$.

Corollary 1. *We have*

$$D(n) = 2(2 - \varphi)F_n + o(1) \quad (2(2-\varphi) \approx 0.7639),$$

and also, for $n \geq 5$,

$$D(n) = N\left(\frac{2\varphi^{n-1}}{\sqrt{5}} - 2\right).$$

Fibonacci words have been used for giving lower bounds on the number of repeated primitive rooted squares. We are unaware of previous results for $D(n)$, but estimates for $R(n)$ were given previously. In [**Cro1981**], Lemma 10, it was shown that $R(n) \geq \frac{1}{6}F_n \log_2 F_n$ for $n \geq 5$. Theorem 11 in [**CrR1995**] says that the number of primitive rooted square prefixes of a word w is $< \log_\phi |w|$. For $w = f_n$ this implies $R(n) < F_n \log_\phi F_n = F_n \log_2 F_n / \log_2 \phi \leq 1.441 F_n \log_2 F_n$. In [**IMS≥1997**] the structure of squares in Fibonacci words, especially infinite Fibonacci words, was studied.

Rather than just estimating the number of squares in a finite Fibonacci word, as done previously, we give an exact formula for it.

Theorem 2. *The function $R(n)$ is given by*

$$R(n) = \frac{4}{5}nF_n - \frac{2}{5}(n+6)F_{n-1} - F_{n-2} + n + 1$$

for $n \geq 3$.

Theorem 2 improves the above two results [**Cro1981**], [**CrR1995**]. Asymptotically it gives,

Corollary 2. *We have*

$$R(n) = \frac{2}{5}(3-\varphi)nF_n + O(F_n) = \frac{2(3-\varphi)}{5\log_2\varphi}F_n \log_2 F_n + O(F_n) \quad \left(\frac{2(3-\varphi)}{5\log_2\varphi} \approx 0.7962\right).$$

Comparing this with Corollary 1, we see that the number of repeated squares is logarithmically larger than the number of distinct squares in f_n with almost the same numerical multiplicative constant. The maximum number $M(n)$ of distinct squares a binary word of length n can have, is larger than that found in Fibonacci words: in [**FrS≥1997**] we proved that $M(n) < 2n$ for all positive integers n, and constructed a special set of words to show that $M(n) = n - o(n)$ for infinitely many n. These two results hold for both nonprimitive and primitive rooted squares.

2. Distinct Squares in Fibonacci Words

We use several lemmas to establish Theorem 1. The Fibonacci words f_n were defined in the Introduction. The first few are listed in Table 1.

TABLE 1.

$f_0 = 0$
$f_1 = 1$
$f_2 = 10$
$f_3 = 101$
$f_4 = 10110$
$f_5 = 10110101$
$f_6 = 1011010110110$
$f_7 = 101101011011010110101$
$f_8 = 1011010110110101101011011010110110$
$f_9 = 1011010110110101101011011010110110101101011011010110101$

Lemma 1. *For $m \geq 5$, f_m contains f_{m-4}^2; and for $m \geq 7$, f_m contains f_{m-4}^3.*

Proof.

$$f_m = f_{m-1}f_{m-2} = f_{m-2}f_{m-3}f_{m-2} = f_{m-3}f_{m-4}^2f_{m-5}f_{m-2}.$$

The suffix f_{m-2} of this formula can be iterated to yield $f_{m-2} = f_{m-6}f_{m-7}f_{m-6}f_{m-5}f_{m-4}$, so $f_m = f_{m-3}f_{m-4}^3f_{m-7}f_{m-6}f_{m-5}f_{m-4}$. ■

Definition 1. For any word $x = x_1 \ldots x_n$ and any $j \in \{0, \ldots, n-1\}$, a word of the form $x_{j+1} \ldots x_n x_1 \ldots x_j$ is called a *conjugate* or a *rotation* of *degree* j of x.

A word of length n has n conjugates, including itself, which is a conjugate of degree 0.

Lemma 2. *Let $u = xxy$ for subwords x, y with $|y| < |x|$. Let i be any integer satisfying $0 \leq i \leq |y|$. Then u contains a square v_{i+1}^2 of length $2|x|$, beginning at position $i+1$ of u, where v_{i+1} is a conjugate of degree i of x, if and only if the prefix of x of length i is identical to the prefix of y of length i.*

Proof. See Fig. 1. Since $i \leq |y| < |x|$, v_{i+1}^2 is a square of length $2|x|$ beginning at position $i+1$ of u if and only if the prefix of length $i+1$ of the second occurrence of x = suffix of length $i+1$ of the first occurrence of v_{i+1} = suffix of length $i+1$ of the second occurrence of v_{i+1} = prefix of length $i+1$ of y. If the condition is satisfied, it's clear that v_{i+1} is a conjugate of degree i of x. ■

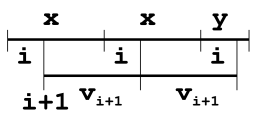

FIGURE 1.

Remark. Lemma 2 is true also for $|x| = |y|$. We excluded that case since for $i = |x| = |y|$, the square v_{i+1}^2 is x^2, which was already counted for $i = 0$. It will become clear from Lemma 8 below that, when u is a Fibonacci word, the squares v_{i+1}^2 are *distinct* in pairs for $i \in \{0, \dots |x| - 1\}$.

Notation. Let u, v be factors (subwords) of a word w such that $w = uv$. Then wv^{-1} denotes the prefix u of w with its suffix v deleted. Thus $u = wv^{-1}$. By $s_i(m)$ we denote the suffix of length i of f_m, and by $\overline{s}_i(m)$ its complement. For example, $f_4 = 10110$, so $f_4 s_2(4)^{-1} \overline{s}_2(4) = 10101$.

Lemma 3. *For $m \geq 5$ and $i \in \{1, \dots, m-4\}$, all the F_i conjugates u of f_i have the property that f_m contains u^2.*

Proof. Since $f_1 = 1, f_2 = 10$ and $1^2, (10)^2, (01)^2$ appear in f_m for every $m \geq 5$, the result holds for $i \in \{1, 2\}$. We may thus assume $i \geq 3$ (whence $m \geq 7$). Clearly f_m contains f_{i+4} as a subword. Now $i + 4 \geq 7$, so by Lemma 1, f_{i+4} contains f_i^3, hence f_m contains f_i^3. In particular, f_m contains xxy with $x = f_i$ and $y = f_i s_1(i)^{-1}$. The result now follows from Lemma 2. ■

Lemma 4. *Let k, m be integers with $k \geq 3$.*
For $m \geq 2k + 2$, we have

(1)
$$f_m = f_{m-3}^2 f_{m-4} f_{m-2k}^{-1} f_{m-2k-2} f_{m-2k-1} f_{m-2}.$$

For $m \geq 2k + 1$,

(2)
$$f_m = f_{m-2}^2 f_{m-3} f_{m-2k+1}^{-1} f_{m-2k-1} f_{m-2k}.$$

Proof. We remark that for the words (1) and (2) to even exist, f_{m-2k} has to be a suffix of f_{m-4}, and f_{m-2k+1} has to be a suffix of f_{m-3}. In fact, f_{m-2k} is trivially a suffix of f_{m-4} for $k = 2$. If it holds for k, then it holds also for $k+1$, since $f_{m-2k} = f_{m-2k-1} f_{m-2k-2}$. The same argument shows that f_{m-2k+1} is a suffix of f_{m-3}.

For proving (2) we write, for $k \geq 3$,

(3)
$$f_m = f_{m-2}^2 f_{m-4} f_{m-6} \cdots f_{m-2k+2} f_{m-2k-1} f_{m-2k},$$

whose validity we prove by induction on k:

$$\begin{aligned} f_m &= f_{m-1} f_{m-2} = f_{m-2} f_{m-3} f_{m-3} f_{m-4} = f_{m-2} f_{m-3} f_{m-4} f_{m-5} f_{m-5} f_{m-6} \\ &= f_{m-2}^2 f_{m-5} f_{m-6} f_{m-7} f_{m-6} = f_{m-2}^2 f_{m-4} f_{m-7} f_{m-6}, \end{aligned}$$

so (3) holds for $k = 3$. Assume it holds for k. From

$$\begin{aligned} f_{m-2k-1} f_{m-2k} &= f_{m-2k-1} f_{m-2k-1} f_{m-2k-2} = f_{m-2k-1} f_{m-2k-2} f_{m-2k-3} f_{m-2k-2} \\ &= f_{m-2k} f_{m-2k-3} f_{m-2k-2}, \end{aligned}$$

(with $m \geq 2(k+1)+1$) we see that it also holds for $k+1$, establishing (3).

We use (3) to write

$$f_m = f_{m-2}^2 f_{m-4} f_{m-6} \cdots f_{m-2k+2} f_{m-2k+1} f_{m-2k+1}^{-1} f_{m-2k-1} f_{m-2k}.$$

Now $f_{m-4} f_{m-6} \cdots f_{m-2k+2} f_{m-2k+1} = f_{m-3}$, by "collapsing" $f_{m-2k+3} = f_{m-2k+2} f_{m-2k+1}$ all the way through till $f_{m-3} = f_{m-4} f_{m-5}$. This proves (2). Finally, replace m by $m-1$ in (2), and concatenate f_{m-2} to the right end of both sides of the resulting identity, and then use $f_m = f_{m-1} f_{m-2}$ for the left side. This gives (1). ■

Lemma 5.

(a) *For* $m \geq 6$, $f_m = f_{m-3}^2 f_{m-4} s_2(m-4)^{-1} \overline{s}_2(m-4) f_{m-2}$.

(b) *For* $m \geq 5$, $f_m = f_{m-2}^2 f_{m-3} s_2(m-3)^{-1} \overline{s}_2(m-3)$.

Proof. The assertion (b) is seen to hold for $m = 5$ and $m = 6$ (Table 1). Thus we may assume $m \geq 7$.

Suppose first that m is odd. Then $m = 2j + 1$ for some $j \geq 3$. Since Lemma 4 holds for all $k \geq 3$, we may substitute $k = j = (m-1)/2$ into (2), to get $f_m = f_{m-2}^2 f_{m-3} f_2^{-1} f_0 f_1 = f_{m-2}^2 f_{m-3} (10)^{-1} 01 = f_{m-2}^2 f_{m-3} s_2(m-3)^{-1} \overline{s}_2(m-3)$ as asserted.

Secondly, let m be even. Then $m = 2j + 2$ with $j \geq 3$. Substituting $k = j = (m-2)/2$ into (2), we get $f_m = f_{m-2}^2 f_{m-3} f_3^{-1} f_1 f_2 = f_{m-2}^2 f_{m-3} (101)^{-1} 110 = f_{m-2}^2 f_{m-3} (01)^{-1} 10 = f_{m-2}^2 f_{m-3} s_2(m-3)^{-1} \overline{s}_2(m-3)$. For proving (a) we again replace m by $m-1$ in (b), followed by concatenating f_{m-2} to the right of both sides. ■

Lemma 6. *For* $m \geq 6$*, there are precisely* $F_{m-4}+1$ *conjugates* u *of* f_{m-3}*, of degrees* $0, \ldots, F_{m-4}$*, such that* u^2 *is in* f_m.

Proof. We scan f_m, from left to right, in a search for all the conjugates u of f_{m-3} such that u^2 is in f_m.

For $m \geq 5$, f_{m-4} is a prefix of $f_{m-3} = f_{m-4} f_{m-5}$. By Lemma 5 (a), we can thus apply Lemma 2 with $x = f_{m-3}$ and $y = f_{m-4} s_2(m-4)^{-1} \overline{s}_2(m-4)$, to conclude that for $m \geq 6$, f_m contains $F_{m-4} - 1$ conjugates u of f_{m-3}, of degrees $0, \ldots, F_{m-4} - 2$, such that u^2 is in f_m. These squares begin at positions $1, \ldots, F_{m-4} - 1$ of f_m.

By Lemma 2, the words of length $2F_{m-3}$ beginning at the positions in the closed interval $[F_{m-4}, F_{m-4} + F_{m-3}] = [F_{m-4}, F_{m-2}]$ are not squares, since the second half of each such word contains, by Lemma 5 (a), either a bit at position $2F_{m-3} + F_{m-4} - 1 = F_{m-1} - 1$ or at position F_{m-1}, whose complements appear at positions $F_{m-1} - 1 - F_{m-3} = F_{m-2} - 1$ and F_{m-2} in the first half.

We are now up to position $F_{m-2} + 1$. For $m \geq 4$, $f_m = f_{m-1} f_{m-2} = f_{m-2} f_{m-3}^2 f_{m-4}$. From Lemma 2 with $x = f_{m-3}$, $y = f_{m-4}$, we thus deduce that f_m contains the first $F_{m-4} + 1$ conjugates u of f_{m-3} such that u^2 is in f_m. These squares begin at $F_{m-2} + 1, \ldots, F_{m-2} + F_{m-4} + 1$. The last of these ends at position $F_{m-2} + F_{m-4} + 2F_{m-3} = F_m$. We have thus completed our scan, which yielded $F_{m-4}+1$ distinct squares of length $2F_{m-3}$. The conjugates of f_{m-3} of degrees $0, \ldots, F_{m-4} - 2$ appeared twice each in these squares, and the conjugates of degrees $F_{m-4} - 1$ and F_{m-4} appeared once each. ■

Lemma 7. *For* $m \geq 5$*, there are precisely* $F_{m-3}-1$ *conjugates* u *of* f_{m-2}*, of degrees* $0, \ldots, F_{m-3} - 2$*, such that* u^2 *belongs to* f_m.

Proof. The proof is similar to that of Lemma 6, but simpler. For $m \geq 4$, f_{m-3} is a prefix of $f_{m-2} = f_{m-3} f_{m-4}$. By Lemma 5 (b), we can thus apply Lemma 2 with $x = f_{m-2}$, $y = f_{m-3} s_2(m-3)^{-1} \overline{s}_2(m-3)$, to conclude that for $m \geq 5$, f_m contains $F_{m-3} - 1$ conjugates u of f_{m-2}, of degrees $0, \ldots, F_{m-3} - 2$, such that u^2 is in f_m. The last of those begins at position $F_{m-3} - 1$ and ends at position $(F_{m-3} - 1) + (2F_{m-2} - 1) = F_m - 2$ of f_m. Thus there are only two remaining words of length $2F_{m-2}$ in f_m. By Lemma 2 these could be squares only if the penultimate bit of f_m were the same as the penultimate bit of f_{m-3}; but it is not by Lemma 5 (b). ■

Definition 2. A *border* of a word x is a subword which is both a prefix and suffix of x. Denote by b_m the longest border of f_m, and let $B_m = |b_m|$. For example, $b_5 = 101$, so $B_5 = 3$.

In Lemma 2.5 of [**IMS≥1997**] it is proved that $B_m = F_{m-2}$. (Actually their method shows that even $b_m = f_{m-2}$.) Their Lemma 2.6 states that every conjugate of f_m of positive degree is different from f_m. We need the following generalization of this result:

Lemma 8. *For every integer $m \geq 0$, all the conjugates of f_m are distinct in pairs.*

Proof. The result holds for $m \leq 2$, so we may assume $m \geq 3$. Let $f_m = x_1 \ldots x_k$ $(k = F_m)$. Suppose that $x_{i+1} \ldots x_k x_1 \ldots x_i = x_{j+1} \ldots x_k x_1 \ldots x_j$ for some $i, j \in \{0, \ldots, k-1\}$ with $i < j$. Removing the suffix of length i and concatenating it to the beginning, we get $f_m = x_1 \ldots x_k = x_{j-i+1} \ldots x_k x_1 \ldots x_{j-i}$. Thus $x_{j-i+1} \ldots x_k$ and $x_1 \ldots x_{j-i}$ are both borders of f_m and one of them has length at least $F_m/2 > F_{m-2}$, contradicting the fact that $B_m = F_{m-2}$. ■

Proof of Theorem 1. Table 2 shows that the assertion holds for the first few values of n (but not for $n = 4$). Assume $n \geq 7$; then all the lemmas of this section can be applied.

TABLE 2.

n	F_n	$\lvert D(n)\rvert$
4	5	1
5	8	4
6	13	8
7	21	14
8	34	24

By Lemma 3, for $i \in \{1, \ldots, n-4\}$, the *maximum* number F_i of conjugates u of f_i have the property that u^2 is in f_n. By Lemma 6, precisely $F_{n-4} + 1$ conjugates u of f_{n-3} have the property that u^2 is in f_n. By Lemma 7, precisely $F_{n-3} - 1$ conjugates u of f_{n-2} have the property that u^2 is in f_n. Thus,

$$D(n) \geq \sum_{i=1}^{n-4} F_i + (F_{n-4} + F_{n-3}) = (F_{n-2} - 2) + F_{n-2} = 2(F_{n-2} - 1).$$

No conjugate of f_{n-1} can be a square in f_n, since $2F_{n-1} > F_n$. On the other hand, there is the following "folklore" result, proved in [**Séé1985**] (see Theorem 2.3 of [**IMS≥1997**]): *For every nonempty subword u^2 of f_n, u is a conjugate of f_m for some $m \geq 1$.* It follows that the last inequality is actually an equality. ■

Proof of Corollary 1. From Theorem 1, $D(n) = 2(F_{n-2} - 1) = 2\varphi^{-2} F_n + o(1) = 2(2 - \varphi) F_n + o(1)$. The second part follows from the well-known identity

$$F_n = \frac{\varphi^{n+1} - \overline{\varphi}^{n+1}}{\sqrt{5}} \quad (n \geq 0). \qquad ■$$

In fact, $D(F_n) \geq 0.75F_n$ for $n \geq 11$.

3. The Exact Number of Repeated Squares in Fibonacci Words

The first few values of $R(n)$, the number of repeated squares in f_n, found by computer, are listed in Table 3.

TABLE 3.

n	$R(n)$
0,1,2,3	0
4	1
5	4
6	11
7	26
8	57
9	118
10	235
11	454
12	857
13	1588
14	2899

Since $f_m = f_{m-1}f_{m-2}$, repeated squares in f_m can be of three kinds. First, they can lie entirely within the first F_{m-1} bits of f_m; there are clearly $R(m-1)$ such repeated squares. Secondly, they can lie entirely within the last F_{m-2} bits of f_m; there are $R(m-2)$ such repeated squares. Thirdly, they can intersect both the first F_{m-1} bits and the last F_{m-2} bits of f_m. This third kind of repeated squares will be called *boundary squares* of f_m, since they cross the boundary between f_{m-1} and f_{m-2}, and $f_m = f_{m-1}f_{m-2}$. Their number is denoted by $B(m)$. Thus,

$$R(m) = R(m-1) + R(m-2) + B(m). \tag{4}$$

We will get a recurrence relation for $R(m)$ by obtaining an expression for $B(m)$. To do so, we partition the boundary squares of f_m into two subsets.

Boundary squares of type 1. For $m \geq 5$ we have,

$$f_m = f_{m-1}|f_{m-2} = f_{m-2}f_{m-3}|f_{m-3}f_{m-4} = f_{m-2}f_{m-3}|f_{m-4}f_{m-5}f_{m-4}, \tag{5}$$

where the boundary is indicated by a vertical bar.

We see that every boundary square of $f_{m-2} = f_{m-3}f_{m-4}$ is also a boundary square of f_m. A boundary square of type 1 is a boundary square of f_{m-2}, which straddles the boundary of f_m. In other words, it is a boundary square which is contained in the copy of f_{m-2} which straddles the boundary of f_m. The number B_1 of such squares is $B(m-2)$, which, by (4), is

$$B_1 = B(m-2) = R(m-2) - R(m-3) - R(m-4). \tag{6}$$

Boundary squares of type 2. A boundary square of type 2 is a boundary square which does not lie inside the copy of f_{m-2} which straddles the boundary of f_m.

Clearly, every boundary square is either of type 1 or of type 2. In Lemmas 9–11 we count the number B_2 of boundary squares of type 2.

Lemma 9. *For $m \geq 8$, f_m contains no boundary squares of type 2 of length $2F_i$ for $i \leq m-6$. There are $F_{m-7}-1$ boundary squares of type 2 of length $2F_{m-5}$.*

Proof. For proving the first part, note that if a square of length $2F_i$ for $i \leq m-6$ would straddle the boundary of f_m, it would be contained in f_{m-2}, since $2F_{m-6} < F_{m-4}$ and $f_{m-2} = f_{m-3}|f_{m-4}$, so it would be of type 1, not type 2.

By Lemma 3, f_m contains $\geq F_{m-5}$ repeated squares of length $2F_{m-5}$. From (5),

$$\begin{aligned} f_m &= f_{m-2}f_{m-3}|f_{m-4}f_{m-5}f_{m-4} \\ &= f_{m-2}f_{m-4}f_{m-5}|f_{m-5}f_{m-6}f_{m-5}f_{m-4} \\ &= f_{m-2}f_{m-4}f_{m-5}|f_{m-5}f_{m-6}f_{m-6}f_{m-7}f_{m-4} \\ &= f_{m-2}f_{m-4}f_{m-5}|f_{m-5}f_{m-6}f_{m-7}f_{m-8}f_{m-7}f_{m-4} \\ &= f_{m-2}f_{m-4}f_{m-5}|f_{m-5}f_{m-5}f_{m-8}f_{m-7}f_{m-4}. \end{aligned} \tag{7}$$

Note that the last two bits of f_{m-4} are complements of the last two bits of f_{m-5}. Hence it follows from the last line of (7) and from Lemma 2 that precisely F_{m-5} squares of length $2F_{m-5}$ straddle the boundary. From the first and last lines of (7) it is easy to deduce that exactly $F_{m-4} - F_{m-5} + 1 = F_{m-6} + 1$ of the boundary squares of length $2F_{m-5}$ lie in the copy of f_{m-2} which straddles the boundary of f_m, so are of type 1. The rest, namely $F_{m-5} - (F_{m-6}+1) = F_{m-7} - 1$, are of type 2. ■

Lemma 10. *For $m \geq 7$, f_m contains $F_{m-6}-1$ boundary squares of type 2 of length $2F_{m-4}$.*

Proof. From (5) and the end of the proof of Lemma 1, we have, for $m \geq 7$,

$$\begin{aligned} f_m &= f_{m-2}f_{m-3}|f_{m-4}f_{m-5}f_{m-4} \\ &= f_{m-3}f_{m-4}^3 f_{m-7}f_{m-6}f_{m-5}f_{m-4}. \end{aligned} \tag{8}$$

A square of length $2F_{m-4}$ beginning at any position $\leq F_{m-1} - 2F_{m-4} + 1$, lies entirely to the left of the boundary of f_m, so is not a boundary square. Next consider the interval $I = [F_{m-1} - 2F_{m-4} + 2, F_{m-2}]$. Note that every square of length $2F_{m-4}$ in this interval is a boundary square. None of them is of type 1, since the last begins at the last bit of F_{m-2}, which, by the first line of (8), is just before the beginning of the copy of f_{m-2} which straddles the boundary.

Now $F_{m-1} - 2F_{m-4} + 2 = 2F_{m-3} - F_{m-4} + 2 > F_{m-3}$, and the length of I is $F_{m-2} - F_{m-1} + 2F_{m-4} - 1 = F_{m-6} - 1$. We thus see from the second line of (8), using Lemma 2, that each bit in this interval is the beginning of a square of length $2f_{m-4}$. All of these are type 2 squares.

A square of length $2F_{m-4}$ beginning at any position in the interval $[F_{m-2}+1, 2F_{m-2}-2F_{m-4}] = [F_{m-2}+1, 2F_{m-3}]$ lies entirely within the copy of f_{m-2} which straddles the boundary, and so is of type 1.

Now consider words of length $2F_{m-4}$ beginning in the interval $[2F_{m-3}+1, F_{m-1}] = [F_{m-1} - F_{m-4}+1, F_{m-1}]$. The first half of every such word w has a bit at position F_{m-1}, and its second half has a bit at position $F_{m-1}+F_{m-4} = 2F_{m-2}$. These are the terminal bits of the words f_{m-1} and f_{m-2} respectively by (2) of Lemma 4. For w to be a square, these two bits have to be the same. Hoewever, it is easily verified, by induction, that f_i ends in 0 if i is even, and in 1 if i is odd, so they cannot be the same. Thus there are no squares in this interval.

A square of length $2F_{m-4}$ beginning at any position $\geq F_{m-1}+1$ lies entirely to the right of the boundary, so is not a boundary square. This completes the left-to-right scan of f_m. ■

Lemma 11. *For $m \geq 6$, f_m contains $F_{m-4}+1$ boundary squares of type 2 of length $2F_{m-3}$, and $F_{m-3}-1$ boundary squares of type 2 of length $2F_{m-2}$.*

Proof. Lemma 6 gives a complete panorama of the locations of all the conjugates of f_{m-3}. By the "folklore" result, cited at the end of the proof of Theorem 1, this picture exhausts all squares of length $2F_{m-3}$.

The first batch of squares of length $2F_{m-3}$ ends at $(F_{m-4}-1)+2F_{m-3}-1 = F_{m-1}-2$, so these are not boundary squares. The next batch begins at $F_{m-2}+1$. Since $F_{m-2}+1+2F_{m-3}-1 = F_{m-1}+F_{m-3}$, the first of them is a boundary square. Since the last begins at $F_{m-2}+F_{m-4}+1 < F_{m-1}$, all these $F_{m-4}+1$ squares are, in fact, boundary squares of type 2: they cannot be of type 1, since they are too long to fit in the central copy of f_{m-2}.

Since $2F_{m-2} > F_{m-1}$ and $F_{m-3}-1 < F_{m-1}$, we see from Lemma 7 that all of the $F_{m-3}-1$ squares of length $2F_{m-2}$ are boundary squares, and clearly they are of type 2. ■

Collecting the results of Lemmas 9–11, we get,

$$B_2 = (F_{m-7}-1) + (F_{m-6}-1) + (F_{m-4}+1) + (F_{m-3}-1) = F_{m-5} + F_{m-2} - 2 = 2(F_{m-3}-1).$$

From (4) and (6) we thus get the recurrence,

$$R(n) - R(n-1) - 2R(n-2) + R(n-3) + R(n-4) = 2(F_{n-3}-1). \tag{9}$$

Note that (9) agrees with Table 4 for $n \geq 7$ (but not for $n = 6$).

Our last lemma is of a technical nature.

Lemma 12. *For $n \geq 2$, $\varphi^n = F_{n-1}\varphi + F_{n-2}$, $\overline{\varphi}^n = F_{n-1}\overline{\varphi} + F_{n-2}$.*

Proof. Induction on n. ■

Proof of Theorem 2. Replacing n by $n-1$ in (9), we get:

$$R(n-1) - R(n-2) - 2R(n-3) + R(n-4) + R(n-5) = 2(F_{n-4}-1).$$

Subtracting from (9) we then get:

$$R(n) - 2R(n-1) - R(n-2) + 3R(n-3) - R(n-5) = 2F_{n-5}. \tag{10}$$

This recurrence agrees with Table 3 for $n \geq 8$, but not for $n = 7$. The characteristic polynomial of (10) is: $x^5 - 2x^4 - x^3 + 3x^2 - 1$, and its roots are -1, 1 (double root), φ and $\overline{\varphi}$. Thus for the homogeneous part we have a solution of the form $c_1(-1)^n + (c_2 + c_3 n) + c_4\varphi^n + c_5\overline{\varphi}^n$. (See e.g., [**Liu1968**] for methods to solve linear recurrences.)

We assume that a particular solution has the form $a_0 n F_n + a_1(n-1)F_{n-1}$. Substituting into (10) and collecting terms, we then get:

$$\begin{aligned} & n\big(a_0(F_n - 2F_{n-1} - F_{n-2} + 3F_{n-3} - F_{n-5}) + a_1(F_{n-1} - 2F_{n-2} - F_{n-3} + 3F_{n-4} - F_{n-6})\big) \\ & + a_0(2F_{n-1} + 2F_{n-2} - 9F_{n-3} + 5F_{n-5}) + a_1(-F_{n-1} + 4F_{n-2} + 3F_{n-3} - 12F_{n-4} + 6F_{n-6}) \\ & = 2F_{n-5}. \end{aligned}$$

The coefficient of n ought to vanish, and it is easy to verify that $F_n - 2F_{n-1} - F_{n-2} + 3F_{n-3} - F_{n-5}$ is indeed identically 0. Simplifying we get, $(F_{n-3} + F_{n-5})a_0 + (2F_{n-3} - 3F_{n-5})a_1 = 2F_{n-5}$. Putting $a_0 - 3a_1 = 2$, $a_0 + 2a_1 = 0$, gives $(a_0, a_1) = (4/5, -2/5)$, so a particular solution has the form $\frac{4}{5}nF_n - \frac{2}{5}(n-1)F_{n-1}$. Thus the solution of (10) $(n \geq 8)$ has the form

$$R(n) = c_1(-1)^n + c_2 + c_3 n + c_4\varphi^n + c_5\overline{\varphi}^n + \frac{4}{5}nF_n - \frac{2}{5}(n-1)F_{n-1}.$$

We now evaluate this solution for $n \in \{8, \ldots, 12\}$, using Lemma 12 to simplify the powers of φ and $\overline{\varphi}$. This leads to the matrix equation $MC = B$, where

$$M = \begin{bmatrix} 1 & 1 & 8 & \frac{47}{2} + \frac{21}{2}\sqrt{5} & \frac{47}{2} - \frac{21}{2}\sqrt{5} \\ -1 & 1 & 9 & 38 + 17\sqrt{5} & 38 - 17\sqrt{5} \\ 1 & 1 & 10 & \frac{123}{2} + \frac{55}{2}\sqrt{5} & \frac{123}{2} - \frac{55}{2}\sqrt{5} \\ -1 & 1 & 11 & \frac{199}{2} + \frac{89}{2}\sqrt{5} & \frac{199}{2} - \frac{89}{2}\sqrt{5} \\ 1 & 1 & 12 & 161 + 72\sqrt{5} & 161 - 72\sqrt{5} \end{bmatrix},$$

and

$$B = R(n) - \big(\frac{4}{5}nF_n - \frac{2}{5}(n-1)F_{n-1}\big)\big|_{n=8}^{12} = \left[\, \frac{-509}{5},\ \frac{-846}{5},\ -279,\ \frac{-2286}{5},\ \frac{-3731}{5} \,\right],$$

where C is the vector of the coefficients c_i. The MAPLE package then produced the solution

$$C = \left[\, 0,\ 1,\ 1,\ -2 - \frac{4}{25}\sqrt{5},\ -2 + \frac{4}{25}\sqrt{5} \,\right],$$

which gives,

$$R(n) = \frac{4}{5}nF_n - \frac{2}{5}(n-1)F_{n-1} - \Big(2 + \frac{4}{5\sqrt{5}}\Big)(F_{n-1}\varphi + F_{n-2}) + \Big(-2 + \frac{4}{5\sqrt{5}}\Big)(F_{n-1}\overline{\varphi} + F_{n-2}) + n + 1.$$

After simplification this leads to the assertion in Theorem 2. Though it provably solves (10) for $n \geq 8$, it actually gives the correct values of $R(n)$ for $n \geq 3$ (but not for $n = 2$), as can be verified using Table 3. ■

Proof of Corollary 2. We combine the two leading terms of the solution given in Theorem 2 in the form,

$$R(n) = \frac{2}{5}nF_n + \frac{2}{5}nF_{n-2} - \frac{12}{5}F_{n-1} - F_{n-2}) + n + 1.$$

This gives,

$$R(n) = \frac{2}{5}(3-\varphi)nF_n + O(F_n) =$$

$$\frac{2}{5}(3-\varphi)F_n \log_\varphi F_n + O(F_n) = \frac{2(3-\varphi)}{5\log_2\varphi}F_n \log_2 F_n + O(F_n). \quad \blacksquare$$

References

1. [Cro1981] M. Crochemore, An optimal algorithm for computing the repetitions in a word, *Inform. Process. Letters* **12** (1981) 244–250.

2. [CrR1994] M. Crochemore and W. Rytter, *Text Algorithms*, Oxford University Press, New York, 1994.

3. [CrR1995] M. Crochemore and W. Rytter, Squares, cubes, and time-space efficient string searching, *Algorithmica* **13** (1995) 405–425.

4. [dLV1990] A. de Luca and S. Varricchio, A combinatorial theorem on p-power-free words and an application to semigroups, *RAIRO Inform. Thor. Appl.* **24** (1990) 205–228.

5. [FMT1978] A.S. Fraenkel, M. Mushkin and U. Tassa, Determination of $[n\theta]$ by its sequence of differences, *Canad. Math. Bull.* **21** (1978) 441–446.

6. [FrS≥1997] A.S. Fraenkel and R.J. Simpson, How many squares can a string contain? Preprint.

7. [IMS≥1997] C.S. Iliopoulos, D. Moore and W.F. Smyth, A Characterization of the squares in a Fibonacci string, *Theoret. Comput. Sci.*, to appear.

8. [Liu1968] C. L. Liu, Introduction to Combinatorial Mathematics, McGraw-Hill, New York, 1968.

9. [MiP1992] F. Mignosi and G. Pirillo, Repetitions in the Fibonacci infinite word, *RAIRO Inform. Thor. Appl.* **26** (1992) 199–204.

10. [Séé1985] P. Séébold, Propriétés combinatoires des mots infinis engendrés par certains morphismes, Report No. 85-16, LITP, Paris, 1985.

11. [Sto1976] K.B. Stolarsky, Beatty sequences, continued fractions, and certain shift operators, *Canad. Math. Bull.* **19** (1976) 473–482.